2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)

Evaluating IVR in Primary School Classrooms

Yvonne Chua* Priyashri Kamlesh Sridhar Haimo Zhang Vipula Dissanayake
Suranga Nanayakkara

Augmented Human Lab, Auckland Bioengineering Institute, University of Auckland

Figure 1: Empirical study to understand the influence of immersion and interactivity in learning in a real classroom context using performance, physiological and self-report measures.

Abstract

In recent years, the availability of affordable mobile Virtual Reality (VR) viewers has resulted in strong interests to incorporate Immersive Virtual Reality (IVR) within classrooms. However, studies on the effect of IVR on primary schoolers' learning are few, and they have often used equipment and settings far removed from everyday classroom instruction. We explored the role of interactivity and immersion in learning in a primary school classroom with 36 children aged 11–13 years, using commercially available devices that are ready-to-scale. We co-created content with different levels of immersion and interactivity together with teachers and investigated student engagement and learning. We present and discuss the use of multiple data sources (performance, physiological responses, observations and self-report) in a real-world classroom evaluation.

Index Terms: Human-centered computing—Interaction paradigms—Virtual reality; Human-centered computing—Human computer interaction—Empirical studies in HCI

1 Introduction

Much past research on VR for learning in classrooms has focused on non-immersive desktop VR [5, 13], largely because it is a mature technology and suited to widespread deployment in schools [14]. However, in recent years, IVR has become much more accessible to classrooms [16]. Low cost mobile VR viewers take advantage of smartphones that enable real-time display of high-resolution content from supporting educational IVR content platforms.

Presence, the sense of 'being there', is the game changer that IVR promises to bring [25]. Many researchers have linked presence to learning [14, 28] but empirical studies have failed to establish that experiencing a greater sense of presence results in increased learning [10]. Interactivity and immersion both contribute towards providing users with a sense of presence [25]. However, past studies examining these factors using IVR have not found any conclusive effect on learning [15, 20] and much remains to be investigated [14].

Pioneering research on IVR for childrens' learning have mainly utilized sophisticated equipment [2, 19] to explore the full potential of the technology. Those systems are quite different from the mobile VR viewers in classrooms today. Furthermore, such explorations are situated either in labs [19] or conducted as pull-out sessions [8, 9]. Since learning is highly influenced by the environment, there is a strong need to conduct such explorations in real-world classrooms.

In this paper, we discuss our experiences co-designing VR content with primary school teachers, its integration into their classroom curriculum and present the results of a study with 36 students. We used observations, performance, physiological and self-report measures to gain a more holistic understanding of learners' experiences at two levels of interactivity: passive consumption of video content and learner-paced interactive exploration. We investigated these two levels of interactivity in three viewing conditions providing increasing levels of immersion: a normal tablet screen (iPad), 'Magic Window' mode on the tablet and mobile VR viewer (Google Cardboard). 'Magic Window' on the tablet offers increased immersion compared to a normal screen by matching the movement of the tablet in space to the view displayed. Hwang et al. [7] found that a similar motion-based hand-held screen display increased users' self-reported presence compared to normal screen displays. We chose devices that are readily available to classrooms today so that findings would have immediate applicability for educators and content developers.

In very recent years, a few studies have started to explore the use of low-cost mobile viewers in low-resource classrooms [17, 26], and there has been strong interest in large-scale, in-the-wild investigations [23]. Our study lies between the lab and the wild, as a semi-controlled study conducted in a real-world setting, with learning content designed to fit into the classroom curriculum. We contribute with 1) A discussion of the role of multiple data sources in understanding the influence of immersion and interactivity in learning in a real-world classroom context; and 2) Insights and lessons learned from conducting such a study in the classroom.

2 Method

2.1 Content Creation: Co-designing with Teachers

The learning content used in the study was co-designed with two teachers from a blended Year 7 and 8 primary school classroom to supplement their inquiry (science) lessons. Three co-design sessions (20~50min) took place over 5 weeks. The inquiry topic for the term was bridges and structures. The key learning outcomes for the topic were, 1) to understand how structures are designed and for what purpose; and 2) to investigate what makes a good structure and how to evaluate the success of its function. We began by having teachers

*e-mail: yvonne@ahlab.org

978-1-7281-4765-9/19/$31.00 ©2019 IEEE
DOI 10.1109/ISMAR-Adjunct.2019.00-53

Figure 2: a. Build the First Bridge, b. Time Travelling Mailman, c. Content types for the experiment.

put together a picture of the ideal learning experience they wanted for their students with the aid of picture prompts (Fig.3). Next we discussed gaps and challenges in the current learning experience, as well as what they wanted IVR to bring to the table. Teacher motivations for exploring the use of IVR included:

- Giving students a better way to *"connect to content"*, as *"what they are learning needs to make sense to them and be relatable"*
- Opening doors to experiences and places students can't otherwise access due to financial, time and practical constraints
- Generating excitement, fun and interest; as a way to 'hook' students into learning

We came up with 6 concepts based on the useful learning outcomes identified by teachers in our discussion and their lesson plan for the term, then shortlisted and refined 2 concepts based on feedback from both teachers. Both concepts were based around the history of a bridge located near the school and are described below:

Content 1: Build the First Bridge (BB)
Learning Goal: To understand how structures are designed, and more specifically the history and construction of the first bridge.
Experience: Students have to build the first bridge across the river (this bridge is no longer standing today). In the interactive version, they select materials and components to start building the bridge. Steps need to be selected in the correct sequence for the bridge to be successfully constructed. For example, if the bridge deck is chosen before piles are driven to support it the deck splashes on the river and floats away. Audio narration provides guidance on what to consider if a step is selected in the wrong sequence (e.g., *"try first adding some supports for the deck"*) and also explains how materials were used to build the bridge after they have been selected (e.g., *"wooden piles were driven 5 metres deep into the riverbed"*). In the passive version, students watch the bridge being built with audio narration providing the same information. Students in the Magic Window (MW) and IVR Condition had control over their view (similar to watching a 360 video) but were not able to otherwise interact with the environment or control the pace of the content. Students in the screen condition did not have any control over the view.

Figure 3: One teacher's ideal learning experience

Content 2: The Time Travelling Mailman (MM)
Learning Goal: To understand what purposes structures are designed for and what makes a good structure.
Experience: Students get a first hand experience of what it was like to cross the river in the past. They are tasked with delivering a letter and parcel across the river at two time points in the 1800s. At the first time point the bridge has not been built yet and they have to use a punt to cross the river. At the second time point students have to pay a toll and wait for the swing span of the bridge to open and close to let a boat pass before crossing the river, just like what foot passengers had to do at the time. Audio narration provides guidance on what actions to take and supplies information on the punt and bridge. In the interactive version, students are able to explore the environment at their own pace by selecting objects and teleporting to various locations using a button press or tapping the screen. In the passive version, students watched a first-person view of events unfolding with audio narration providing the same information.

2.2 Participants

Thirty-six Year 7 and 8 primary school students (23 male, 13 female) participated in the experiment. Students ranged from 11 to 13 years old (mean age = 11.97 years, SD = 0.71). All students were part of the same classroom but were separated into 2 teaching groups for inquiry lessons, with one group (n=12) generally receiving more structured and scaffolded instruction. Two weeks prior to the study, this group learnt about the history of the bridge in two lessons, and the teacher provided additional scaffolding by having students learn content-related vocabulary (punt, piles, beams, abutment) through a fun activity. Both groups watched a short 4 minute video on the history of the bridge in class the week before the study.

2.3 Design

A between subjects design was used to investigate the effect of 3 viewing conditions (IVR; MW, Screen (SC)) and 2 interactivity conditions (Interactive; Passive). Two different content types (BB and MM) were used and we examined the interactive/passive difference for both types. We used the following measures:

- *Physiological Measures:* The Empatica E4 wristband was used to collect electrodermal activity and heart rate data.
- *Observations:* Video and screen recordings were taken to aid in the triangulation of physiological measures.
- *Self-Report:* The Smileyometer, Again-Again, and Fun Sorter (best to worst learning; most to least fun) components of the fun toolkit [18] were used.
- *Learning Questionnaires:* Learning questionnaires were designed with teacher feedback, based on the specific learning goals for each content, as well as overall learning goals for the topic.
- *Other measures:* Students also completed pre- and post-balance [27] and stereoacuity [3] tests as well as child simulator sickness questionnaires [6].

2.4 Apparatus

iPads (9.7") from the school were used for the MW and SC conditions. Google Cardboard viewers with mobile phones were used for the IVR condition. Stereo earphones were used for all conditions. Interactive and passive versions were developed using Unity3d for the 3 viewing conditions (Fig.2c, Tab.1).

2.5 Procedure

We conducted 2 study sessions on the same day, 1 for each inquiry group (n=12, n=24). All students were present in the same room, clustered at different tables based on viewing condition. Six researchers were present on the day of the study.

First, students completed learning questionnaires. They were given 4 minutes to complete Q1 and Q2 together, and 4 minutes each for Q3, Q4 and Q5. Next, all students viewed *BB* followed by *MM*, with interactive and passive content counterbalanced. They viewed either *interactive BB + passive MM*, or *passive BB + interactive MM*. Students completed Q1 and 2 after viewing *BB* and Q3, 4 and 5 after viewing *MM*. Both types of content lasted 2~4 minutes. After viewing all the content, students completed another balance test as well as the fun toolkit and child simulator sickness questionnaire [6].

Due to a limited number of devices, E4 data was collected from a subset of students. Students were randomly selected within conditions and gender balanced (12 male, 10 female; IVR 8, MW 7, SC 7). These students wore the wristbands and completed a baseline period of sitting quietly and relaxing for 5 minutes before starting the pre-test questionnaires.

Learning questionnaires were completed in pairs so students would be more familiar and comfortable with the activity. We discovered in the co-design sessions that there was a strong emphasis on team support and collaborative learning in the classroom, with almost all assignment in the class carried out in pairs and groups instead of individually. Students were paired in same-gender pairings as much as possible. One week later, we returned to the class and students completed learning questionnaires (Q1 - Q5) again. In addition, the sessions were video-taped from different points of view in the classroom.

2.6 Filtering of data

The review of the video recordings for the session revealed several possible confounding factors. To reduce the effect of these factors on the data, we decided to exclude the data of students based on behaviours observed. The list of video timestamps and videos were passed to 3 evaluators to code for whether the participant was a) watching the content, b) watching but slightly distracted, c) not watching content or highly distracted, d) watching the same content multiple times or e) watching content when answering questions. If two or more evaluators coded a participant as showing behaviours c, d, or e, their data was excluded from the analysis. Based on this, we identified 8 participants. As participants completed the questionnaires in pairs, excluding one in the pair meant removing the data for both. Therefore, 14 participants (2 of them were in the same pair) were excluded through this process. Of the 22 remaining participants, we had E4 data for 15 participants (30 viewing sessions). However, data was missing or too noisy for some viewing sessions (EDA: 4; heart rate: 6) due to poor contact with the sensor, extreme movements and maybe increased oiliness of the skin at the area of contact. Conditions were not balanced after exclusion. Tab.2 shows the number of participants for data analysis.

Table 1: Viewing conditions and interaction methods

Viewing Condition		Object Selection
IVR	Mobile device in a Google Cardboard viewer	*Moving head to align the reticle (white circle in the middle of the screen) with object and pushing a button on the side of the viewer*
MW	Tablet in Magic Window mode	*Moving the tablet to align the reticle with the object and tapping the screen*
SC	Tablet in normal mode	*Tapping on screen*

Table 2: No. of participants for data analysis after exclusion (Participants with complete E4 data)

	IVR	MW	SC	Totals
Interactive BB; Passive MM	2 (2)	6 (3)	0	8 (5)
Passive BB; Interactive MM	4 (2)	6 (1)	4 (2)	14 (5)
Totals	6 (4)	12 (4)	4 (2)	22 (10)

3 Results and Discussion

3.1 Influence of interactivity on learning

3.1.1 Performance:

Scores for each question were first calculated as a percentage of the full score possible. For open-ended questions the highest score amongst the students was used as the full score. Performance was then calculated based on the percentage gain in score from pre to post and delayed time points to account for different starting points between pairs. We mapped questions with content to score for performance. Q1 and Q2 related to BB content. Students viewing interactive content showed larger performance gains between delayed and pre-tests (median = 159%) compared to students viewing passive content (median=103%). A Mann-Whitney U test showed that the difference was not significant (W=22, p=.131). Q3 related to MM content. Students viewing interactive content showed slightly larger performance gains between delayed and pre-tests (median=50%) compared to students viewing passive content (median=33%). A Mann-Whitney U test showed that the difference was not significant (W=18, p=.442).

3.1.2 Self-report:

In the Fun Toolkit, students rated the content they watched in the Smileyometer on a scale of 1 ("Awful") to 5 ("Brilliant"); were asked if they would like to watch them again in Again-Again; and sorted the two types of content they watched in the Funsorter in terms of fun and learning. Students rated both types of content similarly in the Smileyometer. More students stated that they would like to watch interactive content again (n=14) compared to passive content (n=4). In the Funsorter, 2 students rated both types of content as having the same learning and fun. Of the remaining 20 students, 16 rated interactive content as being more fun than passive content (5 BB, 11 MM), and 11 rated interactive content as having better learning (6 BB, 5 MM). Fourteen students rated BB content as having better learning than MM content.

3.1.3 Physiological:

We extracted some physiological markers that have been shown to be potentially sensitive to engagement and increased mental effort in children [24].

For electrodermal activity, we elicited skin conductance responses (SCRs) that refer to peaks in skin conductance and correspond with increased arousal/excitement. Due to individual differences

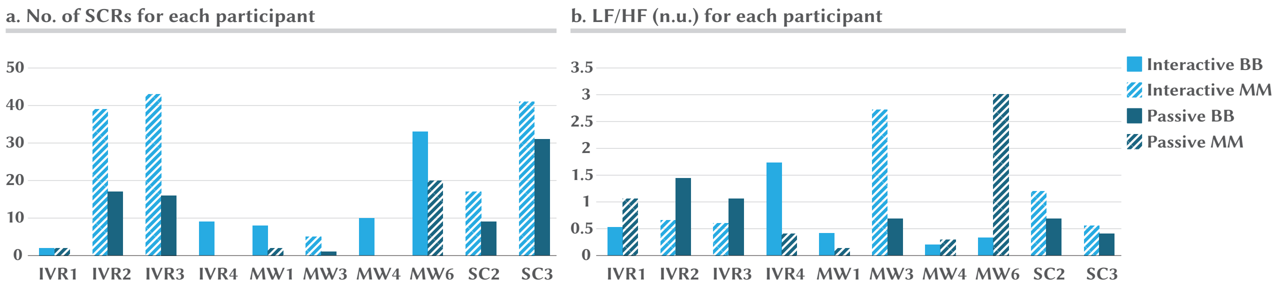

Figure 4: a. Number of Skin Conductance Responses (SCRs) and b. HRV LF/HF for each participant (n=10)

in number of SCRs, averaging values across participants in interactive/passive conditions for a between-subjects comparison did not give a good representation of the effect of interactivity. We compared SCR data on an individual level, for participants who had recorded data for both interactive and passive sessions (n=10) and found that all participants showed higher number of SCRs as they interacted with interactive content (Fig.4a) indicating increased engagement as compared to the passive condition regardless of content type (MM or BB).

From the heart rate measures, we conducted a time-domain and frequency-domain analysis particularly focusing on the low frequency (LF) and high frequency (HF) components of heart rate variability that were calculated as the area under the Power Spectrum Density curve corresponding to 0.04 - 0.15 Hz and 0.15 - 0.4 Hz respectively. We normalized the LF and HF to minimize impact of the difference in total power and computed the LF/HF ratio that provides an estimate of sympathetic modulation and has been shown to be a proxy of mental effort and cognitive load [12]. A higher HRV LF/HF value corresponds to increased mental effort. We found that cognitive load seemed to vary across conditions and individuals (Fig.4b).

3.1.4 Combining data sources and observations:

The interactive condition elicited better performance scores and better engagement, shown as the higher number of SCRs and supported by the higher self-reported fun rating from students. This was in spite of varying mental effort, and mixed self-reported learning ratings.

There were large individual differences in SCRs and cognitive load (HRV LF/HF). Examining data on an individual level revealed insights into students' experience of the content. IVR1 and IVR4 both viewed immersive, interactive BB content. IVR1 had lesser engagement (2 SCRs) than IVR4 (9 SCRs) and a rise and fall of cognitive load over time while IVR4 had a continuous increase in effort over time. The screen recording for IVR1 revealed that higher cognitive load occurred during periods of multiple processing like listening to explanations, scanning choices and choosing a response. Interestingly the highest cognitive load mapped to the time when he made an error and was given corrective feedback and an explanation.

E4 data helped to supplement our understanding of covert behaviours which was useful as there were some challenges in coding facial expressions due to the viewer obstructing the child's eyes. IVR2 despite being calm and reserved when viewing interactive content without displaying salient events in the video, had a high number of steadily increasing SCRs thereby telling us she was attentive/aroused.

A content difference was seen with larger performance score gains for BB compared to MM content, supported by students' self-reported learning ratings. This could be due to the nature of BB content in eliciting a more active involvement from the participant as it required an experimental trial and error approach to figure out the right sequence of steps compared to following a linear narrative in MM content. However no obvious content differences were seen in physiological data.

3.2 Influence of immersion on learning

3.2.1 Performance:

We did not find a clear difference between the combined performance scores for each of the different viewing conditions for both pre-post as well as pre-delayed gains (Fig.5a and b). This could be due to the fact that the learning content did not take advantage of enough IVR affordances such as representing spatial and abstract experiences. While we intended to give students a first-hand experience of the past, it is also possible that the quality of the graphics for the environment was insufficient to create that experience. Another factor to consider is the level of immersion of the IVR system in the study (a cardboard viewer with 3 degrees of freedom (DOF)). Several studies that show immersion increasing learning used more immersive systems (6 DOF HMDs with controller and body movement tracking) [4]and Alhalabi [1] found indications that within IVR, students viewing content in systems with higher immersion performed better.

Looking into individual questions revealed that gains for Q1 and Q2 were considerably higher than that of other questions, with the students in the IVR condition having comparatively higher score gains to other conditions. This could be due to the visual-matching required for Q1 - findings from Rasheed et al. [17] suggest that IVR helped students answer questions related to spatial position and colour better compared to normal instruction. Students mostly performed worse on Q4 in post tests (Fig.5a) but better on delayed tests. We observed that several students questioned the point of completing repeated evaluations and expressed dissatisfaction *"we have to do this again?"*. This was particularly the case for Q4 and Q5, open-ended questions that asked them to generate as many responses as they could think of. Many of them put in shorter replies for these in the post-evaluation.

3.2.2 Self-report:

Students in more immersive viewing conditions rated the content higher and were more inclined to repeat the experience (Fig.5c).

3.2.3 Physiological:

We did not find any trends in SCRs and HRV LF/HF between viewing conditions. There were large individual differences between students in the same viewing condition, particularly the MW condition (Fig.4) and instances of large differences in cognitive load in the same child when viewing different types of content (Fig.4b: IVR4, MW3, 6) as well.

3.2.4 Combining data sources and observations:

Students' self-reported preferences for more immersive content were not reflected in other measures. The large variances in physiological data suggests that individual differences and content differences might outweigh viewing condition effects in our study. This could

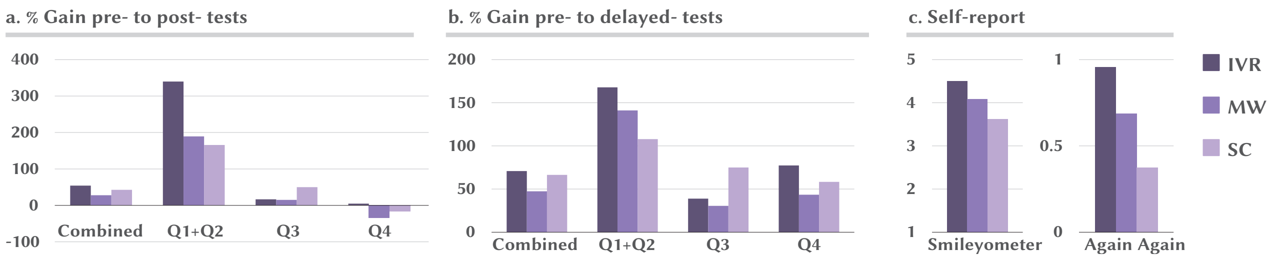

Figure 5: a. Performance score gains from pre to post tests and b. from pre to delayed tests, c. Self report scores (n=22)

be one reason for the lack of a trend in performance data and is supported by video data showing that students responded to the same viewing condition differently. In the Magic Window condition for example, some students remained fairly static, resting the iPad on the table, only lifting or tilting it occasionally (MW1, 4). Others spun around constantly in their seats, moving the iPad at various angles to see different parts of the world (MW2, 3). Others still started leaving their seats and walking around (MW5, 6). Movement could have been a reflection of how engaged they felt, or might have served to increase their engagement in the virtual world. Slater [22] showed that self-reported presence increased with whole body movement and Markowitz et al. [11] found that participants' knowledge improvement on ocean acidification was positively correlated to the amount of exploration of the virtual environment.

3.3 Interesting Observations and Teacher Feedback

We observed a general sharing of experience and excitement amongst students even though they all viewed content on separate screens. Students discussed on-screen content and guided their peers to look at specific parts of the virtual environment (Fig.6). We saw this in all viewing conditions, and when viewing both interactive and passive content. Even without collaboration built into the virtual world, IVR was by no means isolating. One participant pair viewing passive IVR content continuously updated each other on their experience as they both looked into separate viewers:
IVR1: *"I'm now on the boat."*
IVR4: *"I'm on the boat too! I'm going, I'm going!"*

Videos revealed a variety of verbal exclamations in IVR (Wow), appreciation (Oh I love the sky), remarks (This is like 3D, I'm walking, psych) and questions (Oh where to go? I need to pay a toll? How much?) on ongoing events. After participants 'finished' watching the content with the successful completion of the bridge and delivery of the parcel, they continued to explore the scene and discussed what they did with each other, with some in the IVR content swapping viewers as well. The teachers commented that they found IVR to be a very useful tool in teaching the content (bridges and structures) especially with scaffolding.

3.4 Limitations and Lessons Learned

3.4.1 Running controlled studies in classrooms:

While we expected to face challenges in running a classroom based study, we were surprised by how much the dynamics of the class affected data collection and procedure especially for the second and larger group where a 1:4 researcher to student ratio was insufficient to keep track of all students. We had to exclude quite a few participants to reduce the confounding factors such as not watching the content due to distraction or watching it multiple times. The number of participants in each condition was also limited because of the complexity of the experimental design. Further research with larger populations is needed to make conclusions on the effect of immersion and interactivity on learning. We plan to start future studies with more controlled pull-outs to tease out the effects of specific variables of IVR content (type and degree of interactivity; IVR affordances used) followed by class-wide in-the-wild implementation in the hands of teachers to access a larger participant population to see if results can be replicated in a natural setting.


Figure 6: Shared experiences across viewing conditions

3.4.2 Physiological data collection in real world settings:

There are challenges in collecting physiological measures in classrooms (availability of devices, missing HRV data in ambulatory settings, presence of other people and interactions, reduced specificity in SCRs). As a result, we had to rule out some participants from the physiology analysis. However these challenges can be partially mitigated by using combined data and interpreting holistically, using one source of data to clarify another; tracking individually as illustrated to get a deeper understanding into participants. It would also be useful to have this data from a more one-on-one session as a yardstick, monitoring this data in intervals.

3.4.3 Single short session:

With current hardware, learning through IVR is a necessarily short experience due to concerns with visual fatigue and simulator sickness. In this study students only viewed each content once and for short period of time, and that may have limited the potential learning gains. Studying the effect of multiple sessions, perhaps on the reinforcement of a tricky concept, would be a valuable extension of the study.

3.4.4 Topic selection and design of content:

In our study content was co-designed with teachers to fit in with their curriculum for inquiry for the school term. While we did find higher performance gains for IVR condition for some of the questions, we believe that having content that takes advantage of more IVR affordances may evoke even stronger patterns when comparing levels of interactivity and immersion. This can be done by looking at topics planned for the entire year and across subjects, to select one that can take advantage of multiple IVR affordances.

3.4.5 Rethinking Pre- Post- Performance evaluations:

To address the issue of student disatisfaction with completing repeated questionnaires, and the confounding effect that it had on post-test scores, we plan to design performance evaluations for future studies in either a more engaging manner, for example as a game or puzzle [21] or in a way that disguises the repetition of questions.

3.4.6 Continuity of IVR in the classroom:

Prior to the study, we ran a short introductory workshop on how to deploy IVR in the classroom using available platforms (Google Expeditions, Google Poly, Tour Creator, CoSpaces). We were encouraged by the fact that teachers expressed interest in more in-depth training on creating their own content after the end of the study. One of the teachers even commented on using IVR to enhance literacy in her class in the future. We believe that the introductory workshop and co-creation sessions played an important role not just to get their support but also inspire the use of it beyond experiments and research.

4 Conclusion

Our study is one of the few explorations of the effect of interactivity and immersion on learning in a real classroom setting. We used content co-designed with teachers in accordance with the class curriculum. Participants viewed two types of content, one each in interactive and passive modes and through one of the three conditions: Immersive Virtual Reality, Magic Window and screen.We found indications that interactivity was more engaging, more enjoyable and might result in better performance compared to the non-interactive condition. We did not find clear indications on the effect of immersion. We discuss individual differences, the role of different data sources in understanding children's viewing experience and present takeaways for running a study in the classroom, designing pre- post-evaluations and minimizing other confounding variables.

5 Acknowledgements

This work was supported by *Understanding the effects of VR on children* research grant by Soremartec S.A. The content is solely the responsibility of the authors and does not necessarily represent the official views of Soremartec S.A.

References

[1] W. S. Alhalabi. Virtual reality systems enhance students' achievements in engineering education. *Behaviour & Information Technology*, 35(11):919–925, 2016.

[2] D. Allison, B. Wills, D. Bowman, J. Wineman, and L. Hodges. The Virtual Reality Gorilla Exhibit. *IEEE Computer Graphics and Applications*, 17(6):30–38, Dec. 1997. doi: 10.1109/38.626967

[3] E. Birch, C. Williams, J. Drover, V. Fu, C. Cheng, K. Northstone, M. Courage, and R. Adams. Randot Preschool Stereoacuity Test: Normative data and validity. *Journal of American Association for Pediatric Ophthalmology and Strabismus*, 12(1):23–26, Feb. 2008. doi: 10.1016/j.jaapos.2007.06.003

[4] S. W. Greenwald, W. Corning, M. Funk, and P. Maes. Comparing learning in virtual reality with learning on a 2d screen using electrostatics activities. *Journal of Universal Computer Science*, 24(2):220–245, 2018.

[5] K. F. Hew and W. S. Cheung. Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, 41(1):33–55, Jan. 2010. doi: 10.1111/j.1467-8535.2008.00900.x

[6] R. M. Hoeft, J. Vogel, and C. A. Bowers. Kids Get Sick Too: A Proposed Child Simulator Sickness Questionnaire. *Proc. HFES*, 47(20):2137–2141, Oct. 2003. doi: 10.1177/154193120304702013

[7] J. Hwang, J. Jung, and G. J. Kim. Hand-held virtual reality: a feasibility study. In *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '06*, p. 356. ACM Press, Limassol, Cyprus, 2006. doi: 10.1145/1180495.1180568

[8] R. L. Jackson and E. Fagan. Collaboration and learning within immersive virtual reality. In *Proc. CVE*, pp. 83–92. ACM Press, 2000. doi: 10.1145/351006.351018

[9] A. Johnson, T. Moher, S. Ohlsson, and M. Gillingham. The Round Earth Project-collaborative VR for conceptual learning. *IEEE Computer Graphics and Applications*, 19(6):60–69, Dec. 1999. doi: 10.1109/38.799741

[10] G. Makransky, T. S. Terkildsen, and R. E. Mayer. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction*, 2017.

[11] D. M. Markowitz, R. Laha, B. P. Perone, R. D. Pea, and J. N. Bailenson. Immersive virtual reality field trips facilitate learning about climate change. *Frontiers in Psychology*, 9, 2018.

[12] D. J. McDuff, J. Hernandez, S. Gontarek, and R. W. Picard. Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera. In *Proc. CHI*, pp. 4000–4004. ACM, 2016.

[13] Z. Merchant, E. T. Goetz, L. Cifuentes, W. Keeney-Kennicutt, and T. J. Davis. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70(Supplement C):29–40, Jan. 2014. doi: 10.1016/j.compedu.2013.07.033

[14] T. A. Mikropoulos and A. Natsis. Educational virtual environments: A ten-year review of empirical research (19992009). *Computers & Education*, 56(3):769–780, Apr. 2011. doi: 10.1016/j.compedu.2010.10.020

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

[16] E. Olmos, J. F. Cavalcanti, J.-L. Soler, M. Contero, and M. Alcaiz. Mobile Virtual Reality: A Promising Technology to Change the Way We Learn and Teach. In *Mobile and Ubiquitous Learning*, Perspectives on Rethinking and Reforming Education, pp. 95–106. Springer, Singapore, 2018. doi: 10.1007/978-981-10-6144-8_6

[17] F. Rasheed, P. Onkar, and M. Narula. Immersive Virtual Reality to Enhance the Spatial Awareness of Students. In *Proc. IndiaHCI*, IndiaHCI'15, pp. 154–160. ACM, New York, NY, USA, 2015. doi: 10.1145/2835966.2836288

[18] J. C. Read. Validating the Fun Toolkit: an instrument for measuring childrens opinions of technology. *Cognition, Technology & Work*, 10(2):119–128, Apr. 2008. doi: 10.1007/s10111-007-0069-9

[19] M. Roussos, A. Johnson, T. Moher, J. Leigh, C. Vasilakis, and C. Barnes. Learning and Building Together in an Immersive Virtual World. *Presence: Teleoperators and Virtual Environments*, 8(3):247–263, June 1999. doi: 10.1162/105474699566215

[20] M. Roussou and M. Slater. Comparison of the Effect of Interactive versus Passive Virtual Reality Learning Activities in Evoking and Sustaining Conceptual Change. *IEEE Transactions on Emerging Topics in Computing*, pp. 1–1, 2017. doi: 10.1109/TETC.2017.2737983

[21] S. D. Scott, R. L. Mandryk, and K. M. Inkpen. Understanding children's collaborative interactions in shared environments. *Journal of Computer Assisted Learning*, 19(2):220–228, June 2003. doi: 10.1046/j.0266-4909.2003.00022.x

[22] M. Slater, J. McCarthy, and F. Maringelli. The influence of body movement on subjective presence in virtual environments. *Human Factors*, 40(3):469–477, 1998.

[23] E. Southgate, S. P. Smith, C. Cividino, S. Saxby, J. Kilham, G. Eather, J. Scevak, D. Summerville, R. Buchanan, and C. Bergin. Embedding immersive virtual reality in classrooms: Ethical, organisational and educational lessons in bridging research and practice. *International Journal of Child-Computer Interaction*, 2018.

[24] P. K. Sridhar, S. W. Chan, and S. Nanayakkara. Going beyond performance scores: understanding cognitive-affective states in kindergarteners. In *Proc. IDC*, pp. 253–265. ACM, 2018.

[25] J. Steuer. Defining Virtual Reality: Dimensions Determining Telepresence. *Journal of Communication*, 42(4):73–93, Dec. 1992. doi: 10.1111/j.1460-2466.1992.tb00812.x

[26] A. Vishwanath, M. Kam, and N. Kumar. Examining Low-Cost Virtual Reality for Learning in Low-Resource Environments. In *Proc. DIS*, pp. 1277–1281. ACM Press, 2017. doi: 10.1145/3064663.3064696

[27] E. N. Williams, S. G. Carroll, D. S. Reddihough, B. A. Phillips, and M. P. Galea. Investigation of the timed Up & Go test in children. *Developmental Medicine and Child Neurology*, 47(8):518–524, Aug. 2005. doi: 10.1017/S0012162205001027

[28] W. Winn, M. Windschitl, and R. Fruland. When Does Immersion in a Virtual Environment Help Students Construct Understanding? In *Proc. ICLS*, pp. 497–503, 2002.