This is an Accepted Manuscript of an article published by Taylor & Francis in International Journal of Human-Computer Interaction on 20/08/2019 (online publication date), available online: https://www.tandfonline.com/doi/full/10.1080/10447318.2019.1655905

Czarnek, G., Strojny, P., Strojny, A., & Richter, M. (2019). Assessing engagement during rescue operation simulated in virtual reality: A psychophysiological study. *International Journal of Human-Computer Interaction*. doi: 10.1080/10447318.2019.1655905

Assessing Engagement During Rescue Operation Simulated in Virtual Reality: A Psychophysiological Study

Gabriela Czarnek[1,2], Paweł Strojny[1,3], Agnieszka Strojny[1,3], Michael Richter[4]

[1] Nano Games, Cracow (Poland)

[2] Institute of Psychology at Jagiellonian University, Cracow (Poland)

[3] Faculty of Management and Social Communication at Jagiellonian University, Cracow (Poland)

[4] School of Natural Sciences and Psychology at Liverpool John Moores, Liverpool (UK)

Correspondence concerning this article should be addressed to Gabriela Czarnek, gczarnek@nano-games.com, Nano Games, Gromadzka 101 Street, 30-719 Kraków, Poland.

This research was accepted by Ethical Committee at the Jagiellonian University Institute of Applied Psychology, Cracow (Poland).

ABSTRACT

The goal of this study was to investigate the patterns of engagement among professional firefighters during a rescue operation challenge simulated in a virtual reality (VR). The simulator offers a training that would otherwise be impossible or very difficult to arrange in the real world, here a mass-casualty incident. We measured engagement with cardiovascular reactivity as well as subjective perceptions of workload. We found that both a VR rescue challenge and a VR control condition lead to engagement evident in the decrease in parasympathetic activation from baseline (measured as high-frequency heart rate variability). However, the rescue operation lead to a stronger increase in sympathetic activity (shorter pre-ejection period and RZ-interval) than the control condition. Furthermore, the subjective workload ratings corroborate the results from the objective engagement indices. These results demonstrate that it is possible to create a virtual environment that elicits engagement among professional rescuers. [145 words]

Keywords: virtual reality; professional simulators; rescue operation; engagement; cardiovascular measures

Funding body

This work was co-financed under *Widespread Disaster Simulator – research and preparation for implementation* (the Smart Growth Operational Programme, sub-measure 1.1.1. Industrial research and development work implemented by enterprises) received by Nano Games sp. z o.o.

Disclosure statement

The funding body did not interfere with the design of a study, data collection, analysis or interpretation of a data, nor in the decision of publishing the paper or choosing the journal.

First three authors of the manuscript, Gabriela Czarnek, Paweł Strojny, and Agnieszka Strojny, were employed at Nano Games which developed the simulator being tested.

Supplemental online material

This paper has supplemental materials which have been submitted to the review.

Acknowledgments

We would like to thank Natalia Dużmańska, Natalia Lipp, Krzysztof Rębilas, and Radosław Sterna for their work on data collection and pre-processing.

Data availability

The data that support the findings of this study are openly available in Open Science Framework at https://osf.io/mja7p.

1. INTRODUCTION

1.1. Goal and importance

Firefighting is a demanding profession, both physically and psychologically (e.g., Bos, Mol, Visser, & Frings-Dresen, 2004; Guidotti, 1992; Williams-Bell et al., 2015). Due to the myriad of incidents that firefighters are required to deal with along with the serious consequences of their actions, their training is extensive and demanding. However, responses to events of low frequency and high damage, are very difficult to train. This is because of the high costs and little flexibility of live training of such events (e.g., impossibility to repeat a specific rescue operation). Thus, virtual reality (VR) technology offer a valuable complement to the current firefighting curriculum. They offer a cost-effective opportunity for the training of otherwise difficult or impossible situations. However, an important question is whether the rescue challenge in VR simulators can actually elicit engagement among firefighters. Many researchers claimed that VR technology provides an engaging educational environment (Dede, 2009; Psotka, 1995; Schutte & Stilinovic, 2017) but others doubted that all VR applications result in engagement (Harris & Reid, 2005; Martin-Niedecken & Gotz, 2017; Mineo, Ziegler, Gill, & Salkin, 2009) and suggested that the efficacy of VR applications to engage users needs yet to be proven (Huang, Rauch, & Liaw, 2010). Hence, the goal of the current study was to examine whether a VR application simulating a rescue operation elicits engagement among professional firefighters.

1.2. Training in Virtual Reality

The firefighter training involves mainly in-class demonstrations and live training (Heldal & Hammar Wijkmar, 2017). Interestingly, only a small proportion of firefighters' work is actual fighting the fire. In fact, the majority of their duties is related to traffic collisions, natural disasters, gas or water leaks, elevator accidents, or industrial accidents (e.g., Austin, Dussault, & Ecobichon, 2001; Lusa, Louhevaara, & Kinnunen, 1994; Park, Jang, & Chai, 2006; Williams–Bell et al., 2015). Workload of professional rescuers is usually high, especially among firefighters who tend to work at great pace, long hours, with heavy equipment and protective gear, often in extreme thermal conditions which puts great cardiovascular and thermoregulatory demands on the body (e.g., Bos et al., 2004; lmer & Gavhed, 2007; Smith et al., 1995). For example, it was shown that firefighting exercise lead to reaching maximal, or near maximal, heart rate (e.g., Smith, Manning, & Petruzello, 2001). Moreover, it was demonstrated that highest demand is placed when firefighters climb or carry heavy objects (Holmer & Gavhen, 2007). Apart from a wide variety of tasks, additional burden of this profession is a changing nature of the incidents and skills required to deal with them, e.g., various constructions of buildings, new materials used, and novel societal challenges such as terrorist attacks (Harrald, 2006; McDevitt, 2017; Meissner et al., 2002). Because of these challenges, professional VR simulators might offer an important supplement to the current training practice. The advantages of VR technology in firefighter training are evident.

First, firefighters, novice and as well as more experienced ones, can test their reactions to various incidents in a safe environment. For example, they may try, repeat, and improve their reactions to fire in different structures and under different weather conditions, such as wind, rain, etc. Second, VR applications are a cost-effective method of simulating low-frequency or mass-casualty incidents (McDevitt, 2017). The costs of organizing real-life simulations include not only hiring the staff, careful preparations, and the materials that is used but also the cost of long-distance travel of the firefighters to a training center (Haldal & Hammar Wijkmark, 2017). Furthermore, the costs of the VR equipment and applications has significantly decreased in recent years (e.g., Evalt, 2017). Third, VR provides an opportunity to train reactions to situations which would be otherwise impossible to perform, for example a large port cannot be closed to simulate a tanker fire (Heldal & Hammar Wijkmark, 2017). Fourth, trainings in VR may offer an instant feedback on performance, which is not possible in real-life simulations. Exposure to errors is especially important during initial skill acquisition and improves retention and performance (Gardner, Abdelfattah, Wiersch, Ahmed, & Willis, 2015; Keith & Frese, 2008). A further benefit of VR is that it provides a high degree of control and flexibility: The VR simulation might be adapted to the needs and the proficiency level of a particular user (e.g., Gallagher et al., 2015). Finally, even if trainees possess the theoretical knowledge and skills to react correctly, they may struggle responding properly during real emergency and time pressure (McDevitt, 2017). VR offers an environment that is as close to an actual emergency as possible and provides thus the best preparation for real-life crisis. In sum, VR technology provides a unique opportunity to train skills required in rescue challenges, especially situations which are low in frequency, dangerous, and associated with increased time pressure in decision-making.

In a few occupations, especially those where mistakes lead to fatal consequences, professional simulators have already been adopted for training. Professional simulators are used in medicine, army, air traffic control, power plant management and emergency management trainings, to name but a few (e.g., Boulet et al., 2003; Dimakis, Filippoupolitis, & Gelenbe, 2010; Gatto, Mól, dos Santos, Jorge, & Legey, 2013; Khan et al., 2013). For example, in healthcare, there are VR applications allowing users to practice open and laparoscopic surgeries (e.g., Bric, Lumbard, Frelich, & Gould, 2016; Lam, Sundaraj, & Sulaiman, 2013), drug administration (e.g., Dubovi, Levy, & Dagan, 2017), or interpersonal and conversational skills for novice clinicians (e.g., Kenny & Parsons, 2011).

Currently, there are several simulators available for firefighter training (see Williams – Bell et al., 2015, for an overview). For example, St Julien and Shaw (2003) created the Firefighter Command Training Virtual Environment, an application for command training for officers. In the simulator, trainees lead a group of virtual agents during extinguishing a fire. Another VR application, the Advanced Disaster Management Simulator (The Environmental Tectonics Corporation), offers a training on how to respond in emergency situation, such as a fire or a terrorist attack. It provides a training on crisis management, including coordination of on-ground personnel and distribution of equipment. Furthermore, Backlund and others (Backlund et al., 2007, 2009) developed SIDH, an immersive game-based firefighter training application, which is simulated using cave virtual environment. Trainees can be exposed to up to 13 different scenarios in which they are supposed to search for victims while wearing breathing apparatus. Despite the existence of several VR applications for firefighters training, the majority of them is focused on emergency management, i.e. the training of the leaders (Williams–Bell et al., 2015). In contrast, in the current study we tested a VR application in which firefighters of various experiences and levels can practice their reactions in a demanding rescue challenge, such as a mass-casualty incident in which they have an opportunity to act as an individual member of a basic tactical unit.

1.3. Engagement

Engagement is defined as energy expenditure to perform an instrumental behavior (see Richter, 2013). This is an important component of an efficient training as it leads to a better performance (e.g., Bakker, Vergel, & Kuntze, 2014; Freeman et al., 2014). In the educational context, researchers focus on three aspects of engagement: cognitive, behavioral, and emotional (e.g., Fredericks, Blumenfeld, & Paris, 2004). In the current setting, the most important aspects of engagement are the cognitive and behavioral ones as they both relate to effort, attention, and concentration on a given task. It was suggested that the reason why engagement is important in education is that it helps students to be more goal-oriented, which in turn increases their chance for performance and learning success (Bakker et al., 2014). In training settings that are not engaging, it is unlikely that trainees try out different strategies or invest the effort that is required to solve the problem. Consequently, they could miss the opportunity to improve their skills and knowledge. An efficient learning environment, thus, needs to be engaging.

In empirical research, engagement is often measured with questionnaires, which sometimes can be problematic. First, people may easily fake their responses trying to conform to the researcher's expectations (e.g., Furnham & Henderson, 1982; McKibben & Silvia, 2017). Furthermore, people may not be aware of their internal processes, for example they may not be aware of what caused their behavior or how they made a particular decision (e.g., Nisbett & Wilson, 1977). Finally, they might be prone to self-serving biases, e.g., unrealistically perceive their intellect in a way that would enhance their self-esteem (e.g. Paulhus & John, 1998). For those reasons, the relationship between the perception of engagement and actual engagement may not be straightforward (e.g., Harper, Eddington, & Silvia, 2016; Muraven, Tice, & Baumeister, 1998; Smith & Hess, 2015). Thus, in this research we focused on objective, physiological measures of engagement. In particular, we focused on cardiovascular measures of engagement. We also employed a subjective measure of workload to complement the analysis of actual engagement with the subjective perceptions.

Drawing on decades of research on active coping hypothesis (Obrist, 1976, 1981), the motivational intensity theory (Brehm & Self, 1989; Wright, 1996), the autonomic space model (Berntson, Cacioppo, & Quigley, 1991), and on cardiovascular responses to physical exercise (Rowell, 1993; White & Raven, 2014), we assumed that objective engagement is reflected by changes in the activity of the autonomic nervous system. In particular, engagement should be reflected by an increased activity of the sympathetic nervous system and a decreased activity of the parasympathetic nervous system (these are the two branches of autonomic nervous system). Given that the cardiovascular system is affected by the sympathetic and parasympathetic branches, studying its function allows inferences about autonomic nervous system activity and this has been done in multitude of studies (e.g., Obrist, 1976; Van Roon, Mulder, Althaus, & Mulder, 2004; White & Raven, 2014; Wright, 1996). Some indices, like heart rate or blood pressure, reflect the interplay between the sympathetic and parasympathetic nervous systems and provide general information about autonomic activity (Berntson, Quigley, & Lozano, 2007). Other indices, like pre-ejection period or high-frequency heart rate variability, are more specific and reflect more selectively sympathetic or parasympathetic activity.

Among the non-invasive indices of sympathetic impact on the cardiovascular system, pre-ejection period is considered the most sensitive measure (Kelsey, 2012; Sherwood et al., 1990). Pre-ejection period is the time interval from the onset of left ventricular depolarization, referring to Q-onset in the electrocardiogram, until the opening of the aortic valve, the B-point in the impedance cardiography signal (Berntson, Lozano, Chen, & Cacioppo, 2004; Sherwood et al., 1990). The shorter the pre-ejection period, the stronger the sympathetic impact on the heart, indicating higher engagement. A related measure, RZ-interval is the time interval between the R-peak in the electrocardiogram, and the maximum dZ/dt peak in the impedance cardiogram (Lozano et al., 2007; Sherwood et al., 1990). Similarly, to pre-ejection period, the shorter the RZ-interval, the stronger sympathetic impact on the heart. The advantage of RZ-interval over pre-ejection period is that it can be estimated more reliably than pre-ejection period given that it can be difficult to identify the B-point in a noisy signal (Kuipers et al., 2016; Lozano et al., 2007; Silvia, Beaty, Nusbaum, Eddington, & Kwapil, 2014).

In contrast to pre-ejection period and RZ-interval as measures of sympathetic activity, assessing heart rate variability enables the estimation of parasympathetic influence on the heart. Heart rate variability is an umbrella term for a wide array of methods, e.g., time-domain analysis, frequency-domain analysis, rhythm pattern and other types of analyses, assessing variations in the interval between consecutive heart beats (Berntson et al., 1997; Task Force, 1996). Most importantly, the frequency domain analysis allows estimating parasympathetic influence on the heart. While low frequency heart rate variability (0.04-0.15 Hz) represents both sympathetic and parasympathetic activity, high frequency heart rate variability (0.15-0.40 Hz) represents parasympathetic activity selectively (Berntson et al., 1997; Task Force, 1996). Thus, to capture the parasympathetic activity, we focused only on high frequency power range of heart rate variability. As parasympathetic nervous system decreases heart activity (it operates as a brake on the heart), the stronger the decrease in parasympathetic activity, the higher the level of engagement. In other words, the stronger the withdrawal of the “vagal brake”, the higher the engagement.

Physiological measures of engagement have already been used to demonstrate that VR applications lead to high levels of engagement among users. For example, Kothgassner and colleagues (2016) demonstrated that public speaking tasks in virtual and real-life environment elicit similar physiological responses, i.e. elevated salivary cortisol and heart rate. For heart rate, there was an increase of about 20 beats per minute in the beginning of the task in both real and VR public speaking conditions, in comparison to a 10-beat increase in the control condition. There were, however, no such differences for their heart rate variability indices. Similarly, Crescentini, Chittaro, Capurso, Sioni, and Fabbro (2016) showed that exposure to VR emergency situation results in higher heart rate (an increase of around 5 beats per minute from baseline to a task period) and this relationship was moderated by mindfulness training. Furthermore, Parsons, Rizzo, Courtney, and Dawson (2012) showed that heart rate was higher when participants faced a challenging task condition, vs. an easier one, but only when the virtual environment was presented using *immersive* technologies (i.e., head mounted display) but not when it was presented on a computer screen. To the contrary, Egan and others (2016) did not find differences in heart rate when they compared an exposure to a VR vs. non-VR environments. However, users were only watching a virtual city and followed a path pre-defined by the researchers which may explain no differences in heart rate between conditions. Furthermore, another study by Gorini, Capideville, De Leo, Mantovani, & Riva (2011) showed a similar high heart rate response when the task was presented using head mounted display or a computer screen but with a meaningful narrative framework (an increase in heart rate was around 13 and 9 beats per minute, respectively). In narrative condition, participants performed a task as doctors who had to find a container with a rare blood which was to be transfused for a seriously sick child. It was contrasted with a low heart rate reactivity in a non-narrative/immersive as well as non-narrative/non-immersive conditions (an increase of around 2 beats per minute). On the other hand, the reported *presence,* which is a technology-mediated feeling of being present in a simulated environment (e.g., Heeter, 1992), was highest in the narrative/immersive condition. This study demonstrates the importance of both including immersive methods and well as narrative contexts to create most engaging environments. Relatedly, it was shown that heart rate responses were positively but weakly related to higher sense of presence (Meehan, Razzaque, Insko, Whitton, and Brooks, 2005) but other studies did not confirm that link (e.g., Felnhofer et al., 2014). It suggests that although presence and engagement might be related, they are conceptually, and physiologically, distinct.

As reviewed above, several studies documented autonomic responses to tasks in VR environment. Nevertheless, to the best of our knowledge, the present study is a first one to examine the whole spectrum of engagement-related autonomic responses (sympathetic activity and parasympathetic withdrawal) in a VR application.

1.4. Overview of the study

In this study we investigated whether VR application elicits engagement during rescue operation challenge. In particular, we ran an experimental study with two groups of firefighters. In the experimental condition participants engaged in a VR rescue operation challenge. In the control condition, participants were free to explore the same VR setting without a rescue goal. Thus, the differences between the experimental and the control group cannot be attributed to the excitement of the first time VR experience as both groups used VR setup.

We measured objective engagement using cardiovascular responses: high frequency heart rate variability, pre-ejection period, RZ-interval, and heart rate. These measures captured both sympathetic activation (pre-ejection period, RZ-interval) as well as parasympathetic withdrawal (high frequency heart rate variability). Furthermore, we measured the subjective workload that the VR task required. We hypothesized that participants would demonstrate higher levels of objective engagement (shorter pre-ejection period and RZ-interval, lower high frequency heart rate variability) and subjective workload in the VR rescue challenge condition than in the control condition.

2. METHOD

2.1. Participants

Due to external constraints the number of participants was fixed to 60 participants: 59 men and 1 woman aged between 19-24 (*M* = 21.58, *SD* = 1.45) took part in the study. A sensitivity power analysis in *GPower* (ver. 3.1.9.2; Faul, Erdfelder, Buchner, & Lang, 2009; alpha = .05) showed that we had a power of 80% to detect an effect of *f* = 0.37, which is moderate to large in size, for the comparison of the rescue challenge and the control conditions. All the participants were firefighter's trainee in the second year of their education. All of them already had a considerable experience of performing in real rescue operations. Participants were recruited at the College of the State Fire Service in Cracow (Poland). One participant was excluded from the dataset due to problems with electrocardiogram recording. One person more was excluded from the analysis of pre-ejection period and RZ-interval due noisy impedance cardiography signal (it was impossible to correctly mark landmarks in their ICG signal). Nevertheless, his person has been taken into account in the analysis of subjective perceptions of task, heart rate, and high frequency heart rate variability. Thus, the final sample size for analysis of sympathetic activation, *i.e.,* pre-ejection period and RZ-interval, was 58 participants and for any other reported measure the final sample size was 59. The sensitivity analysis for such a restricted sample size (*N*=58) showed that we had a power of 80% to detect an effect of *f* = 0.37, for the comparison of the rescue challenge and the control conditions.

2.2. Measures and Equipment

Task

Participants were immersed in VR with a stereoscopic head-mounted display (HMD). In particular, we used HTC Vive (field of view 110 degrees, refresh rate 90 Hz). The HMD was connected to a PC with a 3.40 GHz Intel Core i7 processor, 16GB of RAM and Nvidia GeForce GTX 1080. To operate within a task, participants used HTC Vive wireless controllers held in both hands. The VR application was developed using Unity engine (Unity Technologies, San Francisco, CA).

In the rescue challenge condition, participants performed a simulated rescue operation. In particular, their task was to perform a standard rescue procedure (detailed in the documentation of the National Firefighting and Rescue System[1]) in VR and mark critically injured victims of the car crash. For example, they were supposed to check victims' consciousness, airways, breathing, circulation, and

---

[1] Available at http://www.straz.gov.pl/english/national_firefighting_rescue_system

mark them if they required further medical help. The time limit for the task was 5 minutes. There were 6 victims with various injuries in the car crash. A sample screenshot from the rescue challenge (experimental) condition is presented in Figure 1 Panel B. In the control condition, participants explored a similar VR environment but without a car crash, victims, or a rescue goal in this task for 5 minutes. A sample screenshot from the control condition is presented in Figure 1 Panel A. In both conditions, participants navigated from a first-person perspective.

[Figure 1 about here, PLEASE USE B&W IN PRINT FOR THIS FIGURE]

Figure 1. Sample screenshots from a task in the control condition (Panel A) and the rescue challenge condition (Panel B).

Self-reported workload

We measured subjective workload with the NASA-TLX (Hart & Staveland, 1988, Polish version by Zieliński & Biernacki, 2010) in which participants rate their perception of task mental workload, physical workload, time pressure, effort, performance, and frustration (there is one item per dimensions). All responses are given using a 21-point scale. In the results section, we present separate scores for each one of the subscales.

Physiological Acquisition

During the experiment, we measured participants' electrocardiogram and impedance cardiogram using a BIOPAC MP160 system (BIOPAC Systems Inc., Goleta, CA, USA). We used a 3-lead electrocardiography setup with the pre-gelled Ag/AgCl spot electrodes attached to a participant torso (on the right and left clavicle as well as on the lower left abdomen). The impedance cardiogram was measured with 4 sets of Ag/AgCl pre-gelled spot electrodes placed on both sides of a base of a neck and both sides of lower abdomen (the distance between inner neck and abdomen electrodes was approximately 30 cm). Before placing the electrodes, participants' skin was abraded with ELPREP gel. Both the electrocardiogram and impedance cardiogram signal were sampled at 1000 Hz with a BioNomadix BN-ECG-2 and a BioNomadix-NICO, respectively (BIOPAC Systems Inc., Goleta, CA, USA). Data was stored with AcqKnowledge 5.0 software (BIOPAC Systems, Goleta, CA, USA).

2.3. Procedure

Participants signed informed consents and answered a set of demographic questions. Next, they were randomly assigned to one of the two conditions. Participants were asked to wear the HTC Vive HMD and to take the HTC wireless controllers. Subsequently, they were presented with the VR environment and were instructed how to move. Participants in the rescue challenge condition were additionally instructed how to move objects and start an interaction with a victim of an accident. After the training, participants were asked to remove the HMD and controllers and rated their affect[2]. Later, participants were asked to wear the HMD again and were watching a relaxing movie for 8 minutes. During that period, we measured participants baseline electrocardiogram and impedance cardiogram. Next, participants were given task instruction and performed the main task for 5 minutes. After the task, they stood still for two minutes while being immersed in VR. After this, participants removed the HMD and controllers and completed questionnaires measuring their affect and perception of the task. In short, the procedure was as follows: consent and instructions, pre-task self-report ratings, cardiovascular baseline measurement (8 minutes), task performance (5 minutes), cardiovascular recovery measurement (2 minutes), and post-task self-report ratings.

2.4. Offline Physiological Analysis

In a first step, we filtered the electrocardiogram (0.5–40 Hz) and the impedance cardiogram signals (0.5–50 Hz, Hurwitz et al., 1993). Then, QRS complex boundaries were automatically located with a Pan-Tomkins (Pan & Tomkins, 1985) algorithm. The electrocardiogram signal was visually inspected and corrected. C-points were automatically detected with an adaptive template matching method (BIOPAC, 2016) and B-points were identified using the R-C polynomial method (Lozano et al., 2007). Impedance cardiogram signals with the detected B- and C-points were visually inspected and corrected, if needed. In order to derive pre-ejection period and RZ-intervals, we calculated coherent averages over 1-minute periods (Hurwitz, Shyu, Reddy, Schneiderman, & Nagel, 1990). Pre-ejection period was defined as the interval between Q-onset and B-point, whereas RZ-interval was defined as the interval between R-peak and C-point (e.g., Sherwood et al., 1990).

For heart rate variability and heart rate, we derived RR-intervals (the time difference between two consecutive R-peaks, measured in milliseconds). The RR-intervals were submitted into the HRVAS software (Ramshur, 2010). Ectopic beats were identified as those RR-intervals which were larger than 20% or 3 standard deviations in comparison to a preceding RR-interval (Akhter, Gite, Tharewal, & Kale, 2015). RR-intervals marked as ectopic were removed from analysis (Lippman, Stein, & Lerman, 1994). Furthermore, we used wavelet-based method for a trend removal (Thuraisingham, 2006). The RR-interval signal was interpolated with 6 Hz (Singh, Vinod, & Saxena, 2004). Spectral decomposition was performed with Welch periodogram method. The very low frequency band of heart rate variability was defined as up to 0.04 Hz, the low frequency band as 0.04-0.15 Hz, and the high frequency band as 0.15-0.40 Hz (Task Force, 1993). High frequency heart rate variability was calculated in normalized units

[2] In the current study, we measured affect, emotion, stress and several dimensions of task percepcion. These measures, as not of a vital importance for the current manuscript, are presented in the Supplemental material section.

(relative to the total power in the LF and HF band). Cardiovascular baseline scores were calculated as an arithmetic mean of the data collected during the last 5 minutes of the baseline period. Cardiovascular task scores were calculated as an arithmetic mean of the data collected during the 5 minutes of the task. Next, we computed cardiovascular change scores (delta) by subtracting baseline scores from task scores (Llabre, Spitzer, Saab, Ironson, & Schneiderman, 1991). The delta scores for pre-ejection period, RZ-interval, heart rate variability, and heart rate served as our dependent variables of task-related cardiovascular reactivity.

2.5. Statistical Analysis

To examine whether performing a task in our VR application leads to engagement, we first ran a set of repeated measures ANOVA comparing cardiovascular baseline and task scores. To examine whether the rescue challenge led to higher engagement than the control condition, we compared both groups with one-way ANOVAs. With those two steps of the analysis (for the cardiovascular measures) we could separately examine the impact of VR on engagement and verify whether the engagement is especially pronounced in the rescue challenge condition in comparison to the control condition.[3]

3. RESULTS

3.1. Cardiovascular reactivity

Cardiovascular baseline values and change scores are presented in Table 1.

		Baseline values		Task change score		Significance
Measure	Condition	Mean	SD	Mean	SD	
High frequency heart rate variability	Control	0.38	0.17	-0.14	0.13	*ns*
	Rescue challenge	0.41	0.16	-0.18	0.17	
Pre-ejection period	Control	123.04	11.13	-1.40	6.58	*
	Rescue challenge	122.24	10.04	-7.32	6.89	
RZ-interval	Control	146.52	14.93	-5.86	7.60	*
	Rescue challenge	147.10	15.96	-21.52	12.62	

[3] An introduction to this statistical analysis strategy can be found in the books by Field (2013; especially chapters 11 and 14) and Field, Miles, and Field (2012; especially chapter 10 and 13).

Heart rate	Control	69.92	8.25	5.85	3.42	*
	Rescue challenge	66.84	10.70	11.41	4.72	

Table 1. Cell means for cardiovascular baseline values and change scores during the task. ***Note.*** High frequency heart rate variability is in normalized units, pre-ejection period and RZ-interval are in ms, heart rate is in bpm. An asterisk indicates a significant difference between the two conditions (that is, a *p*-value less than or equal to .05), "ns" indicates a non-significant difference.

For high frequency heart rate variability, task scores were lower than baseline scores in both the control, $F(1,28) = 33.814$, $MSE = 0.009$, $p < .001$, $\eta_G^2 = .211$, and the rescue challenge conditions, $F(1,29) = 33.096$, $MSE = 0.014$, $p < .001$, $\eta_G^2 = .327$. For pre-ejection period, while task scores did not differ from baseline in the control condition, $F(1,28) = 1.315$, $MSE = 21.672$, $p = .261$, $\eta_G^2 = .005$, they were lower than baseline scores in the rescue challenge condition, $F(1,29) = 32.713$, $MSE = 23.732$, $p < .001$, $\eta_G^2 = .117$. For RZ-interval, task scores were lower than baseline scores in both the control, $F(1,28) = 17.220$, $MSE = 28.885$, $p < .001$, $\eta_G^2 = .035$, and the rescue challenge conditions, $F(1,29) = 84.333$, $MSE = 79.656$, $p < .001$, $\eta_G^2 = 0.305$. Finally, for heart rate, task scores were higher than baseline scores in both the control, $F(1,28) = 84.699$, $MSE = 5.855$, $p < .001$, $\eta_G^2 = .012$, and the rescue challenge conditions, $F(1,29) = 175.050$, $MSE = 11.149$, $p < .001$, $\eta_G^2 = .242$.

For high frequency heart rate variability, we did not observe the predicted difference between the rescue challenge and the control conditions, $F(1,57) = 0.734$, $MSE = 0.023$, $p = .395$, $\eta_G^2 = .013$. However, for pre-ejection period, $F(1,56) = 11.175$, $MSE = 45.404$, $p = .001$, $\eta_G^2 = .166$, RZ-interval, $F(1,56) = 32.791$, $MSE = 108.542$, $p < .001$, $\eta_G^2 = .369$, and heart rate, $F(1,57) = 26.646$, $MSE = 17.097$, $p < .001$, $\eta_G^2 = .319$, we found the predicted effect. In comparison to the control condition, participants had shorter pre-ejection period, shorter RZ-interval, and higher heart rate in the rescue challenge condition. Cardiovascular change scores are presented in Figure 2.

[Figure 2 about here]

Figure 2. Cardiovascular change scores for high frequency heart rate variability, heart rate, pre-ejection period and RZ-interval in the control and rescue challenge conditions.

3.2. Subjective Engagement

For NASA-TLX, we found the effects of the task condition on mental workload, $F(1,57) = 10.533$, $MSE = 428.728$, $p = .002$, $\eta_G^2 = .160$, time pressure, $F(1,57) = 60.372$, $MSE = 446.796$, $p < .001$, $\eta_G^2 = .514$, effort, $F(1,57) = 13.785$, $MSE = 378.960$, $p < .001$, $\eta_G^2 = .195$, performance, $F(1,57) = 37.857$, $MSE = 409.964$, $p < .001$, $\eta_G^2 = .399$, and frustration ratings, $F(1,57) = 13.149$, $MSE = 328.937$,

$p < .001$, $\eta_G^2 = .187$. In contrast, we did not find an effect of the task condition on physical workload, $F(1,57) = 0.214$, $MSE = 196.600$, $p = .645$, $\eta_G^2 = .004$. In particular, we found that participants reported higher workload on all NASA-TLX subscales, except physical workload, in the rescue challenge condition in comparison to control condition. Finally, for the composite score of all NASA-TLX subscales averaged, we found an effect of the condition, $F(1,57) = 33.616$, $MSE = 207.038$, $p < .001$, $\eta_G^2 = .370$, with people in the rescue challenge condition reporting higher overall NASA-TLX scores. The average scores of the NASA-TLX subscales are presented in Table 2 and Figure 3.

Subscale	Condition	Mean	SD	Significance
Mental Workload	Control	30.00	19.13	*
	Rescue challenge	47.50	22.12	
Physical Workload	Control	19.31	14.38	*ns*
	Rescue challenge	21.00	13.67	
Time Pressure	Control	16.90	15.61	*
	Rescue challenge	59.67	25.36	
Performance	Control	16.72	16.05	*
	Rescue challenge	49.17	23.60	
Effort	Control	20.35	17.78	*
	Rescue challenge	39.17	20.97	
Frustration	Control	16.21	12.15	*
	Rescue challenge	33.33	22.45	

Table 2. Cell means for subjective workload ratings in NASA-TLX. ***Note.*** An asterisk indicates a significant difference between the two conditions (that is, a *p*-value less than or equal to .05), "ns" indicates a non-significant difference.

[Figure 3 about here]

Figure 3. Subjective workload ratings in NASA-TLX.

4. DISCUSSION

The goal of this study was to investigate the patterns of engagement elicited by a rescue operation simulated in VR. We found that cardiovascular responses to VR application both in the control and the rescue challenge conditions differed from baseline across all indices (except for pre-ejection period in the control condition). This suggests that performing a task in the VR environment led to a decrease in parasympathetic activity and an increase in sympathetic activity. More importantly, we found that performing the rescue challenge resulted in shorter pre-ejection period and shorter RZ-interval than the mere presence in the VR environment (control condition). This means that performing a simulated rescue operation led to an additional increase in sympathetic activation. These results are in line with the theories of engagement suggesting that on lower levels of task difficulty, engagement is driven by parasympathetic withdrawal but to support performance at higher levels of difficulty, there is an additional increase in the sympathetic activity (White & Raven, 2014; Van Roon et al., 2004). Furthermore, we found larger heart rate in the rescue challenge condition in comparison to the control condition. The overall pattern of results suggests that this increase in heart rate was mainly driven by the increase in sympathetic activity.

Apart from quantifying engagement physiologically, we used measures of subjective engagement (NASA-TLX). The results from the self-report mirrored the effects observed for the objective engagement. Specifically, we found that people who took part in a simulated rescue operation reported higher subjective engagement than those in the control condition. We found this effect for perception of mental demands, time pressure, performance, effort, and frustration but not for the perception of physical demands. Overall, our results demonstrated that the simulated rescue operation was not only perceived as but also objectively engaging to professional firefighters. This is important finding because, as already mentioned, engagement is the prerequisite for efficient learning and higher engagement leads to better learning outcomes (Bakker et al., 2014; Freeman et al., 2014).

Unexpectedly, for the subjective engagement measures, we found that the simulated rescue challenge was perceived as more frustrating in comparison to the control condition. This is a noteworthy finding as the real-life rescue challenges evoke more negative emotions than workshops (Strojny, Strojny, & Rębilas, 2017). Thus, the training using a professional simulator should not be perceived only as fun activity but probably could also be used to elicit negative emotions to mirror real-life emergency. Furthermore, research suggests that in computer-based learning environments frustration may lead to better learning outcomes than other low-arousal negative emotions such as boredom (e.g., Baker, D'Mello, Rodrigo, & Graesser, 2010).

The current findings suggest that a VR rescue simulator is not only engaging but it allows for emulating characteristics of the real-life incidents, i.e., frustration, mental demand, and time-pressure. On the other hand, we did not find differences between virtual rescue and control conditions for the perceived physical workload of the task. This could suggest that the VR application, or more precisely the application being currently tested, does not imitate the requirements of the real-life emergency situations regarding physical workload. However, we believe that even if VR technology might be useful for trainings focused on achieving goals of physical nature, such as rehabilitation (e.g., Bryanton et al., 2006) or anti-obesity interventions (Fung et al., 2006), the greatest potential of VR professional simulators lies in reflecting the *psychological* situation of a rescue challenge. This is because the physical aspects of a task could be more easily trained in live training. As it was suggested, psychological experience that is similar to the real life supports knowledge and skill transfer (Bacon, Windall, & MacKinnon, 2012; Romano & Brna, 2001). We, thus, believe that the future professional simulators should aim at recognizing the psychological states of the users and trying to imitate them as close as possible and be less concerned with the physical resemblance, e.g., weight of the objects. In other words, in our opinion the greatest potential of VR applications is in nurturing the *psychological fidelity,* which is the degree to which users subjectively perceive that a simulator reproduces real-life scenario (Dahl, Alsos, & Svanæs, 2010; Rehmann, 1995; *cf.* Hamstra, Brydges, Hatala, Zendejas, & Cook, 2014).

We also would like to address an important implication of our study. Namely, the feasibility of utilizing measures assessing activation of autonomic nervous system during task performance in VR applications and its consequences for designing human-computer interfaces in the future. As already mentioned, the methods used in the current study enables assessing both sympathetic and parasympathetic activation. Monitoring activation of autonomic nervous system on a basis of blood pressure or heart rate may be insufficient here because it is determined by an interplay of both sympathetic and parasympathetic influences (Berntson et al., 2007). For example, the relationship between heart rate and actual engagement might be complex as engagement at lower levels of difficulty may manifest in parasympathetic withdrawal, while sympathetic activation may occur only at higher levels of task difficulty (White & Raven, 2014; Van Roon et al., 2004). Assessing activity of both branches of autonomic nervous system to assess engagement has been already done (e.g., Richter, 2010; Silvia et al, 2014) but to our best knowledge it has never been simultaneously assessed in professional VR applications. We believe that such an approach provides more detailed and useful information for developing VR applications, especially the adaptive ones. In particular, it is a first step toward integration of physiological indices into interfaces adapting the course of the virtual experience to the needs of a particular user (Vaughan, Gabrys, & Dubey, 2016). Such an adaptation could involve dynamic adjusting a difficulty of a game (e.g., Hunicke & Chapman, 2004), manipulation of arousal that a user experiences to prevent too extreme states, similarly to adaptive automation applications (e.g., Pope, Bogart, & Bartolome, 1995; Schaefer, Haarmann, & Boucsein, 2008; Yamamoto & Isshiki, 1992) or balancing task characteristics to prevent simulation sickness (e.g., Cobb, Nichols, Ramsey, & Wilson, 1999; Murata, 2004). For example, if a user responds only with a parasympathetic withdrawal or low sympathetic activation, the app would choose a more challenging task.

The results of the current study are encouraging but we need to address several limitations. First, our sensitivity power analysis demonstrated that we were able to find statistically significant results for the effects of moderate to large magnitude. We did not have the sample size needed to detect small effects. Thus, our sample, ideally, would have been larger. Second, our sample was mainly composed of male participants. This reflects the composition of professional firefighters in Poland: women represent around 4% firefighter workforce, with this number be even lower for firefighters taking part in the everyday operations (Ministry of the Interior and Administration, 2014). Furthermore, as there is little research on the topic of gender differences in engagement in VR technology context, we refrain from speculating on the potential moderating effects of gender. This is also related to the broader question of generalizability of our findings across other dimensions of demographics and individual differences such as age, rank, experience, or personality traits of professional rescuers. We believe that further studies should address this question. Third, in the current study, we were focused on engagement, both subjective and objective. In the future studies, however, it would be crucial to include measures of performance in a simulated rescue operation. This is because research on the effectiveness of VR training show mixed results. Some of them suggest that training in VR increases performance (e.g., Dubovi et al., 2017; Gamberini et al., 2003). In contrast, other studies commands skepticism rather than enthusiasm regarding the usefulness of VR technology for educational purposes (e.g., Bliss, Tidwell, Guest, 1997; Dehn et al., 2018). Thus, the effectiveness of the training in the currently tested VR application need to be checked in the future studies. Fourth, it is important to note that although our measures of engagement provide information on sympathetic and parasympathetic activation, they reflect *overall* engagement. They do not allow to distinguish between particular psychological states, e.g., presence or frustration. Furthermore, we focused on the average engagement throughout the whole task period. However, it is possible that there are temporal dynamics of engagement during the course of a task. For example, users may initially engage but withdraw from a task once they realize that the challenge is too low or too high. Future studies might want to focus on the temporal aspects of task engagement.

4.1. Conclusions

The main goal of the current paper was to investigate the patterns of engagement during rescue operations performed in VR. We demonstrated that objective measurement of engagement during task performance in VR application is possible and we suggest that it is worth exploring in further research. In particular, we showed the importance of assessing both branches of autonomic nervous system, parasympathetic and sympathetic, to make inferences about engagement. This could be especially useful for future human-computer interfaces such as adaptive applications. Apart from that, we also stress the importance of focusing on psychological fidelity of the professional simulators to foster training and performance.

Finally, using VR technology might be especially useful to prepare professional rescuers to situations of low frequency but high risk for which the training is usually insufficient (Bertram et al., 2015; Heldal & Hammar Wijkmark, 2017; McDemott, 2017). Using VR applications, professional rescuers might train their skills, receive feedback, and repeat their actions in a safe environment. For example, such a training might be included in the firefighters' training curriculum for the novice officers. It would be especially useful as the vast majority of the existing VR applications focus on training management skills. On the other hand, we believe that VR applications cannot entirely replace the live and on-the-job trainings with real threats and real victims. Thus, VR technology can be a valuable addition to the regular training but it cannot replace it.

5. REFERENCES

Akhter, N., Gite, H., Tharewal, S. & Kale, K. (2015). Computer based RR-interval detection system with ectopy correction in HRV data. *Proceedings of International Conference on Advances in Computing, Communications and Informatics* (pp. 1613–1618). New York: ACM.

Austin, C.C., Dussault, G., & Ecobichon, D.J. (2001) Municipal firefighter exposure groups, time spent at fires and use of self-contained-breathing-apparatus. *American Journal of Industrial Medicine, 40*(6), 683–692

Backlund, P., Engstrom, H., Hammar, C., Johannesson, M. & Lebram, M. (2007). Sidh - a game-based firefighter training simulation. *Proceedings of International Conference on Information Visualization* (pp. 899–907). Los Alamitos, CA: IEEE.

Backlund, P., Engström, H., Gustavsson, M., Johannesson, M., Lebram, M. & Sjörs, E. (2009). Sidh: A game-based architecture for a training simulator. *International Journal of Computer Games Technology*, *2009*, 1–9.

Baker, R. Sj., D'Mello, S. K., Rodrigo, M. M. T. & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241.

Bakker, A. B., Vergel, A. I. S. & Kuntze, J. (2015). Student engagement and performance: A weekly diary study on the role of openness. *Motivation and Emotion*, *39*(1), 49–62.

Beaubien, J. M., & Baker, D. P. (2004). The use of simulation for training teamwork skills in health care: how low can you go?. *BMJ Quality & Safety*, *13*, i51-i56.

Berntson, G. G., Cacioppo, J. T., & Quigley, K. S. (1991). Autonomic determinism: The modes of autonomic control, the doctrine of autonomic space, and the laws of autonomic constraint. *Psychological Review, 98*, 459-487

Berntson, G. G., Lozano, D. L., Chen, Y. J. & Cacioppo, J. T. (2004). Where to Q in PEP?. *Psychophysiology*, *41*(2), 333–337.

Berntson, G. G., Quigley, K. S. & Lozano, D. (2007). Cardiovascular psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Ed.), *Handbook of Psychophysiology* (pp. 182–210). Cambridge, UK: Cambridge University Press.

Bertram, J., Moskaliuk, J., & Cress, U. (2015). Virtual training: Making reality work?. *Computers in Human Behavior*, *43*, 284-292.

Bliss, J. P., Tidwell, P. D. & Guest, M. A. (1997). The effectiveness of virtual reality for administering spatial navigation training to firefighters. *Presence: Teleoperators & Virtual Environments*, *6*(1), 73–86.

Bos, J., Mol, E., Visser, B., & Frings-Dresen, M. (2004). Risk of health complaints and disabilities among Dutch firefighters. *International Archives of Occupational and Environmental Health*, 77(6), 373–382.

Boulet, J. R., Murray, D., Kras, J., Woodhouse, J., McAllister, J., & Ziv, A. (2003). Reliability and validity of a simulation-based acute care skills assessment for medical students and residents. *Anesthesiology: The Journal of the American Society of Anesthesiologists*, *99*(6), 1270-1280.

Bric, J. D., Lumbard, D. C., Frelich, M. J., & Gould, J. C. (2016). Current state of virtual reality simulation in robotic surgery training: a review. *Surgical Endoscopy*, *30*(6), 2169-2178.

Brehm, J. W. & Self, E. A. (1989). The intensity of motivation. *Annual Review of Psychology*, *40*(1), 109–131.

Bryanton, C., Bossé, J., Brien, M., Mclean, J., McCormick, A., & Sveistrup, H. (2006). Feasibility, Motivation, and Selective Motor Control: Virtual Reality Compared to Conventional Home Exercise in Children with Cerebral Palsy. *CyberPsychology & Behavior, 9*(2), 123-128.

Cobb, S. V., Nichols, S., Ramsey, A. & Wilson, J. R. (1999). Virtual reality-induced symptoms and effects (VRISE). *Presence: Teleoperators & Virtual Environments, 8*(2), 169–186.

Crescentini, C., Chittaro, L., Capurso, V., Sioni, R. & Fabbro, F. (2016). Psychological and physiological responses to stressful situations in immersive virtual reality: Differences between users who practice mindfulness meditation and controls. *Computers in Human Behavior*, *59*, 304–316.

Dahl, Y., Alsos, O. A., & Svanæs, D. (2010). Fidelity considerations for simulation-based usability assessments of mobile ICT for hospitals. International. *Journal of Human–Computer Interaction*, 26(5), 445-476.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science*, *323*(5910), 66–69.

Dehn, L. B., Kater, L., Piefke, M., Botsch, M., Driessen, M. & Beblo, T. (2018). Training in a comprehensive everyday-like virtual reality environment compared to computerized cognitive training for patients with depression. *Computers in Human Behavior*, *79*, 40–52.

Dimakis, N., Filippoupolitis, A., & Gelenbe, E. (2010). Distributed building evacuation simulator for smart emergency management. *The Computer Journal*, *53*(9), 1384-1400.

Dorsey, D., Campbell, G., & Russell, S. (2009). Adopting the instructional science paradigm to encompass training in virtual environments. *Theoretical Issues in Ergonomics Science*, *10*(3), 197-215.

Dubovi, I., Levy, S. T., & Dagan, E. (2017). Now I know how! The learning process of medication administration among nursing students with non-immersive desktop virtual reality simulation. *Computers & Education*, *113*, 16-27.

Egan, D., Brennan, S., Barrett, J., Qiao, Y., Timmerer, C., & Murray, N. (2016). An evaluation of Heart Rate and ElectroDermal Activity as an objective QoE evaluation method for immersive virtual reality environments. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)* (pp. 1-6). IEEE.

Evalt, D. M. (2017, March 1). Oculus Slashes Prices On Rift Hardware. *Forbes.* Retrieved from: https://www.forbes.com/sites/davidewalt/2017/03/01/oculus-slashes-prices-on-rift-hardware/#35945c6a53af

Faul, F., Erdfelder, E., Buchner, A. & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–60. doi:10.3758/BRM.41.4.1149

Field, A. (2013). Discovering statistics using IBM SPSS. Thousand Oaks, CA: SAGE Publishing Inc.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, CA: SAGE Publishing Inc.

Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*, 59–109

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Fung, J., Richards, C. L., Malouin, F., McFadyen, B. J., & Lamontagne A. (2006). A Treadmill and Motion Coupled Virtual Reality System for Gait Training Post-Stroke. *CyberPsychology & Behavior 9*(2), 157-162.

Furnham, A., & Henderson, M. (1982). The good, the bad and the mad: Response bias in self-report measures. *Personality and Individual Differences*, *3*(3), 311-320.

Gallagher, A. G., Ritter, E. M., Champion, H., Higgins, G., Fried, M. P., Moses, G., ... & Satava, R. M. (2005). Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of Surgery*, *241*(2), 364-372.

Gamberini, L., Cottone, P., Spagnolli, A., Varotto, D. & Mantovani, G. (2003). Responding to a fire emergency in a virtual environment: different patterns of action for different situations. *Ergonomics*, *46*(8), 842–858.

Gardner, A. K., Abdelfattah, K., Wiersch, J., Ahmed, R. A., & Willis, R. E. (2015). Embracing errors in simulation-based training: the effect of error training on retention and transfer of central venous catheter skills. *Journal of Surgical Education*, *72*(6), e158-e162.

Gatto, L. B. S., Mól, A. C. A., dos Santos, I. J. L., Jorge, C. A. F., & Legey, A. P. (2013). Virtual simulation of a nuclear power plant's control room as a tool for ergonomic evaluation. *Progress in Nuclear Energy*, *64*, 8-15.

Gorini, A., Capideville, C. S., De Leo, G., Mantovani, F., & Riva, G. (2011). The role of immersion and narrative in mediated presence: the virtual hospital experience. *Cyberpsychology, Behavior, and Social Networking*, *14*(3), 99-105.

Guidotti, T.L. (1992). Human factors in firefighting: ergonomic-, cardiopulmonary-, and psychogenic stress-related issues. *International Archives of Occupational and Environmental Health,* 64 (1), 1–12.

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine*, *89*(3), 387-392.

Harper, K. L., Eddington, K. M. & Silvia, P. J. (2016). Perfectionism and Effort-Related Cardiac Activity: Do Perfectionists Try Harder? *PloS One*, *11*(8), e0160340.

Harrald, J. R. (2006). Agility and discipline: Critical success factors for disaster response. *The Annals of the American Academy of Political and Social Science*, *604*(1), 256-272.

Harris, K. & Reid, D. (2005). The influence of virtual reality play on children's motivation. *Canadian Journal of Occupational Therapy*, *72*(1), 21–29.

Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139–183). Amsterdam: North Holland.

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators & Virtual Environments*, *1*(2), 262-271.

Heinrichs, W. L., Youngblood, P., Harter, P. M., & Dev, P. (2008). Simulation for team training and assessment: case studies of online training with virtual worlds. *World Journal of Surgery*, *32*(2), 161-170.

Heldal, I. & Hammar Wijkmark, C. H. (2017). Simulations and Serious Games for Firefighter Training: Users' Perspective. *Proceedings of the ISCRAM Conference* (pp. 868 – 878). Albi, France: ISCRAM.

Holmer, I., & Gavhed, D. (2007). Classification of metabolic and respiratory demands in fire fighting activity with extreme workloads. *Applied Ergonomics, 38*(1), 45-52.

Huang, H. M., Rauch, U., & Liaw, S. S. (2010). Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Computers & Education*, *55*(3), 1171-1182.

Hunicke, R., & Chapman, V. (2004). AI for Dynamic Difficulty Adjustment in Games. *Proceedings of Challenges in Game Artificial Intelligence AAAI Workshop* (pp. 91–96). Menlo Park, CA: AAAI Press.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: a meta-analysis. *Journal of Applied Psychology*, *93*(1), 59-69.

Kelsey, R. M. (2012). Beta-adrenergic cardiovascular reactivity and adaptation to stress: The cardiac pre-ejection period as an index of effort.

Kenny, P., & Parsons, T.D. (2011). Embodied conversational virtual human patients. In *Conversational Agents and Natural Language Interaction: Techniques and Effective Practices*. In C. Perez-Marin & I. Pascual-Nieto (Eds.), *pp.* 254–281. Hershey, PA: IGI Global.

Kothgassner, O. D., Felnhofer, A., Hlavacs, H., Beutl, L., Palme, R., Kryspin-Exner, I. & Glenk, L. M. (2016). Salivary cortisol and cardiovascular reactivity to a public speaking task in a virtual and real-life environment. *Computers in Human Behavior*, *62*, 124–135.

Kuipers, M., Richter, M., Scheepers, D., Immink, M. A., Sjak-Shie, E. & Van Steenbergen, H. (2017). How effortful is cognitive control? Insights from a novel method measuring single-trial evoked beta-adrenergic cardiac reactivity. *International Journal of Psychophysiology*, *119*, 87–92.

Lam, C. K., Sundaraj, K., & Sulaiman, M. N. (2013). A review of computer-generated simulation in the pedagogy of cataract surgery training and assessment. *International Journal of Human-Computer Interaction*, *29*(10), 661-669.

Lippman, N. E. A. L., Stein, K. M., & Lerman, B. B. (1994). Comparison of methods for removal of ectopy in measurement of heart rate variability. *American Journal of Physiology-Heart and Circulatory Physiology*, *267*(1), H411-H418.

Llabre, M. M., Spitzer, S. B., Saab, P. G., Ironson, G. H., & Schneiderman, N. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. *Psychophysiology*, *28*(6), 701-711.

Lozano, D. L., Norman, G., Knox, D., Wood, B. L., Miller, B. D., Emery, C. F. & Berntson, G. G. (2007). Where to B in dZ/dt?. *Psychophysiology*, *44*(1), 113–119.

Lusa, S., Louhevaara, V., & Kinnunen, K. (1994). Are the job demands on physical work capacity equal for young and aging firefighters?. *Journal of Occupational Medicine*, *36*(1), 70-74.

Martin-Niedecken, A. L. & Götz, U. (2017). Go with the Dual Flow: Evaluating the Psychophysiological Adaptive Fitness Game Environment "Plunder Planet." In *Joint International Conference on Serious Games* (pp. 32–43). Cham, Switzerland: Springer.

McDevitt, D. M. (2017). *Searching for effective training solutions for firefighting: the analysis of emergency responses and line of duty death reports for low frequency, high risk events*.

McKibben, W. B., & Silvia, P. J. (2017). Evaluating the distorting effects of inattentive responding and social desirability on self-report scales in creativity and the arts. *The Journal of Creative Behavior*, *51*(1), 57-69.

Meehan, M., Razzaque, S., Insko, B., Whitton, M., & Brooks, F. P. (2005). Review of four studies on the use of physiological reaction as a measure of presence in stressful virtual environments. *Applied Psychophysiology and Biofeedback*, *30*(3), 239-258.

Meissner, A., Luckenbach, T., Risse, T., Kirste, T. & Kirchner, H. (2002). Design challenges for an integrated disaster management communication and information system. *Proceeding of The First IEEE Workshop on Disaster Recovery Networks*. Los Alamitos, CA: IEEE.

Ministry of the Interior and Administration (2014, March 8). Kobiety w służbach MSWiA [*Women in service at Ministry of the Interior and Administration*]. Retrieved from: http://www.mswia.gov.pl/pl/aktualnosci/11877,Kobiety-w-sluzbach-MSW.html

Mineo, B. A., Ziegler, W., Gill, S. & Salkin, D. (2009). Engagement with electronic screen media among students with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(1), 172–87. doi:10.1007/s10803-008-0616-0

Murata, A. (2004). Effects of duration of immersion in a virtual reality environment on postural stability. *International Journal of Human-Computer Interaction*, *17*(4), 463-477.

Muraven, M., Tice, D. M. & Baumeister, R. F. (1998). Self-control as a limited resource: Regulatory depletion patterns. *Journal of Personality and Social Psychology*, *74*(3), 774–789.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231-259. doi: 10.1037/0033-295X.84.3.231

Obrist, P. A. (1976). The cardiovascular-behavioral interaction—As it appears today. *Psychophysiology*, *13*(2), 95–107.

Obrist, P. A. (1981). *Cardiovascular psychophysiology: A perspective*. Plenum: New York, NY.

Park, C. H., Jang, G., & Chai, Y. H. (2006). Development of a virtual reality training system for live-line workers. *International Journal of Human-Computer Interaction*, *20*(3), 285-303.

Parsons, T. D., Rizzo, A. A., Courtney, C. G., & Dawson, M. E. (2012). Psychophysiology to assess impact of varying levels of simulation fidelity in a threat environment. *Advances in Human-Computer Interaction*, *2012*, ID 831959.

Paulhus, D. L. & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, *66*(6), 1025–1060.

Pope, A. T., Bogart, E. H., & Bartolome, D. S. (1995). Biocybernetic system evaluates indices of operator engagement in automated task. *Biological Psychology, 40*(1-2), 187-195.

Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional Science*, *23*(5-6), 405–431.

Rehmann A. (1995). *A Handbook of Flight Simulation Fidelity Requirements for Human Factors Research. Technical Report No. DOT/FAA/CT-TN95/46.* Wright-Patterson AFB, OH: Crew Systems Ergonomics Information Analysis Center.
Richter, M. (2013). A closer look into the multi-layer structure of motivational intensity theory. Social and Personality Psychology Compass, 7(1), 1–12. doi:10.1111/spc3.12007.
Richter, M., Gendolla, G. H. E., & Wright, R. A. (2016). Three decades of research on motivational intensity theory: What we have learned about effort and what we still don't know. In A. J. Elliot (Ed.), *Advances in motivation science: Volume 3,* pp. 149-186). San Diego, CA: Elsevier Academic Press.
Romano, D. M., & Brna, P. (2001). Presence and reflection in training: Support for learning to improve quality decision-making skills under time limitations. *CyberPsychology & Behavior*, *4*(2), 265-277.
Rowell, L. B. (1993). *Human cardiovascular control*. New York, NY: Oxford University Press.
Rueda, R., O'Neil, H. F., & Son, E. (2016). The Role of Motivation, Affect, and Engagement in Simulation/Game Environment: Proposed Model. In H. F. O'Neil, E. Baker, & R. S. Perez (Eds), *Using games and simulations for teaching and assessment: Key issues.* New York, NY: Routlege.
Schaefer, F., Haarmann, A. & Boucsein, W. (2008). The usability of cardiovascular and electrodermal measures for adaptive automation. In J. H. D. M. Westerink, M. Ouwerkerk, T. J. M. Overbeek, W. F. Pasveer, & B. de Ruyter (Ed.), *Probing Experience* (pp. 235–243). Dordrecht, The Netherlands: Springer.
Schutte, N. S., & Stilinovic, E. J. (2017). Facilitating empathy through virtual reality. *Motivation and Emotion, 41*, 708-712. doi: 10.1007/s11031-017-9641-7
Silvia, P. J., Beaty, R. E., Nusbaum, E. C., Eddington, K. M., & Kwapil, T. R. (2014). Creative motivation: Creative achievement predicts cardiac autonomic markers of effort during divergent thinking. *Biological Psychology*, *102*, 30-37.
Shamim Khan, M., Ahmed, K., Gavazzi, A., Gohil, R., Thomas, L., Poulsen, J., … Dasgupta, P. (2013). Development and implementation of centralized simulation training: evaluation of feasibility, acceptability and construct validity. *BJU International*, *111*(3), 518–523.
Sherwood, A., Allen, M., Fahrenberg, J., Kelsey, R., Lovallo, W. & others. (1990). Methodological guidelines for impedance cardiography. *Psychophysiology*, *27*(1), 1–23.
Singh, D., Vinod, K. & Saxena, S. (2004). Sampling frequency of the RR interval time series for spectral analysis of heart rate variability. *Journal of Medical Engineering & Technology*, *28*(6), 263–272.
Smith, B. T. & Hess, T. M. (2015). The impact of motivation and task difficulty on resource engagement: Differential influences on cardiovascular responses of young and older adults. *Motivation Science*, *1*(1), 1–15.
Smith, D. L., Manning, T. S., & Petruzzello, S. J. (2001). Effect of strenuous live-fire drills on cardiovascular and psychological responses of recruit firefighters. *Ergonomics*, *44*(3), 244-254.
Smith, D., Petruzzello, S., Kramer, J., Warner, S., Bone, B. & Misner, J. (1995). Selected physiological and psychobiological responses to physical activity in different configurations of firefighting gear. *Ergonomics, 38*(10), 2065–2077.
St Julien, T. U. & Shaw, C. D. (2003). Firefighter command training virtual environment. In *Proceedings of the 2003 Conference on Diversity in Computing* (pp. 30–33). New York: ACM.
Strojny, P., Strojny, A., Rębilas, K. (2017, September). *Męcząca zabawa: wpływ trudności gry komputerowej na wysiłek w nią wkładany [Exhausting game: the impact of difficulty of a computer game on effort mobilization]*. Paper presented at The 14th Annual Convention of Polish Society of Social Psychology, Toruń, Poland.
Task Force (1996). Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use. *Circulation*, *93*(5), 1043–1065.
Thuraisingham, R. (2006). Preprocessing RR-interval time series for heart rate variability analysis and estimates of standard deviation of RR intervals. *Computer Methods and Programs in Biomedicine*, *83*(1), 78–82.
Van Roon, A. M., Mulder, L. J., Althaus, M. & Mulder, G. (2004). Introducing a baroreflex model for studying cardiovascular effects of mental workload. *Psychophysiology*, *41*(6), 961–981.

Vaughan, N., Gabrys, B., & Dubey, V. N. (2016). An overview of self-adaptive technologies within virtual reality training. *Computer Science Review*, *22*, 65-87.

White, D. W. & Raven, P. B. (2014). Autonomic neural control of heart rate during dynamic exercise: revisited. *The Journal of Physiology*, *592*(12), 2491–2500.

Williams-Bell, F. M., Kapralos, B., Hogue, A., Murphy, B. & Weckman, E. (2015). Using serious games and virtual simulation for training in the fire service: a review. *Fire Technology*, *51*(3), 553–584.

Wright, R. A. (1996). Brehm's theory of motivation as a model of effort and cardiovascular response.

Yamamoto, Y. & Isshiki, H. (1992). Instrument for controlling drowsiness using galvanic skin reflex. *Medical & Biological Engineering & Computing, 30*(5), 562–564.

Zieliński, P. & Biernacki, M. (2010). Analiza psychometryczna polskiego przekładu narzędzia do subiektywnej oceny obciążenia zadaniowego NASA-TLX [Psychometric analysis of polish version of a tool assessing subjective task workload NASA-TLX]. *Polski Przegląd Medycyny Lotniczej*, *16*(3), 219–239.

SUPPLEMENTAL MATERIALS

Affect

Participants' affective states were assessed with The Self-Assessment Manikins (Bradley & Lang, 1994). In the Self-Assessment Manikins participants assessed their affect on the dimensions of valence, arousal, and dominance. Participants were presented with three series of pictorial representations of the affective states varying in valence, intensity, and dominance. For each dimension, participants were asked to choose a picture that represent their current state. We used modified version of a task (Irtel, 2007) in which there were 9 pictures for each dimension (instead of 5). Results are presented in Figure S1.

In order to identify emotions of subjects, we used Scale of emotions and Scale of General Mood (Wojciszke, & Baryła, 2005). In the first questionnaire participants answered a question regarding intensity of experienced emotions on 7-point scale (1 = "not at all" to 7 = "extremely intensive"). The list contains 24 single-word names of emotions - four per each of six basic emotions. We also used Scale of general mood in which participants assessed their mood by rating of 10 statements on 5-point Likert scale (1 = "I disagree" to 5 = "I agree"). Results from both scales are presented in Figure S1.

In order to identify the level and nature of stress experienced during simulation, Stress Appraisal Questionnaire (Włodarczyk & Wrześniewski, 2010) was used. It contains two forms - dispositional and situational stress assessment, we used only situational part due to our research questions. It contains the set of 35 adjectival expressions used for describing stressful situations. Subjects respond on 4-point Likert scale (0 = "definitely not" to 3 = "definitely yes"). There were 4 subscales, such as threat (e.g., "terrifying", 9 items, Cronbach's α = .92); challenge - activity (e.g., "mobilizing", 5 items, Cronbach's α = .49); challenge - passivity (e.g., "interesting", 5 items, Cronbach's α = .70); harm / loss (e.g., "unjust", 4 items, Cronbach's α = .76). Results are presented in Figure S1.

Participants' subjective self-efficacy was assessed with General Self-Efficacy scale (Schwarzer, Jerusalem, & Juczynski, 2001) originally aimed to assess dispositional self-efficacy, scale was adapted to the need of momentary self-efficacy assessment. It contains 10 statements (e.g., "It is easy for me to stick to my aims and accomplish my goals", Cronbach's α = .89), to which participants responded on four-point scale (1 = "no" to 4 = "yes"). Results are presented in Figure S1.

Figure S1

Affect ratings

Measure		Condition	Mean	SD
Manikin	Valence [Pre]	Control	7.828	1.104
		Rescue challenge	7.300	1.236
	Arousal [Pre]	Control	4.655	1.446
		Rescue challenge	4.500	1.889
	Dominance [Pre]	Control	6.207	1.048
		Rescue challenge	5.233	1.547
	Valence [Post]	Control	7.379	1.049
		Rescue challenge	6.633	1.351
	Arousal [Post]	Control	4.724	1.688
		Rescue challenge	5.100	1.936
	Dominance [Post]	Control	5.793	1.590
		Rescue challenge	5.733	1.617
	Valence [Post-Pre]	Control	-0.448	0.870
		Rescue challenge	-0.667	1.184
	Arousal [Post-Pre]	Control	0.069	1.163
		Rescue challenge	0.600	0.932
	Dominance [Post-Pre]	Control	-0.414	1.680
		Rescue challenge	0.500	1.526
Emotions	Happiness [Pre]	Control	5.060	0.731
		Rescue challenge	4.925	0.711
	Love [Pre]	Control	4.302	1.107
		Rescue challenge	4.258	0.911
	Anxiety [Pre]	Control	2.431	0.961
		Rescue challenge	2.392	0.944
	Anger [Pre]	Control	2.664	0.987
		Rescue challenge	2.875	1.196
	Guilt [Pre]	Control	1.690	0.664
		Rescue challenge	1.925	0.786
	Sadness [Pre]	Control	1.698	0.745
		Rescue challenge	1.750	0.685
	Happiness [Post]	Control	3.526	0.955
		Rescue challenge	3.558	1.188
	Love [Post]	Control	1.991	1.064
		Rescue challenge	2.625	1.102
	Anxiety [Post]	Control	2.086	0.931
		Rescue challenge	2.525	1.343
	Anger [Post]	Control	1.353	0.600
		Rescue challenge	1.733	0.940
	Guilt [Post]	Control	1.198	0.599
		Rescue challenge	1.550	0.862
	Sadness [Post]	Control	1.198	0.580
		Rescue challenge	1.508	0.850
Mood	Mood [Pre]	Control	4.393	0.660
		Rescue challenge	4.427	0.558
	Mood [Post]	Control	4.445	0.575
		Rescue challenge	4.227	0.758
Stress	Threat	Control	0.205	0.354

		Rescue challenge	0.447	0.591
	Challenge - active	Control	1.705	0.559
		Rescue challenge	1.753	0.438
	Challenge - passive	Control	2.200	0.563
		Rescue challenge	2.037	0.522
	Harm/loss	Control	0.086	0.224
		Rescue challenge	0.192	0.358
General Self-Efficacy		Control	3.348	0.384
		Rescue challenge	3.150	0.434

Subjective Engagement and Workload

In Two-dimensional Effort to Difficulty Ratio (EtoD-2D; A. Strojny, Rębilas, P. Strojny, 2017), participants were asked to rate the perception of task difficulty and effort exerted during a task. They gave their response on a 11 by 11 matrix. The x-axis referred to task difficulty: the scale ranges from 0 (“*The task goal was easy to accomplish”*) to 10 (“*The task goal was nearly impossible to accomplish”*). The y-axis referred to subjective effort: the scale ranges from 0 (“*I did not exert any effort”*) to 10 (“*I exerted a lot of effort”*). Participants were asked to mark a point corresponding to their assessment of difficulty and effort. Results are presented in Figure S2.

In NASA-TLX (Hart & Staveland, 1988; Polish version by Zieliński, & Biernacki, 2010), apart from rating scales, participants responded in the second part of a tool in which they indicated which aspect of the workload was more strongly felt during the task. They were given 15 pairs of workload aspects to compare. For example, they were asked if mental vs. physical demands or time pressure vs. frustration were more strongly felt during task performance. The subjective workload rating are calculated by weighting the workload ratings from a first part of NASA-TLX by the number of times a particular aspect was chosen in a second part. Finally, this weighted score is divided by 15. We present the average weight for each aspect of workload, as well as the weighted scores in Figure S2 (the non-weighted scores are presented in the main text).

Figure S2

Subjective effort

Measure		Condition	Mean	SD
EtoD-2D	Effort	Control	1.865	1.978
		Rescue challenge	2.983	1.634
	Difficulty	Control	1.596	1.077
		Rescue challenge	4.259	2.617
NASA-TLX	Mental Workload [rating]	Control	4.069	0.998
		Rescue challenge	3.207	0.774
		Control	1.931	1.387

Physical Workload [rating]	Rescue challenge	1.103	0.939
Time Pressure [rating]	Control	2.310	1.228
	Rescue challenge	4.034	1.017
Performance [rating]	Control	4.069	1.193
	Rescue challenge	3.724	1.251
Effort [rating]	Control	1.345	0.936
	Rescue challenge	1.138	0.875
Frustration [rating]	Control	1.276	1.601
	Rescue challenge	1.724	1.811
Average rating	Control	8.345	6.115
	Rescue challenge	10.122	5.915
Mental Workload [weight]	Control	2.264	2.843
	Rescue challenge	1.489	1.957
Physical Workload [weight]	Control	3.207	4.172
	Rescue challenge	16.900	9.531
Time Pressure [weight]	Control	4.333	4.097
	Rescue challenge	11.400	5.900
Performance [weight]	Control	1.828	2.064
	Rescue challenge	3.100	3.964
Effort [weight]	Control	1.724	2.507
	Rescue challenge	4.722	6.159
Frustration [weight]	Control	1.865	1.978
	Rescue challenge	2.983	1.634
Mental Workload [score]	Control	1.596	1.077
	Rescue challenge	4.259	2.617
Physical Workload [score]	Control	4.069	0.998
	Rescue challenge	3.207	0.774
Time Pressure [score]	Control	1.931	1.387
	Rescue challenge	1.103	0.939
Performance [score]	Control	2.310	1.228
	Rescue challenge	4.034	1.017
Effort [score]	Control	4.069	1.193
	Rescue challenge	3.724	1.251
Frustration [score]	Control	1.345	0.936
	Rescue challenge	1.138	0.875

Immersion and Realism

We measured participants' sense of presence in VR settings with Igroup Presence Questionnaire, (IPQ; Schubert, Friedmann, & Regenbrecht, 2001, Polish version of Lipp, A. Strojny, & P. Strojny, in preparation). IPQ is a 14-item measure in which participants rate the degree to which they agree with a presented item using a 7-point Likert scale (-3 = "disagree" to 3 = "agree"). This scale consists of three subscales: spatial presence, which is defined as a sense of being in computer-generated environment (e.g. "I felt present in the virtual space"; 6 items, Cronbach's α = .68), involvement which is defined as the amount of attention devoted to virtual reality (e.g. "I was completely captivated by the virtual world"; 4 items, Cronbach's α = .77), and realism which is subjective experience of how realistic virtual environment is (e.g. "Virtual environment seemed absolutely realistic to me"; 4 items, Cronbach's α = .65). Results are presented in Figure S3.

We used Polish version of the German VR Simulation Realism Scale (Poeschl & Doering, 2014, professional Polish translation). The scale contains 14 statements regarding Scenic realism (e.g., "Reflection in virtual space seemed to be natural.", 5 items, Cronbach's α = .75), Audience behavior (e.g., "Gestures of virtual humans was natural.", 4 items, Cronbach's α = .85), Audience appearance (e.g., "Outfit of virtual humans was adequate.", 4 items, Cronbach's α = .74), and Sound realism ("Ambience sound intensity in the virtual room was … (1 = too low to 5 = too loud)", 1 item). Participants use 5-point Likert type scale for answering (-2 = "I totally disagree" to 2 = "I totally agree"). Results are presented in Figure S3.

Figure S3

Ratings of immersion and realism

Measure		Condition	Mean	SD
IPQ	Spatial Presence	Control	1.546	0.804
		Rescue challenge	1.567	0.856
	Involvement	Control	-0.862	1.300
		Rescue challenge	0.125	1.389
	Realism	Control	-0.224	1.018
		Rescue challenge	-0.242	1.056
Realism	Scenic realism	Control	0.855	0.628
		Rescue challenge	0.573	0.584
	Audience behavior	Control	0.724	0.757
		Rescue challenge	0.167	0.789
	Audience appearance	Control	0.534	0.664
		Rescue challenge	0.800	0.631
	Sound realism	Control	0.069	0.593
		Rescue challenge	-0.367	0.809

References for Supplemental materials

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49-59. doi: 10.1016/0005-7916(94)90063-9

Hart S.G., & Staveland L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. [W]: P. A. Hancock, N. Meshkati (Eds.): *Human Mental Workload*. Amsterdam (The Netherlands): North Holland Press.

Irtel, H. (2007). PXLab: The Psychological Experiments Laboratory [online]. Version 2.1.11. Mannheim (Germany): University of Mannheim. Available at http://www.pxlab.de

Lipp, N., Strojny A., & Strojny P. (2017). *Polish adaptation of Ingroup Presence Questionnaire*. Manuscript in preparation.

Poeschl, S., & Doering, N. (2013). The German VR Simulation Realism Scale-Psychometric Construction for Virtual Reality Applications with Virtual Humans. *Annual Review of Cybertherapy and Telemedicine*, *11*, 33-37.

Schwarzer R., Jerusalem M., & Juczyński, Z. (2001) Skala Uogólnionej Własnej Skuteczności GSES [*Self-Efficacy Scale*]. W: Z. Juczyński (Ed.) *Narzędzia pomiaru w promocji zdrowia i psychologii zdrowia* [*Measurement tool in health promotion and health psychology*]. Warsaw (Poland): Pracownia Testów Psychologicznych PTP.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The Experience of Presence: Factor Analytic Insights. *Presence: Teleoperators and Virtual Environments,* *10*(3), 266-281. doi:10.1162/105474601300343603

Strojny, P., Rębilas, K., Strojny, A. (2017). *EtoD-2D. Two-dimensional Assessment of Effort to Difficulty Ratio*. Manuscript in preparation.

Wojciszke, B., Baryła, W. (2005). Skale do pomiaru nastroju i sześciu emocji [Scale for the measurement of mood and six emotions]. *Czasopismo Psychologiczne, 11*, 31-47.

Włodarczyk, D., Wrześniewski, K. (2010). Kwestionariusz Oceny Stresu KOS [The stress appraisal questionnaire]. *Przegląd Psychologiczny, 53(4)*, 479-496.

Zieliński, P., Biernacki, M. (2010). *Analiza psychometryczna polskiego przekładu narzędzia do subiektywnej oceny obciążenia zadaniowego NASA-TLX* [Psychometric Evaluation of NASA-TLX Polish Translation]. Polski Przegląd Medycyny Lotniczej, 16(3), 219 - 239.

Notes on contributors

Gabriela Czarnek (gczarnek@nano-games.com, http://cscs.edu.pl/gabriela-czarnek/) is a social psychologist with an interest in effort mobilization in applied settings; she is a researcher at Nano Games and at the Institute of Psychology of Jagiellonian University.

Paweł Strojny (pstrojny@nano-games.com, http://www.ips.uj.edu.pl/dr_pawel_strojny) is a social psychologist interested in motivation toward virtual experiences; He is head of R&D at Nano Games and assistant professor at the Institute of Applied Psychology of Jagiellonian University.

Agnieszka Strojny (agnieszka.strojny@uj.edu.pl, http://www.ips.uj.edu.pl/dr-Agnieszka-Strojny) is a social psychologist with an interest in video games engagement and learning effectiveness in VR simulators; She is a researcher at Nano Games and assistant professor at the Institute of Applied Psychology of Jagiellonian University.

Michael Richter (M.Richter@ljmu.ac.uk, http://www.effortlab.website) is a psychophysiologist with an interest in the determinants and consequences of mental and physical effort; He is a Reader in Motivation Psychology at the School of Natural Sciences and Psychology of Liverpool John Moores University.