




Comparison of VR and Desktop Game User Experience in a Puzzle Game: “Keep Talking and Nobody Explodes”

Mehmet İlker BERKMAN*, Bahçeşehir University, Communication Design Department,
Dr., ilker.berkman@comm.bau.edu.tr,  0000-0002-2340-9373

Güven ÇATAK, Bahçeşehir University, Digital Game Design Department, Dr.,
guven.catak@comm.bau.edu.tr,  0000-0002-4679-8973

Mıstık Çağın EREMEKTAR, Bahçeşehir University, Game Design Program, MA,
uacagin@hotmail.com,  0000-0002-1250-6653

ABSTRACT

Since the contemporary game production process is based on the Integrated Development Environment (IDE) applications, it is easier for developers to create multiple versions of their game for both VR and desktop platforms. This provided a great opportunity for researchers to conduct comparative studies to explore the user experience of the relatively novel virtual reality applications. In this study, we evaluated a puzzle game through a within-subjects experiment design using objective measures of game success and gameplay duration, as well as plenty of subjective measures in order to assess game user experience, comparing desktop and VR. In addition to selected dimensions of GUESS (Game User Experience Satisfaction Scale), we employed MEC-SPQ (Measurement Effects Conditions - Spatial Presence Questionnaire) to measure presence. Furthermore, we employed NASA-TLX (NASA Task Load Index) to compare the perceived task complexity of the same task executed in VR and desktop gaming environments. Results revealed that there is not a significant difference in objective measures of player performance, comparing the VR and desktop gameplay. The Game User Experience Satisfaction Scale did not reveal any significant difference between the mean scores of VR and desktop experiences. The spatial presence related dimensions of MEC-SPQ revealed significantly higher scores of VR, for Possible Actions and Self Location dimensions. NASA-TLX weighted scores were significantly higher for VR in physical load and for desktop in frustration. Our results show that a puzzle-based game experienced in VR does not lead to a higher level of satisfaction in terms of game user experience but triggers a sense of spatial presence. Due to the different control schemes, players perceive that HMD based gameplay demands more physical task load. However, the gameplay duration and game success rate are not significantly different. The failure in desktop gameplay might have led to higher frustration, since the experience seems more familiar to players. Since the results are partially concordant with previous studies, it is not possible to make a strict conclusion on the effect caused by different immersive technologies on game user experience. Further studies are required through a more consistent methodology with a focus on game design components rather than game genre.

*Corresponding author

Keywords : *Virtual Reality, Game User Experience, Puzzle Game, Task Load, Presence, Video Games*

Bir Bulmaca Oyununda VR ve Masaüstü Oyun Kullanıcı Deneyimi Kıyaslaması: “Keep Talking and Nobody Explodes”

ÖZ

Çağdaş oyun üretim süreci Entegre Geliştirme Ortamı (IDE) uygulamalarına dayandığından, geliştiricilerin oyunlarının hem VR hem de masaüstü platformları için birden çok sürümünü oluşturmaları daha kolaydır. Bu durum, araştırmacılara da, sanal gerçeklik uygulamalarının kullanıcı deneyimini karşılaştırmalı olarak değerlendirme fırsatı sağlamaktadır. Bu çalışmada, oyun başarısı ve oynanış süresi objektif ölçümlerinin yanında, oyun kullanıcı deneyimini değerlendirmek için birçok öznel ölçümü de kullanarak bir bulmaca oyununu masaüstü ve VR’yi karşılaştırması yaparak değerlendirdik. GUESS ölçeğinin (Oyun Kullanıcı Deneyimi Memnuniyeti Ölçeği) seçilen boyutlarına ek olarak, mevcudiyeti ölçmek için MEC-SPQ (Ölçüm Etkileri Koşulları - Mekansal Mevcudiyet Anketi) kullanıldı. Ayrıca, VR ve masaüstü oyun ortamlarında yürütülen aynı görevin algılanan görev karmaşıklığını kıyaslamak için NASA-TLX (NASA Görev Yük Endeksi) ölçeğinden yararlanıldı. Sonuçlar, VR ve masaüstü ortamı karşılaştırıldığında, oyuncu performansının objektif ölçümlerinde önemli bir fark olmadığını ortaya koydu. Oyun Kullanıcı Deneyimi Memnuniyeti Ölçeği, VR ve masaüstü deneyimlerinin ortalama puanları arasında anlamlı bir fark görülmedi. MEC-SPQ’nun mekansal mevcudiyetle ilgili boyutları olan Olası Eylemler ve Öz-Konum boyutları için VR ortamında anlamlı biçimde daha yüksek skorlar gözlemlendi. NASA-TLX ağırlıklı skorlar fiziksel yük boyutunda VR ve hayal kırıklığı boyutunda masaüstü için anlamlı olarak daha yüksektir. Sonuçlar, VR’de deneyimlenen bulmaca tabanlı bir oyunun, oyun kullanıcı deneyimi açısından daha yüksek bir memnuniyet düzeyine yol açmadığını, ancak mekansal mevcudiyet duygusunu tetiklediğini göstermektedir. Farklı kontrol şemaları nedeniyle, oyuncular kafaya takılan ekran tabanlı oyunun daha fazla fiziksel görev yükü gerektirdiğini düşünmektedirler. Bununla birlikte, oyun süresi ve oyun başarı oranı önemli ölçüde farklı değildir. Masaüstü oyunlarındaki başarısızlık, deneyim oyunculara daha tanıdık geldiği için daha yüksek hayal kırıklığına yol açmıştır. Sonuçlar daha önceki çalışmalarla kısmen uyumlu olduğu için, farklı sürükleyici teknolojilerin oyun kullanıcı deneyimi üzerindeki etkisine ilişkin kesin bir sonuç çıkarmak mümkün değildir. Oyun türünden ziyade oyun tasarımı bileşenlerine odaklanan daha tutarlı bir metodoloji yoluyla daha fazla çalışmaya ihtiyaç vardır.

Anahtar Kelimeler : *Sanal Gerçeklik, Oyun Kullanıcı Deneyimi, Bulmaca Oyunu, Görev Yükü, Mevcudiyet, Bilgisayar Oyunları*

INTRODUCTION

As the virtual reality re-gained popularity as a consumer entertainment technology in recent years, many games had been published for this medium. Since the contemporary game production process is based on the Integrated Development Environment (IDE) applications, it is easier for developers to create multiple versions of their game for different platforms. Online game distribution repositories such as steampowered.com provide many games that can be played either on desktop computers with 2D screens or with 3D stereoscopic head-mounted displays. This provided a great opportunity for researchers to conduct comparative studies to explore the user experience of the relatively novel virtual reality applications. However, the effect of VR use is not only an issue in gaming. Effect of 3D stereoscopic HMDs and 2D screens was compared in different disciplines for different contexts in several recent studies, such as sports spectatorship (D. Kim & Ko, 2019), for their potential of inducing awe through watching videos (Chirico et al., 2017), the feasibility of memory assessment (Ventura, Brivio, Riva, & Baños, 2019), remote collaboration (Anton, Kurillo, & Bajcsy, 2018) or safety training (Buttussi & Chittaro, 2018). These studies and their predecessors (see Buttussi & Chittaro, 2018 for a detailed review) mainly focus on performance measures with a little bit of interest in user experience. On the other hand, there is a long tradition of comparing 2D screens to 3D immersive technologies in studies that aim to measure presence (Berkman & Akan, 2019).

Compared to other contexts, HMD vs. 2D monitor examinations are relatively less abundant for gaming, which limits our knowledge of “differences between players’ experience in video games played in immersive modalities and in games played in non-immersive modalities (Pallavicini, Pepe, & Minissi, 2019). In addition, gaming is a broad context that offers different gaming experiences within different game genres. Although VR systems promise to offer a more immersive and enhanced gaming experience, this should be explored for different types of games.

In this study, we evaluated the puzzle game “Keep Talking and Nobody Explodes” (Pestaluky, Kane, & Fetter, 2015) through a within-subjects experiment design with 21 participants, for desktop monitor and HMD conditions.

In addition to objective measures of game success and gameplay duration, subjective measures of game user experience was measured via the GUESS (Game User Experience Satisfaction Scale) (Phan, Keebler, & Chaparro, 2016). MEC-SPQ (Measurement Effects Conditions - Spatial Presence Questionnaire) (Vorderer et al., 2004) was used to compare the sense of presence as an important measure of VR UX. Participants were also queried for their subjective understanding of the task difficulty, using the NASA-TLX (NASA Task Load Index) scale (Hart & Staveland, 1988).

Although there is a relatively rich background regarding the comparison of highly immersive and less immersive technologies for interacting with the virtual environments, the majority of the studies were not executed within the perspective of Game User Research. This study is an

addition to several studies in recent years (Carroll, Osborne, & Yildirim, 2019; Shelstad, Smith, & Chaparro, 2017; C. Yildirim, Carroll, Hufnal, Johnson, & Pericles, 2018), which are trying to systematically compare the game user experience through different immersive technologies. In order to achieve this, we employed state-of-the-art multidimensional subjective measurement tools, such as GUESS and MEC-SPQ, as well as a clear definition of the gameplay and gaming environment. Also, we provided NASA-TLX as a reliable measure of task load, since games in different genres demand different types of abilities and effort. We think that using a similar approach may benefit game user research studies, as the “game genre” or “game difficulty” is not a clear definition of game attributes in many cases, but comparing games that have corresponding temporal, physical or mental demands as well as similar self-assessments of effort, performance and frustration would be helpful to identify the effect of immersive technologies on different games.

RELATED STUDIES

Comparing games played in immersive and less-immersive modalities

Plenty of studies compare the performance of users exploring a virtual environment in immersive VR and less-immersive desktop systems. Some of those studies explore CAVE (Cave Automatic Virtual Environment – a recursive acronym) systems, as one of the earliest that provided evidence for the superiority of virtual environment over a workstation in a structure detection task (Arns, Cook, & Cruz-Neira, 1999), while the user performance was better in the interaction task for the workstation. Elmqvist, Tudoreanu & Tsigas (2008) revealed that users perform better in free navigation on CAVE but constrained navigation leads to a better performance on desktop computers. Bacim et al. (2013) noticed that stereoscopic multidimensional head-tracked CAVE display enhances the user performance even in complex 3D graph investigation tasks, compared to a non-stereoscopic single side screen. Comparison of the CAVE, HMD, and desktop displays in a Stroop task lead to inconsistent findings in terms of user performance, since the CAVE and desktop display lead to an insignificant difference in time on task metrics for low-stress condition, while the task took significantly longer for the users with HMD. On the other hand, there was a significant difference between desktop and CAVE, that task duration was longer for desktop users. However, task accuracy was not affected by display systems (K. Kim, Rosenthal, Zielinski, & Brady, 2014). The study also explored the emotional responses of the participants via self-report measures as well as biometric measures of skin conductance response. Self-reported arousal and valence were significantly higher on CAVE and HMD displays, as well as the skin conductance response. CAVE experience induced the highest sense of presence as the desktop induced the lowest, and differences were significant between all three conditions. Laha, Sensharma, Schiffbauer & Bowman (2012) and (Ragan, Kopper, Schuchardt, & Bowman, 2013) concluded that with higher immersive qualities such as larger field of regard, stereoscopy, and head tracking, users perform better in visual analysis tasks with volume data and small-scale spatial judgments, as the task complexity increases.

Along with the aforementioned CAVE studies that evaluate task performances, some studies evaluated gaming applications in CAVEs. McMahan, Bowman, Zielinski, & Brady (2012) compared the game performance metrics in a first-person shooter game such as task

completion time, damage taken, accuracy, headshot count along with the subjective measures of presence, engagement, and usability. Users performed better with a familiar low-fidelity mouse-keyboard input device interaction on the desktop computer condition, while the high fidelity human-joystick and magic wand interaction that resembles the real world lead to better performance measures in CAVE. Both the high display fidelity and high controller fidelity affected the subjective measures positively.

Supporting the findings of McMahan et al. (2012), Lugin et al. (2013) reported higher user preference a first-person shooter game played on a CAVE, compared to desktop gameplay. On the other hand, their study report significantly lower game performance for CAVE gameplay, which could have been caused by technical limitations of wand button response time which is inferior to mouse configuration in desktop gameplay.

Comparative studies on 2D screens and HMD based VR has a slightly longer history. Pausch, Proffitt, & Williams (1997) explored the task performance in a camouflaged target seek task and denoted that users with HMD performed faster than users with desktop monitors, although there is not a significant difference in the number of targets determined. Waller, Hunt, & Knapp (1998) compared the task time and performance in a real-world maze and did not detect a difference between the participants who were trained with a 6-DOF head-tracked HMD or a desktop monitor. However, this study evaluates training outcomes of different display systems, rather than the VE experience. Ruddle, Payne & Jones (1999) verified that time on task was better for HMD users, while distance traveled is further than desktop users in indoor navigation tasks, as the HMD users did not stop looking around. In addition, HMD users provide a more accurate relative straight-line distance. However, in another study (Ruddle & Péruch, 2004), they had contrary results in favor of desktop users' distance estimation but noted that neither desktop nor HMD users' estimations were correct, while distances traveled in the maze were also equal. Patrick et al. (2000) did not detect any differences in spatial knowledge learned for a virtual environment using a desktop display, a large screen, and an HMD. To assess the different technologies in educational settings, information recall, spatial recall and presence were evaluated after a 15-minute seminar in four different conditions including desktop monitors and HMD without head-tracking (Mania & Chalmers, 2001). The differences were not significant between these two technological conditions. Winn, Windschitl, Fruland & Lee (2002) reported a higher presence in HMD compared to desktop monitors as a predictor of knowledge gain. (Zanbaka, Lok, Babu, Ulinski & Hodges (2005) denoted that presence is lower for the desktop system compared to HMDs with different controller settings for locomotion. Mizell, Jones, Slater & Spanlang (2002) concluded that immersive virtual reality (IVR) technology gives users a measurable advantage over more conventional display methods when visualizing complex 3D geometry, through a task that users recreated the abstract sculptures in the real world, which they had viewed virtually. (Qi, Taylor, Healey & Martens (2006) found that users equipped with a 2D display fish-tank like system perform better in volume visualization task, in terms of accuracy at judging the shape, density, and connectivity of objects and completed the tasks significantly faster than the HMD group, as HMD group had an inside-out view of the volume while fish-tank systems provide an outside view. Astronauts in training performed similarly in HMD

and desktop simulators of space station evacuation for 3D navigation performance measures but performed better with HMD for pointing tasks (Aoki, Oman, Buckland, & Natapoff, 2008). The healthcare workers trained in communication skills rated a higher level of empathy in HMD situations although the observing experts graded their performance higher for large screen projection conditions (Johnsen & Lok, 2008). Navigating through a virtual maze, both user ratings and EEG data indicated a higher level of presence with a single stereo projection display, compared to a desktop display (Kober, Kurzmann, & Neuper, 2012). Li & Giudice (2013) demonstrated that user performance in desktop systems were not significantly different from HMD systems in a task that users point at and navigate to targets in a multi-level virtual building. Through a set of experiments, Slobounov, Ray, Johnson, Slobounov & Newell (2015) presented that higher self-reported presence for 3D screens is also supported by electroencephalographic biometrics. Weidner, Hoesch, Poeschl, & Broll (2017) compared 2D, 3D stereoscopic, and HMD displays in a driving simulation lane change task and they did not detect any significant differences in physiological responses or driving performance. For watching sports, HMD use “amplified flow experience via vividness, interactivity, and telepresence to the greater extent”, compared to a 2D screen, and leads to a substantially enhanced user satisfaction. (D. Kim & Ko, 2019). Comparing the experience of watching videos on a HMDs vs. on desktop displays, Chirico, Ferrise, Cordella & Gaggioli (2018) concluded “that immersive videos significantly enhanced the self-reported intensity of awe as well as the sense of presence, while awe content on VR also leads to a higher parasympathetic activation. (Wilson & Mayhorn, 2019) reported their preliminary results that 3D sports videos provide a significantly higher level of spatial presence while “engagement and enjoyment as well are trending towards significance”, although cybersickness related symptoms were scored significantly higher than desktop monitor condition.

For a serious game for aircraft cabin safety training, two HMD systems and a desktop system were compared in terms of knowledge gain, and self-reported measures of self-efficacy, engagement and presence (Buttussi & Chittaro, 2018). Although engagement and presence were higher for HMD VR training, knowledge acquisition and self-efficacy were not affected.

Numerous studies evaluate display systems within the gaming context. Slater, Linakis, Usoh & Kooper, (1995) revealed that players performed better in a three-dimensional chess game when they have an egocentric view through an HMD, compared to users who had an exocentric view through a TV display. Sousa Santos et al. (2009) verified that users performed better desktop conditions in terms of the number of collected objects, the number of collisions with walls, walked the distance, average speed, and total gaming time while playing a first-person maze exploring game, compared to an HMD VR condition. Comparing a display's 2D and 3D modes, Litwiller & LaViola (2011) did not detect a significant difference in-game performance metrics within five video games, but they reported a higher user preference of 3D gaming. In a similar study, Kulshreshth, Schild & LaViola (2012) reported: “a positive effect on gaming performance based on stereoscopic vision, although reserved to isolated tasks and depending on game expertise”. Based on the user interviews and biometric data, Tan, Leong, Shen, Dubravs & Si (2015) provide qualitative evidence that HMD users had a higher level of immersion and flow in a first-person shooter game despite the cybersickness. Martel et al. (2015) also reported a higher level of immersion based on user ratings, for different HMD

based interaction schemes for a first-person shooter gameplay compared to a desktop monitor. Shelstad, Smith & Chaparro (2017) revealed that HMD based VR enhanced the game user experience in a strategy game through the dimensions of satisfaction, enjoyment, engrossment, creativity, sound, and graphics quality. Pallavicini et al. (2018) evaluated a first-person platform game comparing the tablet and HMD gameplay. They did not encounter any significant difference neither on game performance measures of task time and progress level nor the self-report measure of system usability and heart rate as a biometric indicator of arousal. The only significant difference was on user preference evaluated as Net Promoter Score. Christensen, Mathiesen, Poulsen, Ustrup & Kraus (2018) explored a competitive multiplayer game for desktop, HMD with joystick and HMD with magic wands conditions and exposed that HMD versions provide significantly higher scores for many of the measures in a multidimensional game user experience questionnaire, such as immersion, flow, positive affect, empathy, and behavioral involvement. Monteiro et al. (2018) evaluated a car racing game for 3rd person view on desktop versus first-person view with HMD and reported a slightly higher level of immersion for HMD gameplay. Roettl & Terlutter (2018) showed that their jump-and-run action game yielded a higher sense of presence when played in the HMD VR than in the stereoscopic 3D than in the 2D video game, but they did not explore any effect on arousal or attitude towards the video game. A recent study (C. Yildirim et al., 2018) which compared the game user experience and presence in two HMD systems and a desktop computer conditions for a first-person shooter game, and did not reveal any significant differences. Making another comparison using a strategy and a racing game (Carroll et al., 2019), authors did not observe any significant difference in player experience metrics except presence, which was higher for HMD VR condition.

Measuring UX in Comparison of Displays Systems for Virtual Environments and Games

The aforementioned studies employ several measures to compare immersive and less-immersive technologies from the perspective of user research, which can be dissected as performance measures, subjective measures, and biometric measures.

Two of the frequently employed performance measures are time on task and task success rate (e.g. in (Bacim et al., 2013; K. Kim et al., 2014; Mizell et al., 2002; Pausch et al., 1997; Ragan et al., 2013; Swindells et al., 2004). Time on task is an indicator of efficiency while the success rate is pointing out the effectiveness of a system, which are common objective quantitative measures in user research (Sauro & Lewis, 2012:13-14). Although the task success is usually taken as a dichotomy, some variables such as “number of collected objects” or “number of collisions with walls” are employed to quantify effectiveness as success rate or error rate while variables such as “walked distance” or “average speed” stand as measures for efficiency (e.g. Sousa Santos et al., 2009). Furthermore, response time is evaluated (Qi et al., 2006). Likewise, in-game achievements such as gained points or failures such as damage taken had also been employed (Litwiller & LaViola, 2011; Lugin et al., 2013; McMahan et al., 2012). In order to assess perceived user performance, standardized questionnaires were employed, as usual in user research studies (Sauro & Lewis, 2012:185). Some of these questionnaires query the emotional responses of users (Johnsen & Lok, 2008; K. Kim et al., 2014; Pallavicini et al., 2019). In some cases, researchers also used ad hoc surveys (Chirico et al., 2018; Kulshreshth et al.,

2012; Litwiller & LaViola, 2011; Slater et al., 1995; Sousa Santos et al., 2009). Within the context of the study, some researchers (Pallavicini et al., 2019) employed standardized usability questionnaires such as System Usability Scale (Brooke, 1996) or some preferred (Anton et al., 2018) User Experience Questionnaire (Laugwitz, Held, & Schrepp, 2008), while NASA Task Load Index (Hart, 2006; Hart & Staveland, 1988) was used to obtain workload estimates (Anton et al., 2018; Berkman, Bostan, & Yalçın, 2020). Experiences related to the flow state were evaluated via several adopted measures (D. Kim & Ko, 2019; Mania & Chalmers, 2002; Roettl & Terlutter, 2018; Slobounov et al., 2015; Winn et al., 2002). In addition, standardized presence questionnaires were used such as Slater-Usuh-Steed Presence Questionnaire (Usuh, Catena, Arman, & Slater, 2000) in several studies (Johnsen & Lok, 2008; Kober et al., 2012; McMahan et al., 2012; Pallavicini et al., 2019; Zambaka et al., 2005), Witmer & Singer (1998) Presence Questionnaire in some studies (Carroll et al., 2019; K. Kim et al., 2014; Kober et al., 2012), ITC Sense of Presence Inventory (Lessiter, Freeman, Keogh, & Davidoff, 2001) in (Lugrin et al., 2013; Wilson & Mayhorn, 2019). In order to assess game user experience with a standardized scale, Game Experience Questionnaire (IJsselsteijn, de Kort, & Poels, 2007) was employed in (Christensen et al., 2018; McMahan et al., 2012; Monteiro et al., 2018) and Game User Experience Satisfaction Scale (Phan et al., 2016) was used in (Carroll et al., 2019; Shelstad et al., 2017; C. Yildirim et al., 2018). In order to evaluate cybersickness, which is a detrimental negative effect of VR systems based on visually induced motion sickness, was assessed using Simulator Sickness Questionnaire (Kennedy, Lane, Berbaum, & Lilienthal, 1993) in many of the studies (Christensen et al., 2018; K. Kim et al., 2014; Mania & Chalmers, 2001; Monteiro et al., 2018; Zambaka et al., 2005).

Biometric measures are employed for exploring emotional responses of users triggered by interactions with the virtual environments. Cardiovascular biometric measures such as heart rate (Pallavicini et al., 2018, 2019; Swindells et al., 2004; Weidner et al., 2017) were used as an indicator of arousal or attention, as well as electro-dermal activity measures (K. Kim et al., 2014; Pallavicini et al., 2019; Tan et al., 2015; Weidner et al., 2017) to determine arousal. EEG data was used (Kober et al., 2012; Slobounov et al., 2015) to explore different consciousness states, as being in a relaxed wakefulness state or negative or positive valence of different emotional states such as fear or happiness. Facial electromyography techniques were used (Weidner et al., 2017) to detect facial expressions that reflect several emotional states.

METHODOLOGY

Participants

The experiment involved 21 participants, 13 males and 8 females, with ages ranging from 18 to 27 ($M=22$, SD , 2.05). Participants have a daily average weighted gameplay score of 2.15 ($SD=.935$) for the last 3 months, based on the Lifetime Television Exposure Scale (Riddle, 2010) modified for reporting gameplay activities instead of TV viewing. The distribution of participant scores are depicted in Figure 1 Distribution of Daily Average Gameplay. The scale provides scores between 1 to 7, based on the participants answers on their gaming habits within the last three months on weekdays with a weight of 71.4 and weekends with a weight of 28.6. Participants have previously used a VR HMD before but none of them played the game “Keep Talking and Nobody Explodes” (Pestaluky et al., 2015) on VR or any other medium.

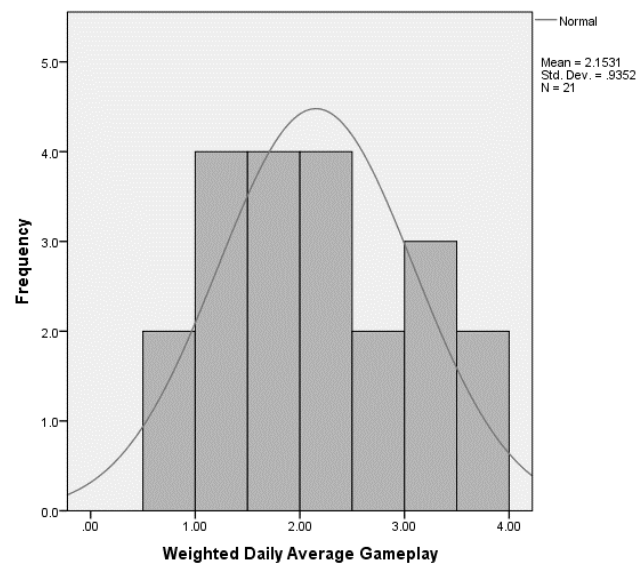


Figure 1 Distribution of Daily Average Gameplay Score

Equipment and Stimulus

Participants used the same laptop computer for gaming. A 27-inch display was attached to the computer, with a standard keyboard and mouse for desktop 2D display condition. An Oculus Rift head-mounted display with a pair of Oculus Touch magic wand controllers were used for stereoscopic head-tracked virtual reality condition. In desktop condition, all the interactions were made using the mouse, while magic wands are the only method of input for VR condition.

Participants played the first level of the game in which players try to de-activate a virtual bomb by following the instructions of another person, the moderator; who does not see the virtual bomb but reads the bomb-defusing instructions to players while players describe what they see on the bomb. To standardize the gameplay conditions, the same person who had prior experience with the game had acted as the moderator.

Since the puzzles in the game are procedurally generated, the level contains different puzzle settings each time the game is launched, but the puzzle complexity remains the same. As seen on the Figure 2, the level contains three modules; battery, wire and cymbal pad, along with a timer that indicates remaining time to explosion, counting down from 5 minutes. On one side of the bomb casing, there is also a battery hole that can be used to determine the “bomb type”. The puzzle was given to users changes randomly, by different settings for the color or number of wires, symbols on the keys and key sequences, activation keys and battery types. Players are free to start from any of those components.

Moderators ask players about what they see on the bomb. Based on their answer, the moderator reads the instructions to the player. For example, if the player says “I see wires”,

the moderator asks the player about the number of wires, and reads the instructions based on the number of wires, using the manual available online at the game's website.

The communication between the moderator and player was established through a VoIP application with a pair of mobile phones. Players wore a pair of earphones for VoIP communication, which is covered by a headphone set for game audio.



Figure 2 The "virtual bomb" in "Keep Talking Nobody Explodes" game. Screen capture (Steel Crate Games, 2015)

Measures

As objective metrics of user performance, the success in the gameplay was recorded as successful when the user deactivated the bomb before the timer reaches zero. Gameplay duration recorded based on the timer on the "virtual bomb".

In order to assess the presence, we preferred to use MEC-SPQ (Measurements, Effects, Conditions- Spatial Presence Questionnaire)(Vorderer et al., 2004; Wirth, Vorderer, Hartmann, Klimmt, & Schramm, 2003), as previous studies demonstrated that it is more sensitive for detecting differences between different VEs (Ç. Yildirim, Bostan, & Berkman, 2019). Participants respond to MEC-SPQ items through a Likert scale, ranging from 1 (I do not agree at all) to 5 (I fully agree). In addition to process subscales of Attention Allocation and Spatial Situation Model, the spatial presence module has two subscales that refer to Self-location and Possible Actions, a.k.a. SPES (Spatial Presence Experience Scale) (Hartmann et al., 2015). The Higher Cognitive Involvement and Suspension of Disbelief subscales refer to higher state actions indicating absorption, while Domain-Specific Interest and Visual-Spatial Imagery subscales are trait-like enduring user-related variables. Each of these subscales can be applied using 4-item, 6-item or 8-item versions. The trait-like latent variables of Domain-Specific Interest and Visual-Spatial Imagery were excluded from our study regarding the within-subjects experimental design. Since these variables are enduring traits of participants rather than being related to the singular experience evaluated.

GUESS (Game User Experience Satisfaction Scale) was preferred to assess gaming experience since it has a clear multidimensional structure that had been psychometrically evaluated (Phan et al., 2016). GUESS has 8 dimensions which are Social Connectivity, Narratives, Usability/Playability, Play Engrossment, Creative Freedom, Personal Gratification, Audio Aesthetics and Visual Aesthetics. As suggested by its developers, we did not employ the Narratives dimension because there is not a background story related with the bomb situation in the game. Besides, we did not collect data using Social Connectivity dimension, since the social connection is not established through a channel that is a part of the game application.

NASA-TLX (National Aeronautics and Space Administration – Task Load Index) (Hart, 2006; Hart & Staveland, 1988) is used to measure the workload of a task, perceived subjectively, by whom executed the task. It involves a total of six subscales, with items evaluated on a 20-point horizontal line scale and originally scored through a two-step process, as a weighted sum of paired comparisons of the six dimensions. After the participants evaluate the task load for their Performance on the task, Physical Demand, Mental Demand and Temporal Demand of the task, as well as Frustration they might have experienced, and the Effort they spent; they compare dimensions pairwise and rate their importance with tallies between 0 to 5. Based on these tallies, a weighted score is calculated for each dimension. Although some researchers (Moroney, Biers, Eggemeier, & Mitchell, 2003) use unweighted NASA-TLX, we employed the tallies method in our study to achieve higher sensitivity. We think that, NASA-TLX is a very suitable tool for game studies, as it provides a weighted measure for different task difficulty measures, because games with different gameplay attributes have different requirements. While some games such as first-person shooter games depend on agility which is a temporal aspect, others such as strategy and defense games seem to depend on mental abilities.

Since there is strong evidence on many previous studies that use of HMD systems lead to a significantly higher level of cybersickness, we kept this issue out of the scope of our study.

Procedure

In order to minimize the effect of prior exposure to the game, half of the participants played the game using VR setup first. The other half started with playing the desktop version. Each player was left alone in the room once the researchers prepared the equipment and game settings. After each gameplay session, participants responded to items of GUESS' selected dimensions, MEC-SPQ, and NASA-TLX with tally comparison, excluding the two enduring trait-like dimensions of MEC-SPQ on the second session. Participants were also asked to indicate their success status and gameplay duration within the questionnaire, which was recorded by experimenters and given to them after they completed the experiment.

RESULTS AND DISCUSSION

A series of paired samples t-tests were conducted to explore and compare the game user experience measures acquired in desktop 2D monitor and HMD based stereoscopic VR conditions.

Gameplay Duration and Success Rate

A paired samples t-test revealed that there is not a significant difference between the gameplay duration on HMD version ($M = 3:35$, $SD = 1:19$) and 2D display version ($M = 3:29.5$, $SD = 0:59.5$); $t(20) = .345$, $p = .733$. Duration is given as minutes and seconds, depicted on Figure 3 (left).

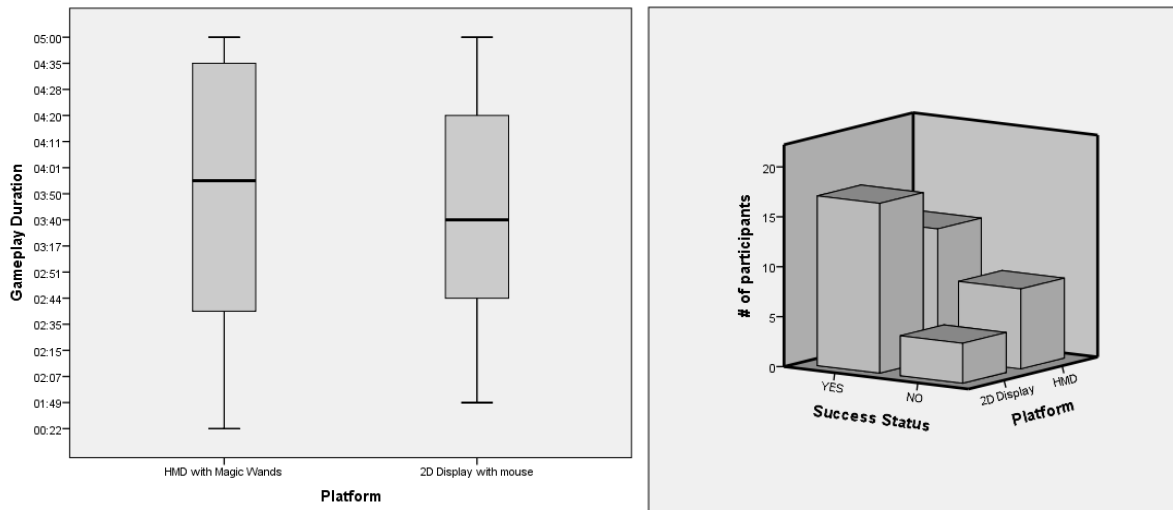


Figure 3 Gameplay duration (left) and success rate (right)

There is not a significant difference in the completion of the gameplay in mission in the HMD version ($M = .62$, $SD = .498$) and the 2D display version ($M = .80$, $SD = .40$); $t(20) = -1.45$, $p = .162$. This parameter was about whether the participants were able to complete the mission (defusing the bomb) or not. Success rate is indicated as percentages. Number of participants according to their success status can be seen on Figure 3 (right).

Considering that gameplay duration is limited to 5 minutes based on the game mechanics that simulates a “live bomb deactivation” participants reached the end of the game around three and a half minutes in both display and controller conditions. When the users interact with the VE with a well-paired controller and display configuration, their efficiency is not affected. Even there is a slight difference between the number of users who successfully “deactivated the bomb” in different conditions (Figure 3 right), it is not a significant difference, as it had also been observed in previous studies. As given in previous studies (Aoki et al., 2008; Kulshreshth et al., 2012; Li & Giudice, 2013; Litwiller & LaViola, 2011; McMahan et al., 2012; Pallavicini et al., 2018; Pausch et al., 1997; Waller et al., 1998), user performance does not dramatically change when they used a mouse or keyboard with a 2D desktop display and magic wands with a head-tracked stereoscopic display, although users usually perform slightly better in 2D display condition. An exception is the results obtained by Sousa Santos et al. (2009) for a maze task, which are not consistent with our findings. However, they did not clearly report the controllers used in HMD and 2D display conditions. This issue can be explained based on users’ prior experiences with 2D displays and relevant controllers. Regardless of the task given, familiarity with output and input devices might be advantageous, although the head-tracked HMD provides more real-world like interactions compared to desktop. On the other hand, Berkman et al. (2020) reported that using either a standard game controller or a pair of magic wand controller along with a HMD did not lead

to a significant difference in difference in subjective measures task load except the unweighted NASA-TLX Effort score, but the difference was on subjective evaluation of usability.

Considering that gameplay duration is limited to 5 minutes based on the game mechanics that simulates a “live bomb deactivation” participants reached the end of the game around three and a half minutes in both display and controller conditions. When the users interact with the VE with a well-paired controller and display configuration, their efficiency is not affected. Even there is a slight difference between the number of users who successfully “deactivated the bomb” in different conditions (Figure 3 right), it is not a significant difference, as it had also been observed in previous studies. As given in previous studies (Aoki et al., 2008; Kulshreshth et al., 2012; Li & Giudice, 2013; Litwiller & LaViola, 2011; McMahan et al., 2012; Pallavicini et al., 2018; Pausch et al., 1997; Waller et al., 1998), user performance does not dramatically change when they used a mouse or keyboard with a 2D desktop display and magic wands with a head-tracked stereoscopic display, although users usually perform slightly better in 2D display condition. An exception is the results obtained by Sousa Santos et al. (2009) for a maze task, which are not consistent with our findings. However, they did not clearly report the controllers used in HMD and 2D display conditions. This issue can be explained based on users’ prior experiences with 2D displays and relevant controllers. Regardless of the task given, familiarity with output and input devices might be advantageous, although the head-tracked HMD provides more real-world like interactions compared to desktop. On the other hand, Berkman et al. (2020) reported that using either a standard game controller or a pair of magic wand controller along with a HMD did not lead to a significant difference in difference in subjective measures task load except the unweighted NASA-TLX Effort score, but the difference was on subjective evaluation of usability.

Game User Experience Satisfaction Scale

None of the mean scores for dimensions of GUESS yielded a significant difference between the HMD gameplay and 2D display gameplay. It should be noted again that we excluded the Narratives dimension and the Social Connectivity dimension of GUESS in our study. The mean scores, standard deviations, mean differences between HMD and 2D display conditions, and t-test parameters for each dimension of GUESS can be followed on Table 1 GUESS scores.

Table 1 GUESS scores. None significant at $p < .05$.

		M	SD	M Diff	Paired T-Test
Usability/Playability	HMD	5.60	.83	-.22	t(20)=-1.66; p=.12
	2D display	5.81	.78		
Play Engrossment	HMD	5.50	.92	.25	t(20)=1.27; p=.22
	2D display	5.25	.98		
Enjoyment	HMD	5.30	.65	.32	t(20)=1.89; p=.08
	2D display	4.99	.88		
Creative Freedom	HMD	5.23	1.17	.40	t(20)=1.68; p=.11
	2D display	4.83	1.36		

Audio Aesthetics	HMD	4.94	1.47	.35	t(20)=1.27; p=.22
	2D display	4.58	1.55		
Visual Aesthetics	HMD	6.13	.78	.22	t(20)=1.23; p=.24
	2D display	5.90	.74		
Personal Gratification	HMD	5.83	.82	.22	t(20)=1.51; p=.15
	2D display	5.61	.99		

Although several studies report user preference for HMD experience compared to 2D display experience (Buttussi & Chittaro, 2018; Litwiller & LaViola, 2011; Martel et al., 2015; Pallavicini et al., 2018; Tan et al., 2015), a limited number of studies that explore the detailed dimensions of gaming experience became available in recent years. Of those who employ GEQ as a multidimensional measure, dimension scores are reported in detail by Christensen et al. (2018) only. Although GEQ has different dimensions, some of them are conceptually comparable to GUESS dimensions. The GUESS Play Engrossment dimension consists of items similar to Flow in GEQ. While Christensen et al. (2018) report that their HMD based VR condition yielded a significantly different score on Flow, we did not detect a similar effect on Play Engrossment, which is a conceptually similar latent variable. On the other hand, there is not a significant effect of display condition neither on GEQ Competence nor on GUESS Personal Gratification. Another pair of conceptually similar dimensions is the Enjoyment of GUESS and Positive Effects of GEQ. However, our results are not consistent with Christensen et al. (2018) who report a significantly higher Positive Affects score for HMD condition. On the other hand, the gaming context seems to be different in their study. Their multiplayer gameplay requires coordination of two players through voice chat, while pulling a series of levers in order to create a specific type of boulder. In this case, the magic wand controlled HMD version could be providing better controls, which lets the user to reach to a higher level of flow and enjoyment. Unfortunately, authors did not report any performance measures in their study, and GEQ does not have a dimension comparable to Usability/Playability dimension of GUESS.

When we compare our results on GUESS with (C. Yildirim et al., 2018), our findings are consistent since they also could not detect any significant difference on any dimensions, with an exception. For Usability/Playability dimension, the FPS game evaluated in their study yielded a significantly higher score for desktop condition, compared to the two different HMD with magic wands conditions. Contrary to our findings, they obtained slightly higher scores for non-immersive desktop conditions on other dimensions. We suggest that their participants did not prefer to play an FPS style game on a HMD due to their prior experiences with similar games, although the authors proposed that “FPS games are in themselves immersive to the extent that playing them on a VR headset and desktop computer are comparably enjoyable.” Unfortunately, the authors did not report any information on their participants’ prior gaming experience and acquaintance with FPS games.

Meanwhile, a previous study (Shelstad et al., 2017) across display conditions reported significantly “higher perceptions of engrossment, enjoyment, creative freedom, audio aesthetic, and visual aesthetic” for strategic gameplay using an HMD. As the authors suggested, future studies as (C. Yildirim et al., 2018) and this study exposed that games in different genres resulted in different experiences through VR, compared to 2D displays.

A recent study (Carroll et al., 2019) that compares two games of different genres using GUESS report that “regardless of the game genre, participants in the VR gaming condition experienced a greater level of sense of presence than did those in the desktop gaming condition”. Since the authors only reported an overall player experience score via GUESS, it is not possible to discuss their results further.

Presence measured via MEC-SPQ

Results show that the mean score for MEC-SPQ dimension of Spatial Presence: Possible Actions is significantly higher for HMD (M = 4.066, SD = .77) compared to 2D display gameplay (M = 3.43, SD = .72); $t(20) = 3.208, p = .004$.

The other Spatial Presence related dimension of MEC-SPQ in which a significant difference ($t(20) = 3.77, p = .001$) between HMD (M = 4.30, SD = .68) and 2D display (M = 3.58, SD = .87) observed is Self Location.

The other dimensions employed in the study did not reveal any significant differences. Mean values and differences can be observed on Table 2 MEC-SPQ scores along with the standard deviations and t-test results.

Table 2 MEC-SPQ scores. * significant at $p > .05$

		M	SD	M Diff	Paired T-Test
Attention Allocation	HMD	4.35	.64	.19	$t(20)=-1.83; p=.082$
	2D display	4.15	.66		
Spatial Situation Model	HMD	4.05	.74	.19	$t(20)=1.51; p=.145$
	2D display	3.86	.70		
Higher Cognitive Involvement	HMD	3.61	.70	.10	$t(20)=.792; p=.438$
	2D display	3.51	.71		
Suspension of Disbelief	HMD	2.94	.77	.06	$t(20)=0.43; p=.68$
	2D display	2.88	.68		
Spatial Presence: Possible Actions*	HMD	4.07	.77	.64	$t(20)=3.21; p=.01$
	2D display	3.43	.72		
Spatial Presence: Self Location*	HMD	4.30	.68	.71	$t(20)=3.78; p=.01$
	2D display	3.59	.87		

As reported in many studies, HMD use leads to a higher sense of presence, spatially. On the other hand, other evaluated dimensions of presence were not significantly affected. The Attention Allocation dimension scores indicate that players can focus their perception at an almost identical level. Their mental model of the virtual environment is not highly affected by a higher level of immersion obtained through the head-tracked stereographic display, as indicated by variables of the Spatial Situation model. Similarly, dimensions related to Absorption did not reveal any significant effect. Based on the Suspension of Disbelief

dimension mean scores, it is possible to claim that both platforms provided plausible experiences for the users without any inconsistencies or errors, while their imagination and thoughts were triggered at a corresponding level by both presentations.

It is not possible to discuss the effect of VR in any further detail since there are plenty of subjective methods for assessing presence. However, previous comparative studies did not employ MEC-SPQ or SPES. Furthermore, they did not report the detailed results for subscales of the multidimensional measures. Although the other questionnaires such as ITC-SOPI (Lessiter et al., 2001) that were employed to assess the difference between the HMD and 2D displays (Wilson & Mayhorn, 2019) combining the measures of spatial presence in a single dimensions, results are consistent with our findings on Possible Actions and Self-location measures.

The Possible Actions score indicates that the players formed an impression as it was possible to be more active in the HMD condition, although both VEs provide the same actions that can be executed on virtual objects, such as cutting the wires, pressing the keys and rating the bomb. We suppose that this effect is partially due to head-tracked stereophonic vision, but it mainly depends on the two hand magic wand based control activities, as findings of another study that comparing a standard game controller and magic wands in HMD condition revealed a similar result (Aksayim & Berkman, 2020). On the other hand, previous work (Schild, LaViola, & Masuch, 2012) revealed that stereoscopic vision in gameplay also leads to higher sense of Self-location and Possible Actions measured via MEC-SPQ. Thus it is not possible to claim a direct relationship between Possible Actions and controllers or Self-location and stereoscopic vision. These two components of spatial presence are complementary to each other and further examination is required in order to understand the effect of technology on each of them.

Task Load via NASA - TLX

Our results revealed that there is a significant difference observed between the Physical Demand scores for the desktop 2D display (M = 32.57, SD = 26.24) and the HMD gameplay version (M = 19.71, SD = 23.95) ; $t(20)=2.24$, $p=.04$.

The Frustration subscale of the NASA-TLX also showed a significant difference ($t(20) = -2.56$, $p = .02$) between the HMD (M = 19.81, SD = 21.98) and 2D display (M = 32.05, SD = 26.19) conditions.

As shown in Table 3 Weighted NASA-TLX scores, other dimensions did not reveal significant differences between the mean scores of experimental conditions, as well as the overall mean score.

*Table 3 Weighted NASA-TLX scores. * significant at $p>.05$*

		M	SD	M Diff	Paired T-Test
Overall Mean	HMD	13.06	3.40	1.15	$t(20)=1.85$; $p=0.08$
	2D display	11.91	3.22		
Mental Demand	HMD	40.29	27.20	3.76	$t(20)=0.55$; $p=0.59$
	2D display	36.52	26.57		

Temporal Demand	HMD	33.10	20.74	7.90	t(20)=1.72; p=.11
	2D display	25.19	17.31		
Performance	HMD	38.14	26.56	-2.43	t(20)=-0.29; p=.78
	2D display	40.57	30.35		
Effort	HMD	31.95	24.57	7.33	t(20)=1.2; p=.25
	2D display	24.62	18.94		
Frustration*	HMD	19.81	21.98	-12.24	t(20)=-2.56; p=.02
	2D display	32.05	26.19		
Physical Demand*	HMD	32.57	26.24	12.86	t(20)=2.24; p=.04
	2D display	19.71	23.95		

As it can be followed on Table 3 Weighted NASA-TLX scores. * significant at $p > .05$, participants provided almost identical scores for Mental Demand in HMD since the players struggled with the same kind of puzzle-solving task in both conditions. The self-assessment based Temporal Demand scores are also in line with the objective time on task measures, which is slightly higher for the HMD version compared to the 2D display version. Players evaluated their own performance as almost identically between conditions, without any significant difference, although the game success rate is a little bit better for desktop 2D gameplay. Although there is a slightly higher score of Effort for HMD use is observed, the difference is not significant.

However, NASA-TLX results revealed that users felt more frustrated through the desktop 2D display condition, although their success rate is slightly higher for the desktop condition. When we inspect the effect of success in the game on NASA-TLX Frustration score through a between-subjects comparison, we observed a significant effect in desktop condition ($t(19)=2.66$; $p < .05$). However, such effect was not observed in HMD condition ($t(19)=-.45$; $p > .05$). Participants who finished the game in 2D screen condition had a lower score of Frustration ($N=17$; $M=25.6$; $SD=24.12$) compared who did not finish the game ($N=4$; $M=59.5$; $SD=15.17$), while the difference between successful participants ($N=13$; $M=17.92$; $SD=18.29$) and failing participants ($N=8$; $M=22.88$; $SD=26.93$) were much closer in HMD condition. We assume that participants had higher expectations of being successful when they were playing the desktop version, probably based on their previous experiences in desktop gaming.

As expected, the reason for higher Physical Demand score HMD with magic wands environment is the more real-world like movements of players as leaning backward and forwards in order to interact with the objects in the game world and use of both hands to manipulate game environment.

CONCLUSION

Our study provided results that contribute to our understanding of gaming across different immersive medium. Although it is previously given that different gameplays yield different

effects on player experiences, our study is revealed that a puzzle game is also affected by the immersive qualities of the gaming medium.

Highly immersive media increases the “sense of being there”, i.e. spatial presence, both in terms of self-location and perceived amount of possible interactions with the VE.

On the other hand, the differences in immersive aspects gaming platforms did not lead to differences in game user experience in the case of our puzzle game that depends on verbal communication and mental progress. Unlike the strategy and racing games (Carroll et al., 2019) or competent multiplayer game (Christensen et al., 2018), playing a puzzle game through the different amounts of immersion did not yield to significant differences in-game user experience or user performance.

For this reason, it is not possible to reach an absolute conclusion about the differences between immersive and less immersive platforms on game user experience since our results show that games in different “genres” may yield different effects. However, the game concept of “game genre” is also controversial, as video games should be classified not only by their narrative components but also according to interactive aspects (Clarke, Lee, & Clark, 2017), and there are several games that are developed as a hybrid of several genres.

Through a game design perspective, the player experience is a consequence of player actions which are result of the design choices of a game. Through a design perspective that evaluates a game as an artifact rather than media (Hunicke, Leblanc, & Zubek, 2004), a video game is composed of mechanics, dynamics, and aesthetics, the MDA framework. Mechanics are “various actions, behaviors, and control mechanisms afforded to the player. In our puzzle game, mechanics were actions of cutting the wires or pressing the keys or rotating the virtual objects, which are quite different from a racing game mechanics such as steering and speed control. On the other hand, the main mechanic in the evaluated puzzle game is “verbal communication with the moderator”, which is not applicable as a game genre. The game dynamics, which are mainly determining the winning conditions or the achievements in the gameplay are also different through the games. Our puzzle game employs time pressure and decision making as dynamic elements, while a racing game has the former, but a strategy game usually has the latter only. While the selection of these dynamics could affect the type of demand, as physical, temporal, or mental; the balance of dynamics may affect the amount of effort, frustration, and player performance. Finally, aesthetics is not limited to sense pleasures depending on artistic elements provided as audio or graphics, but also includes the dramatic components (narrative), pastime submission (play engrossment), fantasy elements, discovery and expression opportunities (creative freedom), challenge and social fellowship (personal gratification) that leads to enjoyment.

Considering the MDA Framework, the immersive technology is a modification in the game mechanics. Instead of using a 2D screen with a limited depth of sense, users have a stereoscopic HMD. Head tracking offers a more “natural” way to change the field of view, unlike the mouse-based rotational camera movements in many games. The two-handed magic wand interactions are easier to adopt compared to mouse clicks and key combinations on controllers

and keyboards. However, if the corresponding game mechanics are well mapped to these input and output methods, the interaction becomes seamless for the user, and players become engaged with the dynamics rather than mechanics. Thus, the main difference of gaming through a highly immersive technology mainly affects the perceived spatial presence, as players get more disconnected from their “real” environment and become surrounded by the virtual one, clearly locating themselves “there”. Besides the more direct mapping of virtual actions to more resembling real counterparts lead to a feeling that they can “act on the virtual world” more freely, compared to proxy-based direct manipulation technologies.

Although these conclusions are legit on the given evidence in our study and the prior work, we suppose that more comparative studies are required to understand the effect caused by the level of immersion on game user experience.

We suggest considering the MDA framework approach and the task load assessment for future studies. In addition to NASA-TLX (Hart & Staveland, 1988), the Player Experience Inventory (Abeele, Spiel, Nacke, Johnson, & Gerling, 2020), which is developed according to MDA framework and can be compared with GUESS. Unfortunately, this measure was unavailable while we collected the data used in this study. Besides, we suggest standardizing reporting of the former gaming habits of participants.

In order to provide an opportunity for a more detailed comparison of findings with future studies, we also decided to publish our dataset, as we are expecting the other researchers to share their data publicly.

REFERENCES

- Abeele, V. Vanden, Spiel, K., Nacke, L., Johnson, D., & Gerling, K. (2020). Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human Computer Studies*. <https://doi.org/10.1016/j.ijhcs.2019.102370>
- Aksayim, A., & Berkman, M. İ. (2020). Effect of Physical Activity on VR Experience: An Experimental Study. In S. Richir (Ed.), *Laval Virtual ConVRgence (VRIC) Virtual Reality International Conference - VRIC 2020*. Retrieved from <https://ijvr.eu/article/view/3316>
- Anton, D., Kurillo, G., & Bajcsy, R. (2018). User experience and interaction performance in 2D/3D telecollaboration. *Future Generation Computer Systems*. <https://doi.org/10.1016/j.future.2017.12.055>
- Aoki, H., Oman, C. M., Buckland, D. A., & Natapoff, A. (2008). Desktop-VR system for preflight 3D navigation training. *Acta Astronautica*. <https://doi.org/10.1016/j.actaastro.2007.11.001>
- Arns, L., Cook, D., & Cruz-Neira, C. (1999). Benefits of statistical visualization in an immersive environment. *Proceedings - Virtual Reality Annual International Symposium*.

- Bacim, F., Ragan, E., Scerbo, S., Polys, N. F., Setareh, M., & Jones, B. D. (2013). The effects of display fidelity, visual complexity, and task scope on spatial understanding of 3D graphs. *Proceedings - Graphics Interface*.
- Berkman, M. İ., Bostan, B., & Yalçın, B. (2020). Controllers in VR Game User Experience: Perceived User Performance on a VR Puzzle Game. In *Contemporary Topics in Computer Graphics and Games: Selected Papers from the Eurasia Graphics Conference Series*.
- Berkman, M. I., & Akan, E. (2019). Presence and Immersion in Virtual Reality. In *Encyclopedia of Computer Graphics and Games* (pp. 1–10). https://doi.org/10.1007/978-3-319-08234-9_162-1
- Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability Evaluation in Industry*.
- Buttussi, F., & Chittaro, L. (2018). Effects of Different Types of Virtual Reality Display on Presence and Learning in a Safety Training Scenario. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/TVCG.2017.2653117>
- Carroll, M., Osborne, E., & Yildirim, C. (2019). Effects of VR gaming and game genre on player experience. *2019 IEEE Games, Entertainment, Media Conference, GEM 2019*. <https://doi.org/10.1109/GEM.2019.8811554>
- Chirico, A., Cipresso, P., Yaden, D. B., Biassoni, F., Riva, G., & Gaggioli, A. (2017). Effectiveness of Immersive Videos in Inducing Awe: An Experimental Study. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-01242-0>
- Chirico, A., Ferrise, F., Cordella, L., & Gaggioli, A. (2018). Designing awe in virtual reality: An experimental study. *Frontiers in Psychology*, 8(JAN), 2351. <https://doi.org/10.3389/fpsyg.2017.02351>
- Christensen, J. V., Mathiesen, M., Poulsen, J. H., Ustrup, E. E., & Kraus, M. (2018). Player experience in a VR and non-VR multiplayer game. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3234253.3234297>
- Clarke, R. I., Lee, J. H., & Clark, N. (2017). Why Video Game Genres Fail: A Classificatory Analysis. *Games and Culture*. <https://doi.org/10.1177/1555412015591900>
- Elmqvist, N., Tudoreanu, M. E., & Tsigas, P. (2008). Evaluating motion constraints for 3D wayfinding in immersive and desktop virtual environments. *Proceeding of the Twenty-Sixth Annual CHI Conference on Human Factors in Computing Systems - CHI '08*, 1769. <https://doi.org/10.1145/1357054.1357330>
- Hart, S. G. (2006). NASA-task Load Index (NASA-TLX). *Proceedings of the Human Factors and Ergonomics Society*.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hartmann, T., Wirth, W., Schramm, H., Klimmt, C., Vorderer, P., Gysbers, A., ... Sacau, A. M. (2015). The spatial presence experience scale (SPES): A short self-report measure for diverse media settings. *Journal of Media Psychology*, 28(1), 1–15. <https://doi.org/10.1027/1864-1105/a000137>

- Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. *AAAI Workshop - Technical Report*.
- Ijsselsteijn, W. A., de Kort, Y. A. W., & Poels, K. (2007). Game Experience Questionnaire: development of a self-report measure to assess the psychological impact of digital games – Eindhoven University of Technology research portal. *PRESENCE 2007*.
- Johnsen, K., & Lok, B. (2008). An evaluation of immersive displays for virtual human experiences. *Proceedings - IEEE Virtual Reality*. <https://doi.org/10.1109/VR.2008.4480764>
- Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology*, 3(3), 203–220. https://doi.org/10.1207/s15327108ijap0303_3
- Kim, D., & Ko, Y. J. (2019). The impact of virtual reality (VR) technology on sport spectators' flow experience and satisfaction. *Computers in Human Behavior*, 93, 346–356. <https://doi.org/10.1016/j.chb.2018.12.040>
- Kim, K., Rosenthal, M. Z., Zielinski, D. J., & Brady, R. (2014). Effects of virtual environment platforms on emotional responses. *Computer Methods and Programs in Biomedicine*. <https://doi.org/10.1016/j.cmpb.2013.12.024>
- Kober, S. E., Kurzmann, J., & Neuper, C. (2012). Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: An EEG study. *International Journal of Psychophysiology*, 83(3), 365–374. <https://doi.org/10.1016/j.ijpsycho.2011.12.003>
- Kulshreshth, A., Schild, J., & LaViola, J. J. (2012). Evaluating user performance in 3D stereo and motion enabled video games. *Foundations of Digital Games 2012, FDG 2012 - Conference Program*. <https://doi.org/10.1145/2282338.2282350>
- Laha, B., Sensharma, K., Schiffbauer, J. D., & Bowman, D. A. (2012). Effects of immersion on visual analysis of volume data. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/TVCG.2012.42>
- Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. <https://doi.org/10.1007/978-3-540-89350-9-6>
- Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence: Teleoperators and Virtual Environments*, 10(3), 282–297. <https://doi.org/10.1162/105474601300343612>
- Li, H., & Giudice, N. A. (2013). The effects of immersion and body-based rotation on learning multi-level indoor virtual environments. *Proceedings of the 5th ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness, ISA 2013*. <https://doi.org/10.1145/2533810.2533811>
- Litwiller, T., & LaViola, J. J. (2011). Evaluating the benefits of 3d stereo in modern video games. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/1978942.1979286>

- Lugrin, J.-L., Cavazza, M., Charles, F., Le Renard, M., Freeman, J., & Lessiter, J. (2013). Immersive FPS games. *Proceedings of the 2013 ACM International Workshop on Immersive Media Experiences - ImmersiveMe '13*, 7–12. <https://doi.org/10.1145/2512142.2512146>
- Mania, K., & Chalmers, A. (2001). The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *Cyberpsychology and Behavior*. <https://doi.org/10.1089/109493101300117938>
- Mania, K., & Chalmers, A. (2002). The Effects of Levels of Immersion on Memory and Presence in Virtual Environments: A Reality Centered Approach. *CyberPsychology & Behavior*, 4(2), 247–264. <https://doi.org/10.1089/109493101300117938>
- Martel, E., Su, F., Gerroir, J., Hassan, A., Girouard, A., & Muldner, K. (2015). Diving Head-First into Virtual Reality—Evaluating HMD Control Schemes for VR Games. *Proceedings of the 10th International Conference on the Foundations of Digital Games - FDG 2015*. Retrieved from https://pdfs.semanticscholar.org/2c7d/b9069822f2599cdb466d51f17b28fd2baad6.pdf?_ga=2.187769836.550852407.1589226699-1224180204.1587306111
- McMahan, R. P., Bowman, D. A., Zielinski, D. J., & Brady, R. B. (2012). Evaluating display fidelity and interaction fidelity in a virtual reality game. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/TVCG.2012.43>
- Mizell, D. W., Jones, S. P., Slater, M., & Spanlang, B. (2002). Comparing immersive virtual reality with other display modes for visualizing complex 3D geometry. In *University College London, ...*
- Monteiro, D., Liang, H. N., Xu, W., Brucker, M., Nanjappan, V., & Yue, Y. (2018). Evaluating enjoyment, presence, and emulator sickness in VR games based on first- and third-person viewing perspectives. *Computer Animation and Virtual Worlds*. <https://doi.org/10.1002/cav.1830>
- Moroney, W. F., Biers, D. W., Eggemeier, F. T., & Mitchell, J. A. (2003). *A comparison of two scoring procedures with the NASA task load index in a simulated flight task*. <https://doi.org/10.1109/naecon.1992.220513>
- Pallavicini, F., Ferrari, A., Zini, A., Garcea, G., Zancacchi, A., Barone, G., & Mantovani, F. (2018). What distinguishes a traditional gaming experience from one in virtual reality? An exploratory study. *Advances in Intelligent Systems and Computing*. https://doi.org/10.1007/978-3-319-60639-2_23
- Pallavicini, F., Pepe, A., & Minissi, M. E. (2019). Gaming in Virtual Reality: What Changes in Terms of Usability, Emotional Response and Sense of Presence Compared to Non-Immersive Video Games? *Simulation & Gaming*, 50(2), 136–159. <https://doi.org/10.1177/1046878119831420>
- Patrick, E., Cosgrove, D., Slavkovic, A., Rode, J. A., Verratti, T., & Chiselko, G. (2000). Using a large projection screen as an alternative to head-mounted displays for virtual environments. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/332040.332479>
- Pausch, R., Proffitt, D., & Williams, G. (1997). Quantifying immersion in virtual reality. *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1997*. <https://doi.org/10.1145/258734.258744>

- Pestaluky, A., Kane, B., & Fetter, B. (2015). *Keep Talking and Nobody Explodes*. Retrieved from <https://www.keeptalkinggame.com/>
- Phan, M. H., Keebler, J. R., & Chaparro, B. S. (2016). The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human Factors*. <https://doi.org/10.1177/0018720816669646>
- Qi, W., Taylor, R. M., Healey, C. G., & Martens, J. B. (2006). A comparison of immersive HMD, fish tank VR and fish tank with haptics displays for volume visualization. *Proceedings - APGV 2006: Symposium on Applied Perception in Graphics and Visualization*. <https://doi.org/10.1145/1140491.1140502>
- Ragan, E. D., Kopper, R., Schuchardt, P., & Bowman, D. A. (2013). Studying the effects of stereo, head tracking, and field of regard on a small-scale spatial judgment task. *IEEE Transactions on Visualization and Computer Graphics*. <https://doi.org/10.1109/TVCG.2012.163>
- Riddle, K. (2010). Remembering Past Media Use: Toward the Development of a Lifetime Television Exposure Scale. *Communication Methods and Measures*, 4(3), 241–255. <https://doi.org/10.1080/19312458.2010.505500>
- Roetl, J., & Terlutter, R. (2018). The same video game in 2D, 3D or virtual reality – How does technology impact game evaluation and brand placements? *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0200724>
- Ruddle, R. A., Payne, S. J., & Jones, D. M. (1999). Navigating large-scale virtual environments: What differences occur between helmet-mounted and desk-top displays? *Presence: Teleoperators and Virtual Environments*. <https://doi.org/10.1162/105474699566143>
- Ruddle, R. A., & Péruch, P. (2004). Effects of proprioceptive feedback and environmental characteristics on spatial learning in virtual environments. *International Journal of Human Computer Studies*. <https://doi.org/10.1016/j.ijhcs.2003.10.001>
- Sauro, J., & Lewis, J. R. (2012). Quantifying the User Experience. In *Quantifying the User Experience*. <https://doi.org/10.1016/C2010-0-65192-3>
- Schild, J., LaViola, J. J., & Masuch, M. (2012). Understanding user experience in stereoscopic 3D games. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/2207676.2207690>
- Shelstad, W. J., Smith, D. C., & Chaparro, B. S. (2017). Gaming on the rift: How virtual reality affects game user satisfaction. *Proceedings of the Human Factors and Ergonomics Society*. <https://doi.org/10.1177/1541931213602001>
- Slater, M., Linakis, V., Usoh, M., & Kooper, R. (1995). Immersion, Presence, and Performance in Virtual Environments: An Experiment with Tri-Dimensional Chess. *Virtual Reality*. <https://doi.org/10.1.1.34.6594>

- Slobounov, S. M., Ray, W., Johnson, B., Slobounov, E., & Newell, K. M. (2015). Modulation of cortical activity in 2D versus 3D virtual reality environments: An EEG study. *International Journal of Psychophysiology*. <https://doi.org/10.1016/j.ijpsycho.2014.11.003>
- Sousa Santos, B., Dias, P., Pimentel, A., Baggerman, J. W., Ferreira, C., Silva, S., & Madeira, J. (2009). Head-mounted display versus desktop for 3D navigation in virtual reality: A user study. *Multimedia Tools and Applications*. <https://doi.org/10.1007/s11042-008-0223-2>
- Swindells, C., Po, B. A., Hajshirmohammadi, I., Corrie, B., Dill, J. C., Fisher, B. D., & Booth, K. S. (2004). Comparing CAVE, wall, and desktop displays for navigation and wayfinding in complex 3D models. *Proceedings of Computer Graphics International Conference, CGI*. <https://doi.org/10.1109/CGI.2004.1309243>
- Tan, C. T., Leong, T. W., Shen, S., Dubravs, C., & Si, C. (2015). Exploring gameplay experiences on the oculus rift. *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. <https://doi.org/10.1145/2793107.2793117>
- Usoh, M., Catena, E., Arman, S., & Slater, M. (2000). Using presence questionnaires in reality. *Presence: Teleoperators and Virtual Environments*. <https://doi.org/10.1162/105474600566989>
- Ventura, S., Brivio, E., Riva, G., & Baños, R. M. (2019). Immersive Versus Non-immersive Experience: Exploring the Feasibility of Memory Assessment Through 360° Technology. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2019.02509>
- Vorderer, P., Wirth, W., Gouveia, F. R., Biocca, F., Saari, T., Jäncke, F., ... Jäncke, P. (2004). *MEC Spatial Presence Questionnaire (MECSPQ): Short Documentation and Instructions for Application*. Retrieved from https://www.researchgate.net/profile/Feliz_Gouveia/publication/318531435_MEC_spatial_presence_questionnaire_MEC-SPQ_Short_documentation_and_instructions_for_application/links/598041b5458515687b4fa65d/MEC-spatial-presence-questionnaire-MEC-SPQ-Short-docume
- Waller, D., Hunt, E., & Knapp, D. (1998). The transfer of spatial knowledge in virtual environment training. *Presence: Teleoperators and Virtual Environments*. <https://doi.org/10.1162/105474698565631>
- Weidner, F., Hoesch, A., Poeschl, S., & Broll, W. (2017). Comparing VR and non-VR driving simulations: An experimental user study. *Proceedings - IEEE Virtual Reality*. <https://doi.org/10.1109/VR.2017.7892286>
- Wilson, R., & Mayhorn, C. B. (2019). Examining the Role of Video in Sports Media Viewing. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. <https://doi.org/10.1177/1071181319631424>
- Winn, W., Windschitl, M., Fruland, R., & Lee, Y. (2002). When Does Immersion in a virtual Environment Help Students Construct Understanding? *ICLS 2002*.
- Wirth, W., Vorderer, P., Hartmann, T., Klimmt, C., & Schramm, H. (2003). Constructing Presence : Towards a two-level model of the formation of Spatial Presence experiences. In *Communication* (Vol. 2003). Retrieved from https://www.researchgate.net/publication/318531733_Constructing_Presence_Towards_a_two-level_model_of_the_formation_of_Spatial_Presence

- Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments*.
<https://doi.org/10.1162/105474698565686>
- Yildirim, Ç., Bostan, B., & Berkman, M. İ. (2019). Impact of different immersive techniques on the perceived sense of presence measured via subjective scales. *Entertainment Computing*.
<https://doi.org/10.1016/j.entcom.2019.100308>
- Yildirim, C., Carroll, M., Hufnal, D., Johnson, T., & Pericles, S. (2018). Video Game User Experience: To VR, or Not to VR? *2018 IEEE Games, Entertainment, Media Conference (GEM)*, 1–9.
<https://doi.org/10.1109/GEM.2018.8516542>
- Zanbaka, C. A., Lok, B. C., Babu, S. V., Ulinski, A. C., & Hodges, L. F. (2005). Comparison of path visualizations and cognitive measures relative to travel technique in a virtual environment. *IEEE Transactions on Visualization and Computer Graphics*, 11(6), 694–705.
<https://doi.org/10.1109/TVCG.2005.92>