Complementing German Vocational Training with VR Training: Investigating the Effect of Time Pressure and Stress on Situation Awareness and Performance of Construction Trainees

Master Thesis

Inka J. M. Raestrup

Educational Science and Technology,

Faculty of Behavioural, Management and Social Sciences,

University of Twente

Supervisors:

Dr. Ilona Friso-van den Bos

Dr. Bas Kolloffel

Keywords: virtual reality, construction, asphalting, trainees, time pressure, subjective stress, physiological stress, situation awareness, performance

Word Count: 16.894

April 10, 2023

Acknowledgement

I would like to take the opportunity to thank everyone who has supported me throughout the process of writing my master thesis. First and foremost, I would like to thank my first supervisor, Dr. Ilona Friso-van den Bos, for her continuous support and valuable feedback. Her insightful comments and critical questions in all phases of my thesis helped me greatly in shaping the direction and improving the quality of my work. Second, I would like to thank my second supervisor, Dr. Bas Kolloffel, for his helpful feedback in the final stage of my thesis. Third, I would like to thank the trainees who participated in this study and their subject teachers at the vocational school and the inter-company training centre for their openness towards the project and for their time and support, especially in translating the simulation into German. Forth, I would like to thank the BMS Lab, in particular Adrian Fagarasanu, for his commitment and invaluable contribution to this project, since he not only made necessary modifications to the VR simulation but also resolved technical issues timely and implemented further, helpful applications in the VR simulation. In this context, I would also like to thank Lucia Rabago Mayer for her critical and insightful comments on the development of an informative storyboard to ensure a smooth implementation of the changes. Last but certainly not least, I would like to thank my family and friends for their ongoing support and encouragement, especially Jonas, Yola and Rengin.

Abstract

To make appropriate decisions and perform work safely and efficiently, construction trainees benefit from a high level of situation awareness (SA), a tacit knowledge generally acquired through repeated on-the-job experiences. Virtual reality (VR) training provides a safe frame to train procedures and acquire SA while being exposed to work-related stressors such as time pressure (TP). In this context, the present study investigated the effect of TP and stress (subjective and physiological) on SA and performance of construction trainees in a VR training. For this end, a within-subjects design was utilised. 53 construction trainees participated in the VR experiment and immersed into a low and a high time pressure condition (TPC) in which they performed the same construction task. Subjective stress was measured via a questionnaire. Participants' skin conductance served as indicator for physiological stress, measured with the Empatica E4 wristband. Participants' SA was assessed using the SAGAT method and their task performance was evaluated based on log-files from the VR simulation.

A stepwise approach was utilised to investigate (1) differences in subjective stress, physiological stress, SA and performance between conditions, (2) the role of subjective and physiological stress for differences in SA and performance between conditions and (3) the effect of subjective and physiological stress on SA and performance within conditions. (1) Paired samples t-tests found significantly higher levels of subjective stress in the high TPC. The same applies to SA and performance, although higher levels were expected for the low TPC. (2) Results of rmANCOVAs could not demonstrate that differences in SA and performance depend on differences in subjective or physiological stress. (3) Multiple linear regression analyses did not find a significant effect of subjective and physiological stress on SA or performance in either TPC.

Even though the study has largely not shown the expected effects, the results indicate that the imposition of acute TP in VR training affects stress-levels and outcome variables. As a limitation, it should be noted that simplified situational demands in the VR training reduced the demands-ability mismatch generally expected for trainees in construction. As a theoretical implication, the results underline the importance of including multiple stress indicators when examining the effects of TP and stress in VR. On a practical level, the findings suggest that imposing stressors in VR training represents not only a safe but also a motivating approach for experiential learning. Future research should build on this study by using integrated model testing, including the level of experience and prior knowledge as covariates and incorporating objective measures for SA.

Keywords: virtual reality, construction, asphalting, trainees, time pressure, subjective stress, physiological stress, situation awareness, performance

Table of Contents

Abbreviations

EDA	Electrodermal activity
Empatica	Empatica E4 wristband
HMD	Head-mounted display
Phystress	Physiological stress
PSQ	Perceived stress questionnaire
SA	Situation awareness
SAGAT	Situation awareness global assessment technique
SCL	Skin conductance level
Substress	Subjective stress
TP	Time pressure
TPC	Time pressure condition
VET system	Vocational education and training system
VR	Virtual reality

1. Complementing German Vocational Training with VR Training: Investigating the Effect of Time Pressure and Stress on Situation Awareness and Performance of Construction Trainees

The German vocational education and training system (VET system) is globally renowned and highly praised (German Federal Ministry for Education and Reserach [BMBF], n.d.). Its success lies primarily in the dual approach which includes both the acquisition of theoretical knowledge at vocational school and practical training at a company. Moreover, vocational training for some sectors, such as crafts, industry and agriculture, is enriched by inter-company training centres, which complement practical training by focusing on activities that are not necessarily carried out in every company for reasons of size and specialty (BMBF, 2020a). According to the BMBF, the dual system is a cornerstone for Germany's flourishing economy (BMBF, 2022a). Especially in view of the shortage of skilled workers, it is therefore crucial that trainees are able to start as well-qualified employees after having finished their education. This entails that trainees gain high levels of required knowledge and skills within practical experiences before they complete their education and enter the labour market.

However, despite its high quality of education, the German VET system reaches its limits when it comes to training tacit knowledge, which is work-related practical know-how acquired through direct experience (Brockmann & Anthony, 2016), and skills in work environments that are inherently complex or even dangerous (Wang et al., 2018). Training under such conditions always involves risks, such as injury (Lenggo Putri et al., 2019) and damage (Bundesanstalt für Arbeitsschutz und Arbeitsmedizin [BAUA], n.d.; Drill, 2013). Likewise, simulating work-related stressful situations for training would increase the risk of serious consequences and is therefore ethically questionable. Lastly, preparing, implementing and supervising training for such contexts is likely to be costly and time-consuming (Deutsche Handwerks Zeitung, 2019; Wang et al., 2018).

To remain competitive in a globalised world, the BMBF strives to continuously develop the German VET system in terms of quality and attractiveness and attaches great importance to means of digitalisation and technology enhanced learning (TEL). The aspiration for increasing digitalisation and TEL in vocational training can be seen in recent publications by the BMBF that concern digitalisation programs (BMBF, 2020b), the dissemination of digital learning and teaching media (BMBF, 2020c), the promotion of innovative training methods (BMBF, 2019) and more. Another initiative that underscores efforts to upgrade vocational training is InnoVET, an innovation competition for excellent vocational education and training funded by the BMBF to the tune of 82 billion euros (BMBF, 2022b). One innovative approach that is in line with the BMBF's efforts to advance vocational training is virtual reality (VR) training. VR involves the immersion into an artificial but realistic world designed to give the user a sense of being present (Bowman & McMahan, 2007). The advantages of VR training are multifaceted, ranging from safety (Sacks et al., 2013) and motivational factors (Sattar et al., 2020) to cost factors (Martín-Gutiérrez et al., 2017). Therefore, VR offers practical, realistic and safe training opportunities for experiential learning that can enrich a wide variety of vocational trainings (e.g., Wang et al., 2018; Xie et al., 2021).

One sector that could particularly benefit from the potentials of VR training is construction, which is known for its dynamic, diverse and often dangerous working environment, including heavy machinery and construction vehicles (e.g., Albert et al. 2014; Hasanzadeh et al., 2018; Kim et al., 2022; Ringen et al., 1995). Moreover, construction is often influenced by various outdoor-related factors which can act as acute stressors, i.e., short-term work demands that elicit a stress response (Stangl, n.d.), making work execution even more demanding. With respect to the person-environment fit theory, individuals are likely to respond with distress, i.e., dysfunctional stress, when the demands of a task outweigh their individual abilities to address these demands (e.g., Edwards & Cooper, 1990; Le Fevre et al., 2003). Therefore, it is vital that trainees learn how to deal with acute stressors, such as time pressure, before entering the labour market. Similarly, construction trainees benefit from high levels of situation awareness (SA), a tacit knowledge that helps them perceive and understand the elements in their environment and predict their status in the near future, but which they usually acquire during repeated on-the-job experiences (Endsley, 1995). Appendix A provides a detailed description of the practical context, including strengths and limits of the German VET system exemplified by the construction sector.

The experience gap, which affects construction trainees' SA-levels and practice in dealing with stressors, is of high relevance since unsafe actions in the construction industry endanger not only the person performing them, but also co-workers or even uninvolved persons such as pedestrians or car drivers (Lenggo Putri et al., 2019). Also, costs are incurred if people are injured, machines are damaged or if the quality of the product is poor (Drill, 2013). Therefore, it should be in the interest of all employees, employers, customers and the state to explore innovative approaches to address this experience gap.

This study started precisely here and investigated whether effects of demanding and stressful working conditions on construction trainees, as they would be expected on a real construction site, can be demonstrated with VR. By investigating the effect of time pressure (TP) as an acute stressor, this study contributes to the exploration of mental stressors for the construction industry as was suggested by Tijani et al. (2021). Specifically, the present study investigated the effect of TP and stress on SA and task performance among construction trainees to answer the following research question: "*Does VR training* *for construction trainees reveal negative effects of time pressure and stress on situation awareness and performance?"*. For this purpose, a realistic, but safe VR environment served to confront German road construction trainees with TP while they executed the task of asphalt compaction and were challenged with SA questions. Multiple hypotheses embedded in three research models were used to investigate the research question.

1.1. VR for Training Professional Skills and Tacit Knowledge

In addition to the games industry, VR has long been discovered for the educational sector and training systems (Rubio-Tamayo et al., 2017). VR involves the immersion into an artificial but realistic world designed to give the user a sense of being present (Bowman & McMahan, 2007). A high level of immersion is attained through the inclusion of different senses (Rubio-Tamayo et al., 2017) and allows for detachment from the real physical world and dive into another reality, whereby intense mental involvement is achieved (Agrawal et al., 2020). A common approach to attain high immersion is the use of a headset (also referred to as VR goggles or head-mounted display [HMD]) which enables the user a 360-degree view around their own axis as well as from top to bottom (Brown & Green, 2016). Another key factor of VR is interaction, which includes the active manipulation and real-time response of virtual objects and is thus directly linked to the idea of immersion (Psotka, 1995).

Previous research has shown the effectiveness of VR in professional skills training in various fields, such as firefighting, medical-surgical training, assembly training, transportation and safety training in construction (Xie et al., 2021). Especially for the medical field, there is ample evidence of performance enhancement through VR training. For example, surgical trainees' performance improved after having received VR training for a medical procedure (Grantcharov et al., 2004). Moreover, VR was found to be effective in training first responders in the emergency medical services. After VR training, first responders' speed in completing tasks increased by 29%, while their accuracy increased by as much as 46% (Koutitas et al., 2021). In the manufacturing sector VR training proved to be superior to traditional training (Abidi et al., 2019). The results of the study show that the participants who were trained with VR ended up making less mistakes and needed less time for the actual product assembly than the participants who received a traditional training. In addition, VR provides a safe and efficient practice environment for driving and dealing with different traffic events. For example, after VR driving training participants were able to improve their reaction time and other driving skills compared to more traditional training methods such as brochures or videos (Lang et al., 2018). Lastly, participants in a multiuser virtual safety training for the correct and safe dismantling procedure of tower cranes learned better than participants using a traditional, non-virtual method (Li et al., 2012).

Another area of interest for VR in the professional context concerns the acquisition of tacit knowledge. Tacit knowledge is a special type of knowledge that instructors or experienced employees face challenges in trying to explain, as it is inherently difficult to verbalize (Weigel et al., 2021). Previous research on VR for training tacit knowledge shows promising results to close this gap. For example, participants in an aerial firefighting supervisor training were able to acquire situation awareness (which is a type of tacit knowledge) after VR training based on HMD (Clifford et al., 2018). Furthermore, VR was found to be helpful for tacit knowledge transfer between companies, i.e., between a manufacturer that designs products which are later assembled by another company. Compared to research on VR for professional skills training, research on VR related to tacit knowledge is still rare, although some research on VR for professional skills training also points to the effectiveness of VR for strengthening tacit knowledge. This includes, in particular, the positive results of VR for first responder training (Koutitas et al., 2021) as well as for driving training (Lang et al., 2018) since high performance in both areas involves tacit knowledge to at least some degree. For example, tacit knowledge as a result of experience (Brockmann & Anthony, 2016) is relevant in anticipating an injured person's health condition in the near future or the result of persistent heavy rain for the road condition. Moreover, tacit knowledge is considered context-specific, personal, practical and procedural (Ambrosini & Bowman, 2001). VR training addresses these attributes of tacit knowledge since it allows for high immersion and practical experience (Xie et al., 2021) as well as for the targeted manipulation of training scenarios and even personalised training routes (Lang et al., 2018).

It is critical to highlight here that the effectiveness of VR training for professional skills and tacit knowledge development is decisively dependent on different quality attributes of the VR simulation. These include that the VR environment reproduces the real working environment and possible work-related challenges to a high degree (Xie et al., 2021). Moreover, it is essential that the VR training allows for a high level of immersion (Weigel et al., 2021). Participants in the VR training for aerial firefighting supervisors mentioned above had a greater ability to acquire SA in a highly immersive environment (e.g., using an HMD) than in a low immersive environment (e.g., using a high-definition TV) (Clifford et al., 2018). Similarly, cognitive skills, psychomotor skills and affective skills could be enhanced after VR training with HMD compared to low immersive methods (Jensen & Konradsen, 2018). Taking these quality attributes into account, VR offers a practical, realistic and mistake-tolerant training opportunity that is particularly suitable for trainees in the construction industry. This includes that trainees can familiarise themselves with the dynamic working environment and task, the construction machines and the unsteady and changing conditions that could make additional demands.

1.2. Situation Awareness

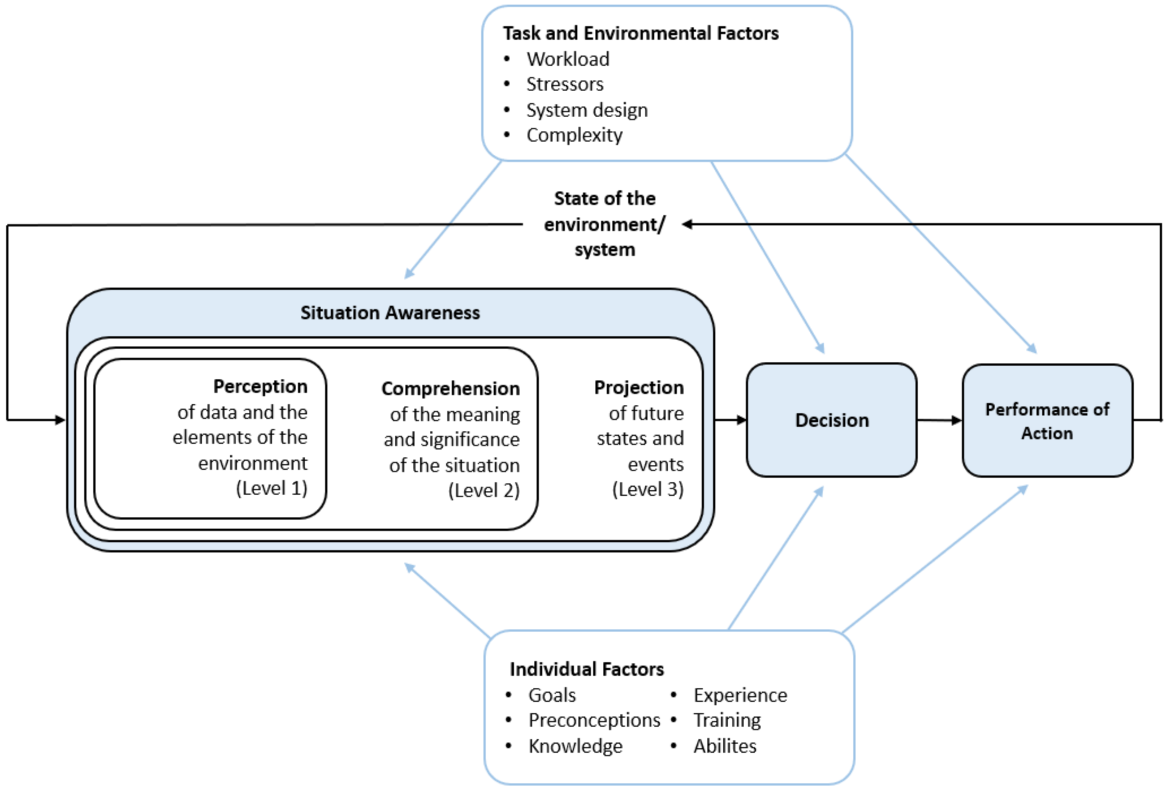
1.2.1. Endsley's Model of Situation Awareness in Complex Decision Making

Situation awareness (SA) is defined as the perception and understanding of elements in the immediate environment and predicting how these elements might change in the near future (Endsley, 1988, as cited in Endsley, 1995). It is a tacit knowledge that informs decision-making and helps to take appropriate actions, especially in complex and dynamic contexts (Endsley, 1995). SA is therefore beneficial for construction trainees to perform their daily work safely and efficiently. Endsley's definition of SA already highlights the three embedded levels of the construct: *perception* of current situation, *comprehension* of current situation and *projection* of future status. For example, a roller operator with a high SA-level would notice a low temperature on the temperature display, understand that the asphalt cools fast, and anticipate that a drop in temperature below five degrees will require a complete compaction stop.

Endsley's concept of SA forms the basis for Endsley's model of situation awareness in dynamic decision making (Endsley, 1995). According to this model, SA provides a person with relevant information to make an appropriate decision, which in turn influences that person's performance of action (see Figure 1). This process is influenced by individual factors, such as differences in knowledge and experience as well as by task or environmental factors, including task complexity, level of automation and stressors. For construction trainees, this means that their limited experience affects their SA-level as do environmental factors, such as traffic, noise, temperature, and weather which can act as stressors.

Figure 1

Endsley's Model of Situation Awareness in Dynamic Decision Making (Adapted from Sapateiro et al., 2008)



1.2.2. Acquisition of Situation Awareness

SA is a type of tacit knowledge which comprises work-related practical know-how acquired through personal experience (Brockmann & Anthony, 2016). In other words, employees develop SA on the job, rather than through formal classroom training. According to Endsley (1995), repeated experiences in an environment form the basis for developing expectations about future events. Consequently, little experienced workers, such as trainees or recently qualified workers, generally do not yet have adequate SA-levels when being confronted with varying, unfamiliar environmental factors.

1.2.3. Relevance of Situation Awareness

The relevance of SA lies especially in understanding and preventing accidents caused by humans in complex and dynamic contexts (Endsley, 1995; Wickens, 2021). Research on SA has long exceeded the original scope of military aircraft industry (Endsley, 1995), addressing related contexts, like civil aviation (Salmon et al., 2016), as well as diverse, unrelated contexts, such as logistics (Choi et al., 2020), healthcare (Green et al., 2017), and the construction industry (Hasanzadeh et al., 2018). The results demonstrate that a high level of SA is beneficial for safe task execution, stressing the relevance of SA for construction workers. In addition to accident prevention and safety promotion for the person performing the task as well as for uninvolved persons, SA can also influence the quality of the task that is to be performed (Endsley, 1995; Jonsson et al., 2021) and reduce the error-rate (Endsley, 1999). For example, SA correlated negatively with human error ($r = -.64$) and mediated the relationship between safety knowledge and human error in a questionnaire-based study among 601 workers of different industries (Mohammadfam et al., 2021). The beneficial effect of SA could also be supported in a study that explicitly used VR technology to investigate the role of SA for error recognition in a virtual operating room (Bracq et al., 2021). The findings demonstrate that scrub nurses who detected most errors had higher SA-levels and identified high-risk errors faster. Finally, high SA-levels are also desirable in view of additional costs or non-compliance with deadlines due to human error (Marquardt, 2019).

1.3. Stress

1.3.1. Person-Environment Fit Theory and Distress

Stress at the workplace is a ubiquitous phenomenon nowadays, caused by various types of workplace stressors, also referred to as job demands (Lepine, 2022). An often-employed explanation for the development of stress is provided by the person-environment fit theory (from here on: PEF theory) (Edwards & Cooper, 1990), whereby “fit” refers to “match”, “congruence” or “correspondence” (Dewe et al., 2012). The PEF theory examines the fit between the characteristics of individuals and their environment and proposes that stress arises from a discrepancy between them (Edwards & Cooper, 1990). The PEF theory incorporates two main perspectives. First, the *demands-ability fit* comprises the degree of correspondence between the demands employees face and their abilities to meet these demands. Second, the *needs-supplies fit* encompasses the correspondence between employees’ needs (e.g., physical and psychosocial needs) and the resources available to them (Dewe et al., 2012). The present work focused on the demands-ability fit only. Moreover, the fit between person and environment, or in this case demands and abilities, can be appraised on both an objective and on a subjective level. The objective appraisal describes the demands of the environment and individual abilities as they actually exist, whereas the subjective appraisal comprises individuals’ perceptions of the two components (Edwards & Cooper, 1990; Pasca, 2014). For the theoretical basis of this work, only objective appraisal was used, i.e., the general situational demands on a construction site were compared with the abilities generally expected of construction trainees.

According to PEF theory, stress arises when the demands of a task or a situation do not match the personal abilities to deal with these demands, whereby a mismatch can generally comprise both too low and too high demands in relation to the personal abilities (Le Fevre et al., 2003). However, at the workplace, a mismatch is commonly associated with highly demanding working conditions or employees who are not fully prepared to handle the demands at the workplace (Pasca, 2014). In this context, the resulting stress response is also referred to as distress, i.e., negative or dysfunctional stress (Le Fevre et al., 2003) and is associated with negative implications for cognitive abilities (e.g., Jokela, 2022) and job performance (e.g., LeBlanc, 2009). In contrast, a good fit between individuals and their environment is associated with a eustress response, i.e., positive or functional stress (Le Fevre et al., 2003). In other words, whereas a low and high level of demands is generally associated with a distress response, a moderate level of demands is associated with a eustress response (Pluut et al., 2022). This is consistent with the Yerkes-Dodson Law (Yerkes & Dodson, 1908, as cited in Van Veldhoven, 2014). According to this theory, an optimal level of job demands leads to an optimal level of activation which is beneficial for performance. In contrast, a too low activation level and a too high activation level, e.g., due to demanding working conditions or the imposition of an additional stressor, have a negative effect on task performance.

With respect to the ability-component, in particular, there are several factors that determine whether individuals can deal with the demands of their environment. Although an elaboration on all these factors would exceed the scope of this study, one fundamental characteristic of the target group in question, i.e., construction trainees, will be described in more detail. The characteristic includes the little experience and the limited prior knowledge of construction trainees, which are relevant as they are associated with the level of cognitive load that individuals experience. Whereas experienced people generally feel a lower cognitive load when confronted with a familiar problem, novices are likely to feel a high cognitive load when confronted with the same task or problem (Kolfschoten & Brazier, 2013). Similarly, learners with little prior knowledge are more cognitively loaded when learning complex tasks than learners with high levels of prior knowledge (Ayres & Paas, 2012). Consequently, trainees' cognitive load is likely to be high, as they have limited experience and little prior knowledge, unlike their colleagues who have completed their training and have been in the profession for years. The level of cognitive load is relevant in this context since cognitive overload is associated with diverse negative effects, including impaired performance and decision-making, difficulty to retrieve, analyse and organize knowledge as well as stress (Eppler & Mengis, 2002).

1.3.2. The Effect of Time Pressure on Construction Trainees: A Distress Response

TP is a frequently experienced workplace stressor, especially in Germany. In 2019, almost 40% of employees in the European Union reported that they "often" or "always" experience TP at work. In Germany, it was almost 50% (Eurostat, 2019). A labour market survey identified TP even as the main stressor for the German workforce (Orizon, 2019). The relevance of TP in the work context is thus obvious. TP is a so-called quantitative job demand which refers to the amount of work and the speed to accomplish a task (Van Veldhoven, 2014).

Whether individuals experience distress or eustress depends on the degree of demands in relation to their personal abilities to cope with these demands (Le Fevre et al., 2003). Construction trainees have limited prior knowledge and little experience in executing procedures and dealing with the demands of the inherently complex and dangerous working conditions on a construction site (e.g., Albert et al., 2014, Kim et al., 2022). With respect to these already demanding working conditions, it is reasonable to assume that the baseline level of demands (without an additional stressor) already leads to a high activation level with respect to the Yerkes-Dodson Law (Yerkes & Dodson, 1908, as cited in Van Veldhoven, 2014). Moreover, limited prior knowledge and little experience probably cause high cognitive load during task execution. Consequently, confrontation with TP as an additional stressor is likely to exceed the optimal activation level and to exacerbate cognitive load or even cause cognitive overload. With respect to the PEF theory, it is therefore probable that exposure to TP exceeds trainees' ability to cope with the situation (e.g., a construction task), leading to a distress response and associated negative effects on cognitive abilities and performance.

1.3.3. Types of Stress by Timescale

The occurrence of stress can be specified along different timescales (Crosswell & Lockwood, 2020): *Chronic* stress pertains to persistent threatening or challenging stressors that affect daily life for an extensive period of time (e.g., role conflict at work or a dissatisfying job). *Daily hassles* comprise daily stressors in the form of interruptions or difficulties that happen on a daily basis (e.g., traffic or work overload). *Acute* stress roots in short-term, event-based exposures to threatening or challenging stressors (e.g., TP). Despite this distinction, it is important to note that the categories often overlap in reality. For example, TP at work might cause acute stress in the first place, but if it occurs every day, it might also be considered a daily hassle or even chronic stress.

Most research in the construction industry does not specify on the type of stress or the precipitating stressor on a timescale. Since TP in the construction industry is often studied in terms of schedule pressure (e.g., Nepal et al., 2006; Webb et al., 2015), it is likely that it relates to the daily, if not even chronic type of TP and stress. A possible reason for the low attention paid to TP as an acute stressor could be the fact that acute stressors in outdoor construction work often cannot be planned (e.g., weather) or require disproportionate effort (e.g., simulating traffic). Similarly, it is difficult to include such acute stressors in practical training. This study used VR to circumvent these obstacles and investigated TP as an acute stressor.

1.3.4. Indicators for Stress

1.3.4.1. Subjective Stress

Subjective stress measurement by means of self-report questionnaires is a low-cost practice with little expense that is often used for stress research in organizations (e.g., Bregenzer & Jimenez, 2021; Moridi et al., 2014) and also specifically for stress research in the construction industry (e.g., Zhang et al., 2023). Questionnaires provide insights into individuals' psychological reactions, such as emotions, as a result of a preceding stressor (Crosswell & Lockwood, 2020) and often take a retrospective perspective (Weckesser et al., 2019). When assessing subjective stress, it is crucial to select an appropriate stress measure that fits the research question and the sample. Also, characteristics of the stressor and stress response, such as timescale and duration, should be clarified first (Crosswell & Lockwood, 2020). Respectively, this study used a questionnaire that is compatible with the investigation of *acute* stress.

Nevertheless, self-report questionnaires for stress measurement are associated with a number of concerns and limitations (Razavi, 2001). For example, consciously perceived stress explains a limited amount of variance in physiological stress reactivity, i.e., well-known reactions of the body, such as increased heartbeat or sweating (Epel et al., 2018). The limited association between subjective stress experience and physiological stress responses can be attributed to several factors, including the willingness to truthfully report the stress state and the fact that some events only affect us on an unconscious level (Epel et al., 2018).

1.3.4.2. Physiological Stress

An alternative approach to measuring stress that overcomes the above mentioned limitations of self-report questionnaires is the assessment of physiological indicators (Crosswell & Lockwood, 2020). Examples are cortisol-levels, heart rate variability and electrodermal activity (Giannakakis et al., 2022). The latter is an umbrella term that concerns "electrical changes, measured at the surface of the skin, which arise when the skin receives innervating signals from the brain" (Empatica, 2021) and usually increases in response to emotional or physical arousal (Empatica, 2022). Electrodermal activity (EDA) has been widely used to assess stress at the physiological level (e.g., Liu & Du, 2018; Setz et al., 2010; Wickramasuriya et al., 2018). Moreover, EDA has often been used to detect *acute* stress under controlled conditions (e.g., Greco et al., 2021). EDA as an indicator of physiological stress can therefore provide additional insights that complement self-report questionnaires on stress.

To better understand how EDA is a meaningful indicator in the context of stress research, the sympathetic and parasympathetic nervous system, as part of the autonomic nervous system, are roughly explained. The autonomic nervous system regulates the processes in the body that cannot be directly influenced by human will (e.g., heartbeat and metabolism). The sympathetic and parasympathetic nervous system function as opponents: The sympathetic nervous system sets the organism up for an increase in activity which is often referred to as "fight or flight". In contrast, the parasympathetic nervous system predominates in phases of rest and regeneration, when the organism 'rests and digests' (McCorry, 2007). EDA is considered the most useful index of changes in the sympathetic arousal as a result of emotional and cognitive states, since it is the only autonomic psychophysiological indicator that is not affected by parasympathetic activity (Braithwaite et al., 2015). Moreover, EDA is commonly divided into two types. The first one is the phasic skin conductance response (SCR), which is characterized by rapidly changing peaks in the measurement resulting from sympathetic neuronal activity (Empatica, 2021). The second one is the tonic EDA-level, which comprises the slowly changing levels of the EDA signal. This type is commonly measured by the *skin conductance level* (SCL), whereby "changes in the SCL are thought to reflect general changes in autonomic arousal" (Braithwaite et al., 2015, p. 4). This means, higher SCLs are related to higher stress-levels. Moreover, the SCL is often used when investigating acute stress (e.g., Knaust et al., 2022). Due to the natural variability of the SCL between individuals, it is advised to use it with a within-subjects design and to calculate the difference in a persons' SCLs between conditions (Braithwaite et al., 2015).

1.4. The Effect of Time Pressure on Stress, Situation Awareness and Performance

TP has often been identified as one of the root causes for work stress in the construction industry (e.g., Aksorn & Hadikusumo, 2007; Campbell, 2006; Pegler, 2021). For example, in a qualitative study on the causes for work stress in construction, it was found that workers suffer from stress due to the complex and intertwined work processes in the construction industry, as well as specific workplace stressors, such as TP (Van der Molen & Hoonakker, 2014).

The influence of TP on SA is suggested in Endsley's model of situation awareness in dynamic decision making (1995) as described earlier. In addition, research found evidence for decreased SA-levels due to TP. For instance, Bustamante et al. (2005) investigated the effect of TP as a potential weather threat—for which pilots have less time to respond adequately—on pilots' SA-levels and found decreased SA-levels when pilots approached the weather threat.

Before discussing the impact of TP on performance, the construct of performance in the context of construction is briefly described. Performance in construction can be assessed at different levels, with *quantity* and *quality* of work as the most basic levels (e.g., Abu Oda et al., 2022). Safety performance is another relevant and frequently studied indicator of performance in construction (e.g., Mohammadi et al., 2018), but was not examined in the present study. Two commonly assessed indicators of construction performance at the quantitative level are labour productivity, which can be described as the number of work units produced per hour of work, and unit rate, which basically describes the opposite, namely the number of working hours required to complete a work unit (Halligan et al., 1994, as cited in Shehata & El-Gohary, 2011). The quality of work, on the other hand, includes the quality of workmanship and outputs as well as the number of defects (Abu Oda et al., 2022). Based on this differentiation, this study assessed the number of work units completed within a specific timeframe (quantitative level) and, at least to some degree, correct workmanship (qualitative level).

Previous research in both construction and other sectors showed that performance suffers when working under schedule pressure. For example, a questionnaire-based survey examined the impact of schedule pressure on perceived construction performance among 102 construction workers (Nepal et al., 2006). It was found that schedule pressure adversely affects the performance indicators of productivity and quality. Similar results were obtained in a meta-analysis by Szalma et al. (2008) which revealed that TP has a positive effect on speed but impairs accuracy in both perceptual and cognitive tasks. Moreover, in an experimental approach to investigate participants' search for complex problem-solving strategies in the presence of TP as an acute stressor, participants had difficulty finding a complex strategy when being exposed to TP (Hiel & Mervielde, 2016).

1.5. The Effect of Stress on Situation Awareness and Performance

Previous research indicates a negative effect of stress on SA. For example, Sneddon et al. (2013) studied the effect of stress on SA and found that higher stress-levels were associated with lower levels of work SA in offshore drilling crews. Agrawal and Peeta (2021) conducted an experimental study using a driving simulator to evaluate the effect of SA and mental stress on the transition from automated driving to human driving. Similar to the previous mentioned study, the authors found significant negative correlations between physiological indicators of SA and psychological stress. Since research on stress and SA resulted in correlation findings only, the scope was expanded to research that addressed attention, instead of SA. This expansion was considered acceptable since attention is required for all of the three SA-levels (Endsley, 1995). Sänger et al. (2014) investigated the influence of acute stress on attention mechanisms. They found that stressed participants showed higher error rates in detecting a stimulus change while ignoring another more prominent, but task-irrelevant stimulus.

Likewise, previous research suggests that performance in construction is negatively affected by stress. For example, according to Stavroula et al. (2003), work-stress impairs performance and productivity and increases unsafe working practices and accident rates. However, research that explicitly focused on the effect of stress on performance in the construction industry is rare. In contrast, the effect of stress on safety behaviour has been widely researched. For example, the impact of stress on safety behaviour and accidents of construction workers was investigated by Leung et al. (2016) who were able to show that physical stress negatively predicts safety behaviour.

1.6. Hypotheses

TP was found to create work stress (e.g., Van der Molen & Hoonakker, 2014) and to decrease SA (e.g., Bustamante et al., 2005) and performance (e.g., Nepal et al., 2006). This holds especially for construction trainees who already face a discrepancy between their abilities and the demands of their job on a daily basis. With respect to the PEF theory (Edwards & Cooper, 1990), exposing construction trainees to TP, as an additional demand to the already existing demands-ability mismatch, is likely to increase their stress-levels even more and to impair their SA and performance. These findings provide the basis for the first research model, which includes hypotheses H1, H2 and H3. The model examined differences in subjective stress, physiological stress, SA and performance between a low time pressure condition (low TPC) and a high time pressure condition (high TPC) (see Figure 2).

H1a: Trainees have higher subjective stress-levels in the high TPC compared to the low TPC.

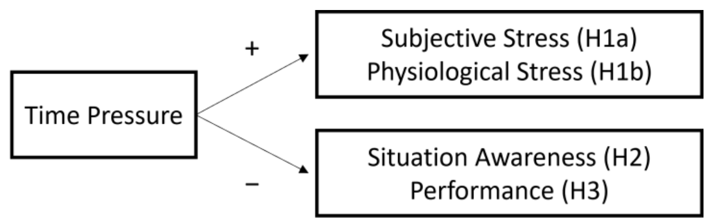
H1b: Trainees have higher physiological stress-levels in the high TPC compared to the low TPC.

H2: Trainees have higher SA-levels in the low TPC compared to the high TPC.

H3: Trainees perform higher in the low TPC compared to the high TPC.

Figure 2

Subjective Stress, Physiological Stress, SA and Performance in the High Time Pressure Condition



Moreover, previous research provides evidence for the negative effect of stress on SA (e.g., Sneddon et al., 2013) and performance (e.g., Stavroula et al., 2003). As a consequence, it is assumed that lower SA-levels and performance in the high TPC are determined by increases in stress-levels. Therefore, in the second research model comprising H4 and H5, a moderation effect was investigated by testing whether differences in SA-levels and performance, as a result of TP, depend on differences in stress-levels (see Figure 3).

H4a: Trainees' subjective stress-levels determine lower SA-levels in the high TPC.

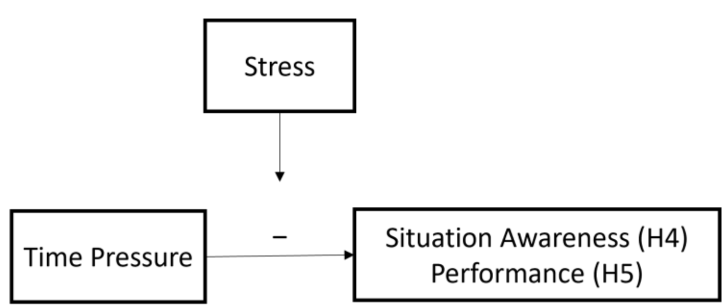
H4b: Trainees' physiological stress-levels determine lower SA-levels in the high TPC.

H5a: Trainees' subjective stress-levels determine lower performance in the high TPC.

H5b: Trainees' physiological stress-levels determine lower performance in the high TPC.

Figure 3

The Effect of Stress for Lower SA-Levels and Performance as a Result of Time Pressure



Lastly, the baseline level of demands in construction is high (e.g., Albert et al., 2014). With respect to the PEF theory and the proposed demands-ability mismatch (e.g., Le Fevre et al., 2003), it is therefore likely that trainees experience stress regardless of the level of TP they face while executing a task. Also, as described for the second model, previous research found evidence for a negative effect of stress on SA and performance. Therefore, the third research model suggested that stress-levels negatively predict SA and performance within conditions (see Figure 4). For this end, both stress variables were simultaneously inserted as predictors to investigate which part of the variance in SA or performance is explained by subjective stress and which part is explained by physiological stress. Although both variables ultimately refer to the same construct, previous research found ample differences between them which supported the simultaneous inclusion of both variables (e.g., Epel et al., 2018). For example, Cusveller et al. (2014) conducted an experimental study in which they monitored participants' physiological response after different types of computer games and compared the results to subjective stress reports. They found increases in both skin conductance and subjective stress-levels, but no significant correlation between the two indicators of stress.

H6a: Trainees' subjective and physiological stress-levels have a negative effect on trainees' SA-levels in the low TPC.

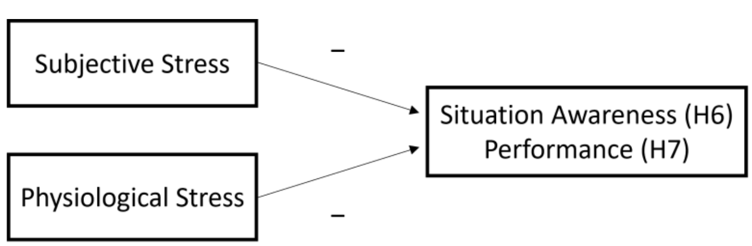
H6b: Trainees' subjective and physiological stress-levels have a negative effect on trainees' SA-levels in the high TPC.

H7a: Trainees' subjective and physiological stress-levels have a negative effect on trainees' performance in the low TPC.

H7b: Trainees' subjective and physiological stress-levels have a negative effect on trainees' performance in the high TPC.

Figure 4

The Negative Effect of Subjective Stress and Physiological Stress on SA and Performance within Conditions



2. Method

2.1. Research Design and Participants

This study utilised an experimental within-subjects-design and quantitative data collection. The experimental design in combination with VR technology served to simulate the real working environment for an asphalt compaction task. The population focus for this study was German road construction trainees. Specifically, only second- and third-year trainees were targeted, as asphalt compaction is not yet part of the curriculum for first-year trainees (BauWiAusbV, 1999). Economic data collection required that trainees were at vocational school or inter-company training centre for data collection rather than split up in their training companies. Therefore, data collection could only take place within specific time frames. Also, there was only one vocational school and inter-company training centre for road construction trainees in the vicinity. For reasons of time constraints and distance, participants were therefore recruited via a convenience sample at one vocational school and one inter-company training centre only. Apart from the year of education no further criteria were defined.

As this study investigated group differences, a minimum sample size of 30 participants per cell was required to achieve high power of 80%. For investigating relationships, a sample size of 50 is considered reasonable (Wilson VanVoorhis & Morgan, 2007). Ultimately, 53 road construction trainees (31 second-year and 22 third-year trainees) participated in the study. None of the participants suffered from motion sickness or dropped out of the study prematurely for other reasons. Trainees were between 17 and 30 years old (M = 20.57, SD = 3.47). About three quarters of participants (75.47%) had no or only one practical experience with asphalt compaction. Similarly, 77.36% had no or only one practical experience with VR. After having participated in the study, almost all trainees rated the integration of VR into their vocational training as “meaningful” (94.34%). The same applies to trainees’ motivation to use VR training in their vocational training: 94.34% reported that they would be “motivated”. Only three trainees rated the integration of VR as “not meaningful” or “partly meaningful” and their motivation to use VR training as “not motivated” or “partly motivated”.

2.2. Equipment and Task

The study used a VR simulation to replicate the task of asphalt compaction. For this purpose, participants wore a head-mounted display which enabled high immersion into the virtual 3D environment (Dhimolea et al., 2022). A joystick and a steering wheel that were attached to the laptop served for forward and backward movement and steering. The task comprised the operation of a roller to compact a newly paved motorway. Precisely, participants followed one of two paving machines and had to decide how much distance they should keep, when they should water the bandages of the roller and when to use the vibration. Figure 5 shows a picture of the setup.

Figure 5

Setup of the VR Simulation

TP, as the independent variable, was induced by means of two different VR scenarios of six minutes each. In the high TPC, the weather conditions were unfavourable for asphalt compaction: The cold temperature of seven degrees and the high probability of rain lead to TP on two levels. First, the asphalt cools down quickly and the roller operator has to stay close to the paving machine. Second, the roller operator must expect that compaction will have to be stopped completely as soon as the temperature drops below five degrees, or it starts to rain (Deutscher Asphaltverband e.V., 2016). In comparison to the high TPC, the low TPC had favourable weather conditions: Due to sunny weather and hot temperatures of 30 degrees, roller operators do not have to hurry with compaction. In contrast, they should keep a large distance from the paving machine so that the asphalt has enough time to cool down. Figure 6 illustrates the low TPC (top) and the high TPC (bottom). Before starting the compaction task, all participants were briefed on the two weather conditions and their impact in terms of TP. For this end, two briefings (see Appendix B) were prepared in advance and read aloud to participants to ensure that everyone received the same information and to compensate for differing levels of prior knowledge.

Figure 6

2D Representation of the Asphalt Compaction Task for Both TP Conditions

Note. The first representation demonstrates a large distance between roller and paving machine in the low TPC. The second representation shows a short distance between roller and paving machine as a result of the high TPC.

2.3. Instruments

2.3.1. Situation Awareness Measurement

The SA-level was assessed with multiple questions based on the Situation-Awareness-Global-Assessment-Technique (SAGAT) (Endsley et al., 1998) which were specifically developed for a hot and sunny and a cold and rainy motorway scenario. Two sets of questions each comprised twelve SA questions, with four questions at the perception-, comprehension- and projection-level (e.g., “Which rolling phase are you currently working on?”). The questions were designed as multiple choice questions with two to four answer options per question. One point was awarded for each correct answer, so that a total of 12 points could be achieved per scenario.

Originally, the questions were developed for and used in the Dutch context. As this study was conducted in Germany, the questions were translated into German. For this end, two subject teachers were consulted to ensure correct translation, especially concerning technical terms. During translation, differences between the Dutch and German guidelines for asphalt compaction were detected. Specifically, some SA questions and answers did not correspond to the way asphalt compaction is regulated and practiced in Germany or are hardly addressed in vocational school. For example, in Germany a different type of asphalt is usually used and dealt with at vocational school than in the Netherlands. Also, contrary to the Netherlands, asphalt is usually not compacted during rain. Therefore, the SA questions and answers were slightly modified to be applicable for asphalt compaction in Germany. Appendix C provides an overview of all SA questions.

2.3.2. Task Performance Measurement

The aim of the asphalt compaction task was to compact the right half of a virtual road four times, which was appraised a reasonable goal for two lanes given the length of the street and the six minutes per scenario. The right half of the road represents an area of 160 squares that serve as measuring points and that register every roller crossing. To evaluate compaction performance, the total number of crossings by each participant and scenario was set in relation to the targeted number of crossings which is 640 (the product of 160 and four). The result is a percentage number indicating to what degree the aim of crossings was attained. This approach ensured that not only the squares compacted, i.e., squares with a value greater than zero, were evaluated, but also the number of crossings on each square. Since the aim was to compact the road four times, i.e., crossing each square four times, any crossing above four did not yield additional points. However, squares with values above four did not result in negative points either, since the redundant crossings were already considered insofar as they have not been used for other squares for which the target of four crossings had not yet been reached. Another argument against negative points was that the software registers a crossing even if only the edge of a square was crossed. Since it is common in the real world to slightly overlap two roller lanes, measuring points with more than four crossings were treated equally to measuring points with four crossings.

It is important to note, that the software takes into account that asphalt takes longer to cool down and to be ready for compaction at hot temperatures. If the virtual asphalt is still too hot, the measuring points would not register a crossing. Therefore, it is necessary to wait 20 seconds before starting compaction. Consequently, participants had 20 seconds less time in the hot scenario than in the cold scenario. This discrepancy was considered when determining the targeted number of crossings in relation to the given time in the hot scenario. Specifically, the targeted number of crossings was reduced by 36 crossings so that the ratio of given time and targeted crossings in the hot scenario (340 seconds for 604 crossings) was the same as in the cold scenario (360 seconds for 640 crossings).

2.3.3. Stress Measures

2.3.3.1. Subjective Stress Measurement

The Perceived Stress Questionnaire (PSQ) from Levenstein et al. (1993) measures the current subjective stress-level. Since the study was conducted among German construction trainees, the German version of the PSQ translated by Fliege et al. (2009) was used to measure trainees' subjective stress-levels in both scenarios. The German version of the PSQ consists of four scales with five items each and was statistically validated (N = 650 participants). Factor-analysis revealed the factors "worries", "tension", "joy" and "demands" with an overall internal consistency of Cronbach's alpha = .85 and a split-half-reliability of r = .86 (Guttmann split-half coefficient) and r = .88 (Spearman-Brown split-half coefficient). The psychological domain of the WHOQOL-Bref, which measures quality of life (Angermeyer et al., 1999), and the F-SOZU, which measures social support (Sommer & Fydrich, 1991), were used to affirm convergent validity with $r = -.79$ and $r = -.61$ respectively. Objectivity of the assessment can be considered to be given, as the examination only requires simple calculations in the lower numerical range. For this purpose, the questionnaire manual provides a formula that was used to calculate the overall score. First, the given answers were coded from 1 (almost never) to 4 (usually). Then, the summed item scores were divided by the number of items. Subsequent transformations served to change the scale range to 0 to 100.

Since this study specifically examined the effect of the stressor TP, only the "demands scale" was used for reasons of questionnaire efficiency and time constraints. This scale was regarded most appropriate for the purpose of this study as it comprises the measurement of demands due to lack of time, deadline pressure or task load and thus reflects key elements of the PEF theory. The demands scale has a convergent validity of $r = -.37$ with the F-SOZU. Although the validity is relatively low, there were two main reasons supporting the use of the demands scale for this study. First, to the best of my knowledge, the demands scale was the most appropriate scale for the context of this study and both validated and freely available. Second, the demands scale still needed to be slightly modified to adapt it to the way TP was induced in this study and to the study design. The modifications, which are described in the next paragraph, served the general purpose to increase validity, however validity was not explicitly tested.

The demands scale has an internal consistency of Cronbach's alpha = .80 and a split-half reliability of r = .69 (Guttmann split-half coefficient) and r = .74 (Spearman-Brown coefficient). Since an internal consistency of Cronbach's alpha = .70 is considered "good", the demands scale fulfils the psychometric property of reliability (George and Mallery, 2003; Streiner, 2003). Although the demands scale was overall quite suitable for the purpose of this study, slight modifications were required for items four and five. First, the fourth item that states "You have enough time for yourself." was replaced by the German translation of "You have enough time.". Second, "Termindruck", which refers to deadline or schedule pressure was exchanged with the German translation of time pressure (i.e., "Zeitdruck"). Moreover, as a result of a small pilot study, the items were changed from present to past tense, since the questionnaire was completed after participants had finished the task and took off the HMD. Participants of the pilot study reported that the present tense was confusing and unanimously suggested the past tense as more appropriate. An overview of the original and adapted items is provided in Appendix D.

As a result of these modifications, Cronbach's alpha was calculated to assess the internal consistency of the demands scale, which consists of five items. A Cronbach's alpha = .70 is generally regarded as "good" (George & Mallery, 2003; Streiner, 2003) and should therefore be achieved as a minimum. The internal consistency for the high TPC met the psychometric property of validity, with Cronbach's alpha = .78, whereas the internal consistency for the low TPC, with Cronbach's alpha = .66, did not. Item-scale-statistics revealed that item four decreased the reliability for the demands scale in both conditions. This is consistent with observations made during study conduction: While filling out item four, it often occurred that participants commented on the six-minute time frame of the two VR scenarios as being too little or just right. It is therefore reasonable to suppose that participants misinterpreted item four which says, "I had enough time.". Precisely, it is assumed that participants related item four to the duration of the VR scenarios (that both were six minutes long), rather than to low or high TP induced within the conditions. Since the exclusion of item four yielded much better results for reliability in both scenarios, it was decided to exclude the item based on the rationale given above. As a result of the exclusion, the demands scale in both conditions fulfilled the psychometric property of reliability, with Cronbach's alpha = .83 for the high TPC and Cronbach's alpha = .72 for the low TPC.

2.3.3.2. Physiological Stress Measurement

To measure physiological stress, the Empatica E4 wristband (from here on: Empatica) was attached to participants' wrist while they executed the task of asphalt compaction. The Empatica is a medical device that enables real-time and continuous collection of physiological data in a convenient way (Empatica, 2020a). It contains four sensors to measure different physiological parameters. For the purpose of this study, electrodermal activity (EDA) was used to assess participants' physiological response in both scenarios and to derive their physiological stress-level. As described earlier, the skin conductance level (SCL), which comprises the slowly changing levels of the EDA signal, was used for physiological stress measurement, whereby higher SCLs are related to higher stress-levels. As proposed by Empatica (2021), the SCL was computed as an average value of the data collected during the asphalt compaction task (Empatica, 2021).

2.4. Experiment Preparation and Procedure

A vocational school and an inter-company training centre which teach and train construction trainees were contacted five months before data collection. Framework conditions and trainees' prior knowledge with regard to the asphalt compaction task were discussed with the subject teachers. The teachers were also asked to refine the German translation of the simulation made for this study, in particular with regard to technical terms. In this context, differences between the Dutch and German guidelines for asphalt compaction were detected and resolved within four iterations, resulting in a slightly modified simulation, especially regarding SA questions and contextual aspects, such as temperature boundaries and compaction levels. Modifications that required changes in the code were mapped in a storyboard and implemented with the help of the BMS Lab of the University of Twente. New features of and tips for the current version (as of November 2022) were recorded in an instruction manual.

Moreover, together with the subject teachers a general briefing for the asphalt compaction task as well as two condition-specific briefings were developed to ensure that all students had the required prior knowledge (see Appendix B). Precisely, second-year trainees generally do not have the same level of knowledge as third-year trainees and some trainees, depending on the specialty and size of their company, commonly have more experience with asphalt compaction than others.

Prior to data collection, ethical approval for the research proposal was requested and obtained from the Ethics Committee of the University of Twente. Preparations, testing, and fixing of software and hardware problems were completed by the start of the study. As a result of a small pilot study (N = 3), a checklist was developed to ensure objectivity and that no step was accidently omitted (see Appendix E). Also, a schedule was created to coordinate the data collection at vocational school and at the inter-company training centre. On the first day at each location, all trainees received the general briefing from the teachers and were informed by the researcher about the purpose and risks of the study. The equipment was set up at both locations in a separate and quiet room and dismantled at the end of each data collection day. Over five weeks, the trainees came one by one to the designated room during selected lessons (vocational school) or time slots (inter-company training centre). Participation required about 30 minutes. First, trainees were guided through the form of consent before signing it. They were specifically told to directly take off the HMD if they did not feel well. Then, trainees were introduced to the hardware, before receiving the briefing for the low TPC and the high TPC. It is important to note here that the order of the low TPC and the high TPC was balanced to control for any practice effects. The start scenario was chosen in advance, i.e., before trainees entered the room. Immediately after the briefing, trainees started the asphalt compaction task, followed by the SA questions and a short stress questionnaire. This procedure was repeated with the second scenario and the session was finished with a short questionnaire on personal data, trainees' practical experience with asphalt compaction and VR as well as their opinion on the integration of VR in vocational training (see Appendix F). The Empatica, that had been attached to trainees' wrist at the beginning of the session, was restarted for each participant, producing one data file per participant. Moreover, timestamps were set once the compaction task was started and finished.

2.5. Data Analysis

First, the automatically recorded output files and screen recordings from the simulation were transferred and organized per participant. The output files entailed the results of the SA questions and a heatmap for the compaction performance. Most SA questions were evaluated automatically by the simulation. Some questions required manual evaluation (see Appendix C). The results of the PSQ demands scale were summed to a score according to the formula given in the manual. Analysis of the Empatica data required the installation of the E4 manager to transfer the data to the E4 connect cloud from which the data could be downloaded as CSV-files. EDA was measured four times per second, resulting in a couple of thousand measurements per participant. To filter the EDA-file for the data collected during the asphalt collection task only, all measurements needed to be linked to unix time first. A self-written Python programme eventually served to enter the timestamps and to mark out the relevant data. The filtered EDA values were then calculated to a mean value (Empatica, 2021). To analyse differences in SA-levels and performance scores as a result of differences in stress-levels (H4 and H5), subjective stress of the low TPC (which served as baseline level of subjective stress) and the difference score of subjective stress (which served as indicator of change between TP conditions) were centred for reasons of multicollinearity. The same procedure was applied to physiological stress.

3. Results

Table 1 shows descriptive statistics and correlations for subjective stress (substress), physiological stress (phystress), SA and performance in both TP conditions.

Table 1

Descriptive Statistics and Correlations for Study Variables

Variable	M	SD	1	2	3	4	5	6	7
1. Substress LTP	15.57	15.93	—						
2. Substress HTP	25.63	23.16	.570**	—					
3. Phystress LTP	1.896	2.586	.076	.091	—				
4. Phystress HTP	2.175	3.091	.061	.129	.752**	—			
5. SA LTP	9.30	1.58	−.012	−.084	−.290*	−.107	—		
6. SA HTP	9.85	1.35	−.127	−.002	−.165	−.058	.302*	—	
7. Performance LTP	25.55	9.18	.212	.154	−.031	−.130	−.271	−.116	—
8. Performance HTP	29.30	9.21	.217	.259	.013	.026	−.199	−.096	.669**

Note. *N* = 53. LTP = low time pressure. HTP = high time pressure.

* p < 0.05 (2-tailed)

** p < 0.01 (2-tailed)

3.1. Model 1: Differences Between Conditions

Prior to testing hypotheses H1, H2 and H3, assumptions for paired samples t-tests were tested and could be confirmed. A detailed description is attached in Appendix G. Table 2 provides an overview of the results of the paired samples t-tests.

Table 2

Differences in Subjective Stress, Physiological Stress, SA and Performance as a Result of TP

Logistic parameter	Low TPC		High TPC		*t*(52)	*p*	Cohen's *d*
	M	*SD*	*M*	*SD*			
Substress	15.57	15.93	25.63	23.16	3.811	<.001	.523
Phystress	1.896	2.586	2.175	3.091	0.990	.163	.136
SA	9.30	1.58	9.85	1.35	2.292	.013	.315
Performance	25.55	9.18	29.30	9.21	3.647	<.001	.501

Note. *N* = 53.

H1a stated that trainees have higher subjective stress-levels in the high TPC compared to the low TPC. H1a could be confirmed based on the results of the t-test (see Table 2). This was also the case after excluding five outliers. Hypothesis H1b said that trainees have higher physiological stress-levels in the high

TPC compared to the low TPC. However, physiological stress-levels were not significantly higher in the high TPC compared to the low TPC (see Table 2). Removing outliers did not make a difference. H2 proposed that trainees have higher SA-levels in the low TPC compared to the high TPC. H2 could not be confirmed (see Table 2). Other than assumed, SA-levels were significantly higher in the high TPC compared to the low TPC. H3 proposed that trainees perform higher in the low TPC compared to the high TPC. H3 could not be confirmed, however the results of the t-test showed significantly higher performance scores in the high TPC compared to the low TPC (see Table 2).

In summary, the result of the t-test confirmed higher subjective stress-levels in the high TPC. Also, other than assumed, higher SA-levels and performance scores were found in the high TPC.

3.2. Model 2: Differences Between Conditions as a Result of Differences in Stress-Levels

Prior to hypotheses testing of H4 and H5, assumptions for repeated measures ANCOVAs (rmANCOVAs) were tested and could be confirmed after the exclusion of outliers. A detailed description is attached in Appendix H. Hypothesis H4a proposed that trainees' subjective stress-levels determine lower SA-levels in the high TPC. The results of the rmANCOVA could not confirm H4a (see Table 3). There was neither an interaction between the level of TP and subjective stress of the low TPC, nor between the level of TP and the difference score of subjective stress, which means that differences in SA-levels as a result of TP did not depend on either of the two indicators of subjective stress.

Table 3

Results of rmANCOVA on the Effect of Subjective Stress for Differences in SA-Levels Between Conditions

Measure	$F(1, 44)$	p	η2
Level of TP	4.423	.041	.091
Level of TP * substress low TPC [a]	0.188	.667	.004
Level of TP * substress DS [b]	0.642	.427	.014

Note. N = 47.

[a] substress low TPC = centred substress of the low TPC (baseline level of substress)

[b] substress DS = centred difference score of substress (indicator of change between TP conditions)

Hypothesis H4b proposed that trainees' physiological stress-levels determine lower SA-levels in the high TPC. The results of the rmANCOVA could not confirm H4b (see Table 4). There was no significant difference in SA-levels between conditions. Also, there was neither an interaction between the level of TP and physiological stress of the low TPC, nor between the level of TP and the difference score of physiological stress. This means that differences in SA-levels as a result of TP were not significant and not determined by either of the two indicators of physiological stress.

Table 4

Results of rmANCOVA on the Effect of Physiological Stress for Differences in SA-Levels Between Conditions

Measure	*F*(1, 36)	*p*	η2
Level of TP	3.074	.088	.079
Level of TP * phystress low TPC [a]	0.039	.845	.001
Level of TP * phystress DS [b]	1.943	.172	.051

Note. N = 39.

[a] phystress low TPC = centred phystress of the low TPC (baseline level of phystress)

[b] phystress DS = centred difference score of phystress (indicator of change between TP conditions)

Hypothesis H5a proposed that trainees' subjective stress-levels determine lower performance in the high TPC. The results of the rmANCOVA could not confirm H5a (Table 5). There was neither an interaction between the level of TP and subjective stress of the low TPC, nor between the level of TP and the difference score of subjective stress. This means that differences in performance as a result of TP did not depend on either of the two subjective stress indicators.

Table 5

Results of rmANCOVA on the Effect of Subjective Stress for Differences in Performance Between Conditions

Measure	*F*(1, 45)	*p*	η2
Level of TP	10.598	.002	.191
Level of TP * substress low TPC [a]	0.010	.922	.000
Level of TP * substress DS [b]	0.000	.995	.000

Note. N = 48.

[a] substress low TPC = centred substress of the low TPC (baseline level of substress)

[b] substress DS = centred difference score of substress (indicator of change between TP conditions)

Hypothesis H5b proposed that trainees' physiological stress-levels determine lower performance in the high TPC. The results of the rmANCOVA could not confirm H5b (Table 6). There was neither an interaction between the level of TP and physiological stress of the low TPC, nor between the level of TP and the difference score of physiological stress. This means that differences in performance as a result of TP did not depend on either of the two physiological stress indicators.

Table 6

Results of rmANCOVA on the Effect of Physiological Stress for Differences in Performance Between Conditions

Measure	$F(1, 37)$	p	η2
Level of TP	7.862	.008	.175
Level of TP * phystress low TPC [a]	0.145	.705	.004
Level of TP * phystress DS [b]	0.105	.748	.003

Note. N = 40.

[a] phystress low TPC = centred phystress of the low TPC (baseline level of phystress)

[b] phystress DS = centred difference score of phystress (indicator of change between TP conditions)

In summary, differences in SA-levels as a result of TP did not depend on subjective or physiological stress. The same applies to differences in performance as a result of TP which were also not determined by subjective or physiological stress.

3.3. Model 3: The Influence of Stress on SA and Performance Within Conditions

The third model examined the influence of stress on SA and performance within conditions. For this purpose, the assumptions for multiple linear regression analyses were tested. Partial regression diagrams yielded unsatisfactory results for the relationship between physiological stress and both outcome variables. Precisely, data points gathered at the lower end of the physiological stress scale and thus hardly showed a relationship between physiological stress and SA or performance in both conditions. Therefore, a logarithmic transformation for physiological stress was conducted. As a result, all assumptions could be confirmed. Appendix I includes a detailed description of the assumption testing.

H6a proposed that subjective and physiological stress have a negative effect on trainees' SA-levels in the low TPC. Multiple linear regression analysis could not confirm H6a. The R^2 for the overall model was .019 (adjusted $R^2 = -.020$), indicative for a small amount of explained variance according to Cohen (1988), however only for R^2. Subjective and physiological stress were not able to significantly predict SA-levles, $F(2, 50) = 0.490$, $p = .616$. Moreover, both regression coefficients were not significant (see Table 7).

Table 7

Results of Multiple Linear Regression Analysis for the Influence of Stress on SA in the Low TPC

						95% Confidence Interval	
Coefficients	*B*	*SE*	β	*T*	*p*	Lower	Upper
Constant	9.246	0.316		29.224	<.001	8.610	9.881
Substress LTP	0.001	0.014	.008	0.059	.953	−0.027	0.029
Phystress LTP	−0.154	0.156	−.140	−0.986	.329	−0.468	0.160

Note. N = 53.

As consequence of assumption testing, it was decided to reconduct the analysis after excluding an outlier. However, removing the outlier did not make a difference.

H6b proposed that subjective and physiological stress have a negative effect on trainees' SA-levels in the high TPC. Multiple linear regression analysis could not confirm H6b. The R^2 for the overall model was .001 (adjusted R^2 = −.039), showing no variance explanation according to Cohen (1988). Subjective and physiological stress were not able to significantly predict SA, $F(2, 50) = 0.034$, $p = .967$. Moreover, both regression coefficients were not significant (see Table 8).

Table 8

Results of Multiple Linear Regression Analysis for the Influence of Stress on SA in the High TPC

						95% Confidence Interval	
Coefficients	*B*	*SE*	β	*T*	*p*	Lower	Upper
Constant	9.867	0.289		34.107	<.001	9.286	10.448
Substress HTP	0.000	0.008	−.008	−0.059	.954	-0.017	0.016
Phystress HTP	0.035	0.137	.037	0.260	.796	-0.239	0.310

Notes. N = 53.

As consequence of assumption testing, it was again decided to reconduct the analysis after excluding an outlier. However, removing the outlier did not make a difference.

H7a proposed that subjective and physiological stress have a negative effect on trainees' performance in the low TPC. Multiple linear regression analysis could not confirm H7a. The R^2 for the overall model was .057 (adjusted R^2 = .019), indicative for a small amount of explained variance according to Cohen (1988). Subjective and physiological stress were not able to significantly predict performance, $F(2, 50) = 1.501$, $p = .233$. Moreover, both regression coefficients were not significant (see Table 9).

Table 9

Results of Multiple Linear Regression Analysis for the Influence of Stress on Performance in the Low TPC

						95% Confidence Interval	
Coefficients	*B*	*SE*	β	*T*	*p*	Lower	Upper
Constant	23.303	1.807		12.893	<.001	19.673	26.934
Substress LTP	0.131	0.080	.228	1.642	.107	−0.029	0.292
Phystress LTP	−0.703	0.893	−.109	−.787	.435	−2.497	1.091

Note. N = 53.

H7b proposed that subjective and physiological stress have a negative effect on trainees' performance in the high TPC. Multiple linear regression analysis could not confirm H7b. The R^2 for the overall model was .084 (adjusted R^2 = .047), indicative for a small amount of explained variance according to Cohen (1988). Subjective and physiological stress were not able to significantly predict performance, $F(2, 50) = 2.286$, $p = .112$. However, subjective stress only could significantly predict performance in the high TPC, whereas the effect of physiological stress on performance in the high TPC was not significant (see Table 10).

Table 10

Results of Multiple Linear Regression Analysis for the Influence of Stress on Performance in the High TPC

						95% Confidence Interval	
Coefficients	*B*	*SE*	β	*T*	*p*	Lower	Upper
Constant	26.295	1.890		13.911	<.001	22.498	30.091
substress HTP	0.112	0.055	.281	2.046	.046	0.002	0.222
phystress HTP	−0.859	0.893	−.132	−0.961	.341	−2.653	0.936

Note. N = 53.

In summary, subjective and physiological stress did not predict SA-levels or performance in either TP condition. Overall, performance as outcome variable appeared to be a better fit for the regression model than SA. Subjective stress only was found to positively predict performance in the high TPC.

4. Discussion

4.1. Interpretation of Findings

The imposition of TP did not produce significant results for differences in physiological stress between conditions and yielded inverse (i.e., positive) results for SA and performance. Furthermore, neither subjective nor physiological stress determined a change in SA or performance between conditions or predicted SA or performance within conditions. Therefore, possible explanations for the non-significant and inverse results are discussed based on theoretical models and previous research.

4.1.1. Model 1: Differences Between Conditions

The first model examined differences in stress, SA and performance between the low and the high TPC. As expected in H1a, trainees had significantly higher subjective stress-levels in the high TPC compared to the low TPC. This means that trainees *appraised* the induced TP as stressful and *perceived* themselves as under stress (Cohen et al., 1997). This finding is in line with previous questionnaire-based research identifying TP as one of the root causes for work stress in the construction industry (e.g., Aksorn & Hadikusumo, 2007; Campbell, 2006; Pegler, 2021).

H1b stated that higher stress-levels in the high TPC can also be detected by skin conductance as indicator of physiological stress. Although the differences in skin conductance between conditions showed a tendency towards higher levels in the high TPC, the difference was not significant. The non-significant result for physiological stress, in the presence of a significant result for subjective stress, can have different reasons. Most likely, the situational demands of the high TPC as a combination of the dynamic work environment, the task of asphalt compaction and the induction of TP were not intense or realistic enough to evoke a physiological response, i.e., a rise in the SCL (Turpin & Grandfield, 2007). Although physiological stress is intertwined with subjective stress (Sommerfeldt et al., 2019), the missing effect for physiological stress does not contradict the significant difference in subjective stress. Subjective stress measurement is generally affected by response styles that result from different factors, such as cognition, social context and social desirability (Kompier, 2005). Physiological stress, in contrast, represents objective data. This means that the measurement of physiological stress is not distorted by individual response styles as subjective stress is. The legitimacy of the deviating results of the two stress indicators is also supported by previous studies which found that subjective stress and physiological stress are affected by different factors (Föhr et al., 2015) and are not necessarily correlated (e.g., Becker et al., 2022). Moreover, the different results in subjective and physiological stress underscore the rationale for examining multiple stress indicators in order to thoroughly assess and discuss the actual impact of an acute stressor in VR. In view of this study, it can therefore not be conclusively clarified whether trainees actually experienced more stress. However, the significant rise in subjective stress in the high TPC in combination with the higher (but non-significant) physiological stress-levels in the high TPC can be appraised as a strong indicator for an overall higher stress experience in the high TPC as a result of the imposed TP.

The results for H2 and H3 could not confirm higher SA-levels and performance scores in the low TPC but showed significant results for the opposite direction. Contrary to expectations, trainees had higher SA-levels and performance scores in the high TPC. Apparently, the induced TP had a positive effect on SA and performance. Although these results were not expected, they can be explained by an interplay of different factors. First, participants' comments during task execution indicated that the VR training simplified the virtual construction site and the compaction task. The simplification comprises the omission of elements in the environment that make out the dynamic, diverse (e.g., Albert et al., 2014) and dangerous (e.g., Kim et al., 2022) nature of construction as well as the fact that mistakes in the compaction process do not result in major consequences for further task execution (e.g., collision with roadblocks would not result in "damage" with impairing consequences). How the characteristics of a VR training influence the learning outcomes was shown in a study that compared VR training to traditional on-the-job training for assembly and maintenance work (Wolfartsberger et al., 2022). Participants in the VR training perceived the task as less cognitively strenuous and achieved lower performance scores compared to traditional on-the-job training. It was concluded that the simplified representation of work steps in the VR environment did not sufficiently challenge users cognitively. This is in line with the Yerkes-Dodson Law which states that low activation levels lead to low performance (Yerkes & Dodson, 1908, as cited in Van Veldhoven, 2014). With respect to the present study, this implies that the situational demands might have not been as demanding as assumed, resulting in low activation levels in the low TPC where trainees also had lower SA-levels and performance scores. Apparently, the imposition of TP led to more optimal activation levels, which, in turn, were beneficial for SA and performance. Moreover, the simplification required trainees to consider fewer elements during task execution. It can thus be assumed that trainees perceived less cognitive load than expected since they required less prior knowledge and skills for conducting the task (Ayres & Paas, 2012; Kolfschoten & Brazier, 2013).

Second, the imposition of TP might have had a beneficial effect on trainees' motivation. Schmitt et al. (2015) investigated boundary conditions of the inverted U-shape relation between TP and work engagement and argued that this relation is influenced by a legitimate or illegitimate perception of the work tasks. The latter comprised unreasonable or unnecessary tasks and is associated with negative feelings, such as injustice or threat. Their findings demonstrated that there is indeed a motivating effect of a moderate level of TP when employees are assigned to tasks that are rated low on unreasonability. With respect to the present study, it is likely that the induced TP was assessed both moderate and legitimate since it rooted in the task itself. Baethge et al. (2018) tested time-exposure effects of TP and found that a short-term exposure of TP was associated with beneficial effects, whereas a long-term exposure of TP reduced work engagement. They concluded that short-term increases of TP may serve as motivating factor.

Third, in addition to the low level of activation and cognitive load in the low TPC, as well as the motivational effect of TP due to its moderate level, legitimacy and its short-term exposure, the positive effect of TP on construction trainees can be explained by the Challenge-Hinderance Stressor Framework (CHSF). The CHSF, similar to the PEF theory, is a theoretical model that is used to understand the effects of workplace stressors, and which is widely accepted in both the academic and practical context (Horan et al., 2020; Lepine, 2022). The CHSF assumes that stressors can be divided into two types. *Hindrance stressors* cause negative affect since they are appraised as detrimental to personal growth and goal achievement. They often comprise obstacles at an organizational level (e.g., role ambiguity, conflict, bureaucratic procedures or job insecurity) and are therefore difficult to overcome through individual efforts (Yang & Li, 2021). *Challenge stressors*, on the other hand, describe work characteristics that evoke positive affect and promote personal growth (Cavanaugh et al., 2000, as cited in Yang & Li, 2021). In this sense, individuals assess a stressor, such as TP, workload or work complexity, as an opportunity for growth, learning and goal achievement (Cavanaugh et al., 2000, as cited in Yang & Li, 2021). While both hindrance and challenge stressors require extra effort and can have harmful effects on employees' psychophysiological health (Wu et al., 2019), challenge stressors can be overcome through individual efforts (Yang & Li, 2021). As a result, employees have a stronger learning motivation, actively respond to the challenge (Spreitzer & Porath, 2013, as cited in Yang & Li, 2021) and thrive at work (Wu et al., 2019). For example, Lepine et al. (2005) used a meta-analytic approach to study the relationship between hindrance and challenge stressors and strains, motivation, and performance. Their results demonstrated that hindrance stressors had a negative direct effect on performance, as well as a negative indirect effect on performance through strains and motivation. Challenge stressors, in contrast, had a positive direct effect on performance as well as a negative indirect effect on performance through strains, but a positive indirect effect on performance through motivation. On the basis of the PEF theory and in view of the inherently demanding working conditions in construction (e.g., Hasanzadeh et al., 2018), the negative effect of TP as shown in previous research (e.g., Sneddon et al., 2013; Stavroula et al., 2003), and the limited experience of trainees, it was reasonable to expect that TP adversely affects SA and performance. However, the results indicate the opposite, namely that TP evoked a eustress response (i.e., a positive stress response). With respect to the CHSF, the induced TP thus acted as a challenge stressor and although it might have evoked strain to some extent, it also had an energizing effect and is likely to have promoted feelings of accomplishment, growth and development (Horan et al., 2020).

4.1.2. Model 2: Differences Between Conditions as a Result of Differences in Stress-Levels

The second model, including H4 and H5, investigated whether differences in SA-levels and performance between conditions can be explained by differences in stress-levels. It was hypothesized that trainees' subjective and physiological stress-levels determine lower SA-levels and lower performance scores in the high TPC. With respect to the unexpected *positive* effect of TP on SA-levels and performance found in the first model, stress-levels would rather determine *higher* SA-levels and *higher* performance scores in the high TPC.

However, the results of the rmANCOVAs could not confirm that stress-levels explain differences in SA-levels or performance between conditions. This applies to both the baseline level of stress and the difference score of stress (i.e., the indicator of change). These results are unexpected as the first model could confirm differences in subjective stress, SA and performance between conditions. However, as a result of the rmANCOVAs, it can be concluded that the whole-group differences found for stress, SA and performance between conditions do not translate to relationships at the individual level.

With respect to physiological stress, the Empatica outputs generally showed a low tonic phase for SCLs in both conditions and paired samples t-tests for the first model could not confirm a significant difference in physiological stress between conditions. Consequently, SCLs as well as changes in SCLs between conditions were probably too low to explain differences in SA-levels or performance as a result of TP. The low levels of physiological stress in the low TPC, which served as baseline stress indicator in the model, can most likely be explained by the overall low situational demands of the VR training that not fully replicated the dynamic and dangerous elements typically found in construction (e.g., Albert et al., 2014; Hasanzadeh et al., 2018; Health and Safety Executive, n.d.). Moreover, the low levels in difference scores, which served as indicator of change between conditions, root in the induced TP. This means, that the induced TP was not intense enough to evoke a significant rise in SCLs in the high TPC (e.g., Turpin & Grandfield, 2007).

Subjective stress-levels, in contrast, showed significant differences between conditions. The reason for the missing effect in this case can probably be explained by the PSQ demands scale which might not have sufficiently measured stress as a result of the imposed TP. Precisely, Crosswell and Lockwood (2020) emphasized the importance to choose an adequate stress measure for the characteristics of the type of stressor and stress. The PSQ was overall assessed to be an adequate instrument, however it did not fully match the context and thus required some adaptions. The fact that item four had to be excluded to achieve a reliability of Cronbach's alpha = .70 already indicated that the PSQ demands scale might not have sufficiently assessed the acute stress reaction as a result of the induced TP. However, this does not apply to SA and performance, which in contrast, did show differences as a result of TP. Consequently, it is likely that changes in subjective stress-levels and changes in SA and performance do not depend to the same extent on the induced TP what could explain the missing effect.

4.1.3. Model 3: The Influence of Stress on SA and Performance Within Conditions

The third model examined the effect of subjective stress and physiological stress on SA and performance within conditions. Specifically, multiple linear regression analyses were used to partition variance in SA and performance, i.e., to investigate which part of the variance can be explained by the results of the PSQ demands scale and which part can be explained by the SCL. However, it was found that subjective and physiological stress did not significantly predict SA in either TP condition. Likewise, subjective and physiological stress did not significantly predict performance in either TP condition. However, subjective stress *only* could positively predict performance in the high TPC. The positive direction of the effect again indicates that trainees experienced a eustress response, rather than the expected distress response.

The missing effects in the multiple linear regression analyses are unexpected, but not surprising given the fact that stress-levels did not determine changes in SA-levels or performance as a result of TP in the second model. The reasons for non-significant results are actually likely to be similar to those for the second model which investigated differences between conditions as a result of differences in stress-levels: The demands scale for subjective stress measurement might have not sufficiently assessed stress as a result of the induced TP which might have distorted the relationship between subjective stress and the outcome variables. Moreover, as a consequence of low situational demands, the low SCLs were probably insufficient to predict SA or performance.

It should be noted here that this study had sound reasons for choosing electrodermal activity, specifically the SCL, as an indicator of physiological stress. However, electrodermal activity is primarily an indicator of changes in sympathetic arousal as a result of emotional and cognitive states (Braithwaite et al., 2015) and can provide insights into a variety of physiological and psychological states that go beyond stress (Empatica, 2022). With respect to the unexpected positive results for both SA and performance in the high TPC as well as the positive effect of subjective stress on performance found in the third model, it is worthwhile to also consider electrodermal activity as an indicator of student engagement. Student engagement roughly describes students' involvement, interest and motivation in a task and is commonly measured at three levels: the emotional, cognitive and behavioral level (McNeal et al., 2020). There is ample evidence that student engagement can be operationalised by electrodermal activity (i.e., skin conductance). For example, skin conductance has proven to be a viable real-time indicator to identify student engagement during lectures (Di Lascio et al., 2018). Moreover, skin conductance was found to be useful as a measure of real-time student engagement among undergraduate students (McNeal et al., 2020). The study revealed that higher SCLs were positively correlated with high-engaging classroom activities (e.g., group work activities) and were consistent with non-significant trends of increased student performance and increased self-reports of engagement. The authors assumed that the non-significance can be explained by the high standard deviations in the data set. Student engagement is of interest in the context of this study from different points of view. First, previous research found that student engagement is closely and positively related to performance (e.g., Bakker et al., 2015; Delfino, 2019). Second, there has been much work to show that immersive VR training, offering experiential and situated learning, promotes student engagement (e.g., Di Natale et al., 2020; Huang et al., 2020). Third, and most relevant with respect to the unexpected positive results for SA and performance in the high TPC, previous research has shown that higher challenge increases student engagement. For example, a study on game-based learning found that higher challenge in the game increased student engagement (i.e., concentration, interest, and enjoyment), which in turn, positively affected learning (Hamari et al., 2016). With respect to this study, this means that student engagement may have differed due to the induced TP, which was perceived as a stimulating challenge due to the low situational demands, resulting in higher levels of SA and performance in the high TPC. Consequently, it is likely that SCLs measured in this study are (partly) related to student engagement.

4.2. Limitations

A strong focus of this study was on the effects of subjective and physiological stress induced by TP. However, the PSQ demands scale for measuring subjective stress might have not sufficiently fitted the context. Most notably is the misinterpretation of item four, asking trainees whether they had enough time. Apparently, trainees did not rate this question with respect to the induced TP, but with respect to the duration of the scenarios. Although item four was excluded, this misunderstanding limits the conclusions for the significant difference in subjective stress between conditions since it is likely that the remaining items have also been affected by this misunderstanding to some degree. For example, item five asks trainees whether they felt under TP. Similar to item four, it is possible to apply this item to the duration of the VR scenarios. However, in contrast to item four, the remaining items yielded satisfactory scores in item-scale-statistics and the difference found for stress-levels between conditions was highly significant. Therefore, the results for subjective stress-levels can still be seen as a strong indicator that the induced TP did evoke a rise in subjective stress-levels.

A second limitation of this study is the simplification of the VR environment and the task. The current version of the VR simulation covered the required elements to meaningfully execute the task of asphalt compaction. However, elements that contribute to the perception of a dynamic, diverse (e.g., Albert et al., 2014) and dangerous (e.g., Kim et al., 2022) working environment, as is the case on a real construction site, were neglected. Likewise, mistakes did not result in major implications for the further execution of the task. Consequently, the VR scenarios, including the working environment, the task and the imposed TP in the high TPC, were not able to sufficiently affect cognitive and emotional states that would result in changes in the sympathetic arousal and thus in more conclusive SCLs (Braithwaite et al., 2015). Therefore, the conclusions regarding physiological stress-levels are limited, as the two scenarios did not sufficiently reflect the real situational demands.

4.3. Theoretical Implications

The present study adds value to current research in the construction industry on different levels. First, whereas previous studies in the context of construction are mainly questionnaire-based (e.g., Nepal et al., 2006; Zhang et al., 2023), the present study used an experimental study design based on a VR simulation that automatically recorded data for SA-levels and performance. Also, while performance measurement in previous studies is often based on subjective assessments (e.g., Nepal et al., 2006), the present study utilised performance measurement on an objective level and could demonstrate two advantages of this approach from which future studies can benefit: First, automatic data collection required fewer human resources, i.e., trained observers (Cornell et al., 2007), to collect objective data that can be used over and over again once programmed. Second, the objective approach offered the advantage of more accurate and less biased assessments (Langevin & Mendoza, 2021).

Second, the findings of this study underscore the importance of combining subjective and physiological stress measurement. This is based on the fact that stress questionnaires often take a retrospective approach (e.g., Weckesser, 2019) and are therefore prone to bias and error (Crosswell & Lockwood, 2020). In the present study, the results of the subjective stress reports were complemented by real-time data acquisition of skin conductance which served as physiological stress indicator and which represented the objective stress experience during the event. Moreover, stress questionnaires generally measure the conscious part of the stress experience (e.g., Fliege et al., 2009). Physiological stress measurement also addresses the unconscious part of a stress experience (Epel et al., 2018) that cannot be directly affected by human will (McCorry, 2007). For the present study, the dual approach revealed deviating results for the two stress indicators, highlighting the importance to examine multiple indicators of stress when investigating the effect of stressors and stress in VR.

Lastly, the present study contributes to the growing interest in the use of VR training in construction (e.g., Wang et al., 2018) and has recognized the need to investigate innovative training methods for the vocational training of construction workers (BMBF, 2020d). In this context, this study explicitly focused on inducing TP, as an acute stressor, using the safe boundaries of VR. In light of the increasing number of studies on consequences of demanding working conditions in construction (e.g., Hasanzadeh et al., 2018) and on the potentials of VR training in the construction industry (Wang et al., 2018), the present study is, to my knowledge, the only study that examined trainees' SA and performance as an outcome of acute TP. An important implication of this study is that the characteristics of the target group, which include little prior knowledge and limited experience, do not carry much weight in terms of a potential distress response when the situational demands are simplified and thus decrease the general mismatch between demands and abilities. This is in line with previous findings (e.g., Wolfartsberger et al., 2022). In this case, the imposition of an acute stressor (TP) might lead to opposite (i.e., positive) results as would be expected on a real construction site.

4.4. Practical Implications

Overall, the use of VR training to expose trainees to occupational stressors would be a valuable addition to current vocational training for sectors of inherently dynamic and complex working conditions, but some aspects should be considered in the design and implementation. First, VR training provides an innovative, mistake-tolerant and safe opportunity to train procedures under different conditions without the risk of injury or damage (e.g., Wang et al., 2018). This means that trainees can learn how to deal with stressors before they encounter them in real life. Also from the trainees' viewpoint, the integration of VR training is appraised both meaningful and motivating as was shown in a brief survey among participants of this study. Almost all participating trainees (94.34%) rated the integration of VR into their vocational training as "meaningful" and reported that they would be "motivated" to participate in VR training.

However, as an implication of this study, it should be thoroughly elaborated which stressors are relevant in a certain job and how these stressors can be implemented best. For this purpose, it is advised to consult subject-matter experts and to integrate sufficient iterations in the design process to test and evaluate the effect of the induced stressor. In this context, it is also important to replicate the real working environment as realistic as possible, i.e., to add sufficient details (and if possible, even sounds) that convey a realistic feeling of the dynamics in and complexity of the working environment. For the construction industry, in particular, this includes moving and reversing construction vehicles, workers walking on site, poor separation of pedestrians or cars and construction vehicles, as well as strenuous lightning conditions after sunset (Health and Safety Executive, n.d.). For a feeling of actually being present (high immersion), it is also advisable to use an HMD, physical equipment (e.g., a steering wheel) and sound (Dhimolea et al., 2022). Although the real working environment should be realistically replicated, the integration of elements should be carefully planned and gradually adjusted to achieve an appropriate level of difficulty with respect to trainees' knowledge and skills. For example, a task that is either too easy or too difficult would be harmful for trainees' motivation (Gallego-Durán et al., 2016). At the same time, research indicates that easy tasks stimulate short term engagement, i.e., not leaving the task immediately, whereas difficult tasks stimulate long term engagement, i.e., prolonged interaction with a system, and strengthen learning (Papoušek et al., 2016). Consequently, it is again important to consult subject-matter experts to identify contextual elements and tasks that would not overstrain and demotivate trainees in the beginning as well as more challenging elements that gradually stimulate long term engagement and strengthen learning. A more advanced but also more expensive version of VR training could automatically customise the difficulty level based on the results of an integrated questionnaire on prior knowledge and skills and depending on the continuous assessment of the trainees' performance during task execution.

Another implication of this study is that trainees should have a certain level of relevant prior knowledge to meaningfully engage in VR training. Trainees who have not gained the relevant knowledge for executing construction tasks, cannot apply this knowledge in a meaningful way and thus do not fully exploit the potentials of the VR simulation. For example, if trainees do not know yet when to use the vibration function and why it is relevant, they will probably not focus on this aspect of the task. If such omission of task-relevant elements has no noticeable effect on further task processing or task quality, this simplification by the trainees can impair the learning effect (Wolfartsberger et al., 2022). Therefore, it is advisable to instruct trainees in advance or to integrate prompts in the simulation that remind them to consider all relevant task elements.

4.5. Future Research

First, future research should build on this study and conduct integrated model testing to obtain even more informative results about the relationship between TP and outcome variables, different stress indicators and different levels of TP. In this context, future research should also investigate the effect of SA on performance, which was not in the scope of this study, but which is a key component in Endsley's model of situation awareness in complex decision making (Endsley, 1995). Moreover, although this study took into account the characteristics of the target group and related them to the PEF theory, they were not included in the statistical model. To gain insights into the extent to which the effects of TP and stress on outcome variables depend on the level of prior knowledge and experience, future research should integrate these as covariates. Also, the positive effect of TP on SA-levels and performance suggests that TP elicited motivation which was similarly found in previous research (e.g., Baethge et al., 2018; Schmitt et al., 2015). Therefore, including motivation into the model could yield additional insights into the dynamics between TP, stress and outcome variables in VR. Likewise, the positive results for SA and performance in the high TPC suggest that the SCL was, at least to some degree, an indicator of student engagement (e.g., Bakker et al., 2015; Delfino, 2019; Hamari et al., 2016). However, future research needs to explore this further to investigate which part of the variance in skin conductance is explained by stress and which by student engagement.

Second, to evaluate performance, this study used both a qualitative indicator (the number of crossings at each measurement point which should not exceed four) and a quantitative indicator (the total number of crossings within six minutes). Although this dual view is considered a strength, future research should consider a more comprehensive approach to qualitative indicators in performance measurement. For example, the rolling pattern, which also determines the quality of the street (Deutscher Asphaltverband e.V., 2016), should also be taken into account when assessing performance.

Third, future research should consider the use of multiple physiological stress indicators to gain a more nuanced understanding of the body's response to stress. The simultaneous investigation of multiple stress indicators is a quite common approach in stress research, since different physiological indicators provide specific information about the body's stress response (e.g., Crosswell & Lockwood, 2020; Marques et al., 2010). For example, the cortisol-level is a stress hormone that provides insight into the hypothalamic-pituitary-adrenal (HPA) axis response to stress (Noushad et al., 2021) which has also been used in studies that explicitly measured acute stress. Boucher and Plusquellec (2019) found that the excess proportion of cortisol load related to the stress-level expected from the characteristics of the task (e.g., novelty and unpredictability) as well as to the perceived stress-level. Moreover, previous research suggests that cortisol is a valid stress indicator in simulated learning (McGuire et al., 2018). However, the measurement of cortisol is typically done by using substrates, such as blood, saliva or urine and is thus less convenient in a school setting than a wearable device as is the Empatica. Another physiological indicator for stress research is the heart rate variability (HRV), which can also be conveniently assessed with the Empatica. HRV provides information about the autonomic nervous system response to stress and was found to be an appropriate indicator for the physiological assessment of stress (Kim et al., 2018). For example, HRV was found to be an insightful stress indicator of acute stress among novice airway mangers (Mefford et al., 2019), indicating its suitability for acute stress measurement and little experienced employees.

Forth, similar to the questionnaire- and physiological-based assessment of stress, the question-based assessment of SA on the basis of SAGAT could be complemented by a physiological measurement to receive a more complete picture of the SA-level. Zhang et al. (2020) conducted a systematic review on physiological measurements of SA and found a frequent use and a high potential of eye tracking techniques. Precisely, correlations between eye movement and direct SA scores could be observed. In addition, eye tracking could reveal dysfunctional attentional allocations for placed hazards on a construction site (Hasanzadeh et al., 2018). An alternative, but less frequently used approach for physiological SA measurement are electroencephalography (EEG) studies that found that EEG was related to changes in SA. Moreover, previous research found evidence for real-time assessment and discrimination of SA-levels by investigating EEG features (Feng et al., 2022). However, it should be noted that the effort and costs of these measurements are much higher compared to physiological measures of stress described earlier.

Fifth, besides the influence of stressors, Endsley (1995) proposes further task and environmental factors which affect SA and performance, such as task complexity. Construction is known for its complex working environment (e.g., Albert et al., 2014), including moving and reversing construction vehicles, workers walking on site, partly poor separation of pedestrians and vehicles and strenuous lightning conditions after sunset (Health and Safety Executive, n.d.). Depending on whether and to which degree these components are present, construction workers are confronted with different levels of task complexity. Experimental study design using VR can provide a safe frame to investigate the influence of these environmental attributes on SA and performance. The results should be considered when designing training scenarios for construction trainees.

5. Conclusion

Despite its high quality, the German VET system must continue to evolve to ensure that Germany remains competitive in a fast-changing and globalised world, notably in view of the shortage of skilled labour. The present study could demonstrate that VR training is a safe and motivating training opportunity with much potential for the training of prospective construction workers as well as for trainees in other professional sectors, which are characteristic for dynamic and dangerous working environments. In this context, the present study specifically focused on the induction of TP as an acute stressor into VR training for construction trainees and revealed partly unexpected results.

On the basis of the PEF theory and an expected demands-ability mismatch, it was assumed that the imposition of TP would increase stress-levels and decrease SA-levels and performance. However, whereas TP indeed increased stress-levels, although only the increase in subjective stress was significant, it also increased SA-levels and performance. This unexpected positive effect of TP suggests a eustress response, rather than a distress response. The reason for the opposite effect of TP probably roots in the attributes of the VR training. Although the VR training, on the basis of HMD-technology, allowed for high immersion and covered all required elements to meaningfully execute the construction task, it neglected elements in the environment that typically make out the dynamic and dangerous working environment in construction and left out impairing consequences of mistakes for further task execution. This simplification of the situational demands might have caused low activation levels in trainees, so that the imposition of TP might have led to more optimal activation levels and increased motivation. Overall, the research question "*Does VR training for construction trainees reveal negative effects of time pressure and stress on situation awareness and performance?"* thus must be negated within the limitations of this study.

The study contributes to current research in the context of VR and construction by explicitly focusing on the induction of acute TP into VR training for construction trainees who are trained on the basis of the German VET-system. Future research should build on this study with integrated model testing and by further investigating the role of simplified situational demands with respect to prior knowledge and experience as well as the role of motivation in the dynamics of TP, stress and outcome variables. The study has to accept limitations for the results of subjective stress-levels and the overall low levels of physiological stress.

6. References

Abidi, M. H., Al-Ahmari, A., Ahmad, A., Ameen, W., & Alkhalefah, H. (2019). Assessment of virtual reality-based manufacturing assembly training system. *The International Journal of Advanced Manufacturing Technology*, *105*, 3743–3759. https://doi.org/10.1007/s00170-019-03801-3

Abu Oda, M. M. A., Tayeh, B. A., Alhammadi, S. A., & Abu Aisheh, Y. I. (2022). Key indicators for evaluating the performance of construction companies from the perspective of owners and consultants. *Results in Engineering*, *15*, Article 100596. https://doi.org/10.1016/j.rineng.2022.100596

Agrawal, S., & Peeta, S. (2021). Evaluating the impacts of situational awareness and mental stress on takeover performance under conditional automation. *Transportation Research Part F: Traffic Psychology and Behaviour*, *83*, 210–225. https://doi.org/10.1016/j.trf.2021.10.002

Agrawal, S., Simon, A., Bech, S., Bærentsen, K., & Forchhammer, S. (2020). Defining immersion: Literature review and implications for research on audiovisual experiences. *Journal of the Audio Engineering Society*, *68*(6), 404–417. https://doi.org/10.17743/jaes.2020.0039

Aksorn, T., & Hadikusumo, B. H. W. (2007). The unsafe acts and the decision-to-err factors of Thai construction workers. *Journal of Construction in Developing Countries*, *12*(1), 1–25. https://www.researchgate.net/publication/43245358_The_Unsafe_Acts_and_the_Decision-to-Err_Factors_of_Thai_Construction_Workers

Albert, A., Hallowell, M. R., & Kleiner, B. M. (2014). Enhancing construction hazard recognition and communication with energy-based cognitive mnemonics and safety meeting maturity model: Multiple baseline study. *Journal of Construction Engineering and Management*, *140*(2), Article 04013042. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000790

Ambrosini, V., & Bowman, C. (2001). Tacit knowledge: Some suggestions for operationalization. *Journal of Management Studies*, *38*(6), 811–829. https://doi.org/10.1111/1467-6486.00260

Ayres, P. & Paas, F. (2012). Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, *26*(6), 827–894. https://doi.org/10.1002/acp.2882

Baethge, A., Vahle-Hinz, T., Schulte-Braucks, J., & Dick van, R. (2018). A matter of time? Challenging and hindering effects of time pressure on work engagement, *Work & Stress*, *32*(3), 228–247. https://doi.org/10.1080/02678373.2017.1415998

Bakker, A. B., Sanz Vergel, A. I., & Kuntze, J. (2015). Student engagement and performance: A weekly diary study on the role of openness. *Motivation and Emotion*, *39*, 49–62. https://doi.org/10.1007/s11031-014-9422-5

BauWiAusbV [Verordnung über die Berufsausbildung in der Bauwirtschaft], Bundesgesetzblatt 28, 1102–1235. (1999). https://www.bibb.de/dienst/berufesuche/de/index_berufesuche.php/regulation/bauwirtschaft_1999.pdf

Becker, S., Spinath, B., Ditzen, B., & Dörfler, T. (2022). Psychological stress = physiological stress? An experimental study with prospective teachers. *Journal of Psychophysiology*, *37*(1), 12–24. https://doi.org/10.1027/0269-8803/a000301

Boucher, P., & Plusquellec, P. (2019). Acute stress assessment from excess cortisol secretion: Fundamentals and perspectives. *Frontiers in Endocrinology*, *10*, Article 749. https://doi.org/10.3389/fendo.2019.00749

Bowman, D. A., & McMahan, R. P. (2007). Virtual reality: How much immersion is enough? *Computer*, *40*(7), 36–43. https://doi.org/10.1109/MC.2007.257

Bracq, M.-S., Michinov, E., Le Duff, M., Arnaldi, B., Gouranton, V., & Jannin, P. (2021). Training situational awareness for scrub nurses: Error recognition in a virtual operating room. *Nurse in Education in Practice*, *53*, Article 103056. https://doi.org/10.1016/j.nepr.2021.103056

Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2015). *A guide for analysing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. Revised version: 2.0.* https://www.birmingham.ac.uk/documents/college-les/psych/saal/guide-electrodermal-activity.pdf

Bregenzer, A., & Jimenez, P. (2021). Risk factors and leadership in a digitalized working world and their effects on employees' stress and resources: Web-based questionnaire study. *Journal of Medical Internet Research*, *23*(3), Article e24906. https://doi.org/10.2196/24906

Brockmann, E. N., & Anthony, W. P. (2016). Tacit knowledge and strategic decision making. *Group & Organization Management*, *27*(4), 436–455. https://doi.org/10.1177/1059601102238

Brown, A., & Green, T. (2016). Virtual reality: Low-cost tools and resources for the classroom. *TechTrends*, *60*, 517–519. https://doi.org/10.1007/s11528-016-0102-z

Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. (n.d.). *Sicher und gesund arbeiten in der Bauwirtschaft. Klassische Gefährdungen und etablierte Verfahren des Arbeitsschutzes*.

https://www.baua.de/DE/Themen/Arbeitsgestaltung-im-Betrieb/Branchen/Bauwirtschaft/_functions/BereichsPublikationssuche_Formular.html?queryResultId=null&pageNo=0

Bundesinstitut für Berufsausbildung. (n.d.). *Ausbildungsrahmenplan.* https://www.bibb.de/de/141443.php

Bundesministerium für Bildung und Forschung (n.d.). *The German vocational training system.* https://www.bmbf.de/bmbf/en/education/the-german-vocational-training-system/the-german-vocational-training-system_node.html

Bundesministerium für Bildung und Forschung (2019). *Lernort gestalten – Zukunft sichern. Digitalisierung in überbetrieblichen Berufsbildungsstätten* [Brochure]. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/31497_Lernort_gestalten_Zukunft_sichern.pdf?__blob=publicationFile&v=4

Bundesministerium für Bildung und Forschung (2020a). *Überbetriebliche Berufsbildungsstätten. Starke Partner der Wirtschaft* [Brochure]. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/31215_Ueberbetriebliche_Berufsbildungsstaetten.pdf?__blob=publicationFile&v=3

Bundesministerium für Bildung und Forschung (2020b). *Überbetriebliche Ausbildung – modern, digital und attraktiv. Sonderprogramm zur Digitalisierung in überbetrieblichen Berufsbildungsstätten* [Brochure]. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/31629_Ueberbetriebliche_Ausbildung.pdf?__blob=publicationFile&v=3

Bundesministerium für Bildung und Forschung (2020c). *eQualification 2021. Lernen und Beruf digital verbinden – Gamification! Projektband des Förderbereiches „Digitale Medien in der beruflichen Bildung“* [Brochure]. https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/3/31653_equalification_2021.pdf?__blob=publicationFile&v=3

Bundesministerium für Bildung und Forschung. (2020d, March 23). *Bekanntmachung. Richtlinie zur Förderung von Forschung und Entwicklung auf dem Gebiet „Interaktive Systeme in virtuellen und realen Räumen – Innovative Technologien für die digitale Gesellschaft“, Bundesanzeiger vom 02.04.2020.*

https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2020/04/2912_bekanntmachung.html

Bundesministerium für Bildung und Forschung (2022a, January 25). *Berufliche Bildung.* https://www.bmbf.de/bmbf/de/bildung/berufliche-bildung/berufliche-bildung_node.html

Bundesministerium für Bildung und Forschung (2022b, March 7). *Wettbewerb InnoVET: Zukunft gestalten – Innovationen für eine exzellente berufliche Bildung.* https://www.bmbf.de/bmbf/de/bildung/berufliche-bildung/foerderinitiativen-und-programme/wettbewerb-innovet/wettbewerb-innovet-zukunft-gestalten.html#searchFacets

Bundesregierung (2019, September 25). *50 Jahre Berufsbildungsgesetz. Erfolgsmodell duale Ausbildung.* https://www.bundesregierung.de/breg-de/aktuelles/erfolgsmodell-duale-ausbildung-1674226

Bundeszentrale für politische Bildung (2016). *Baugewerbe. Bauindustrie*. https://www.bpb.de/kurz-knapp/lexika/lexikon-der-wirtschaft/18786/baugewerbe/

Bustamante, E. A., Fallon, C. K., Bliss, J., Bailey, W. R., & Anderson, L. (2005). Pilots' workload, situation awareness, and trust during weather events as a function of time pressure, role assignment, pilots' rank, weather display, and weather system. *International Journal of Applied Aviation Studies*, *5*(2), 346–367. https://www.researchgate.net/publication/279770166_Pilots'_workload_situation_awareness_and_trust_during_weather_events_as_a_function_of_time_pressure_role_assignment_pilots'_rank_weather_display_and_weather_system

Campbell, F. (2006). Occupational stress in the construction industry. Survey 2006. *The Chartered Institute of Building* [Brochure]. https://www.ciob.org/sites/default/files/CIOB%20research%20-%20Occupational%20Stress%20in%20the%20Construction%20Industry%202006_0.pdf

Choi, M., Ahn, S., & Seo, J. (2020). VR-based investigation of forklift operator situation awareness for preventing collision accidents. *Accident Analysis and Prevention*, *136*, Article 105404. https://doi.org/10.1016/j.aap.2019.105404

Clifford, R. M. S., Khan, H., Hoermann, S., Billinghurst, M., & Lindeman, R. W. (2018). The effect of immersive displays on situation awareness in virtual environments for aerial firefighting air attack supervisor training. *IEEE Conference on Virtual Reality and 3D User Interfaces*, 533–534. https://doi.org/10.1109/VR.2018.8446139

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Cohen, S., Kessler, R. C., & Gordon, L. U. (1997). *Measuring stress: A guide for health and social scientists*. Oxford University Press.

Cornell, G., Boland, W., & Frank, G. A. (2007). Automating human performance measurement using XSLT in simulation-based exercises. *Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC)*, Article 7251. https://www.rti.org/sites/default/files/resources/7251_xslt.pdf

Crosswell, A. D., & Lockwood, K. G. (2020). Best practices for stress measurement: How to measure psychological stress in health research. *Health Psychology Open*, *7*(2), 1–12. https://doi.org/10.1177/2055102920933072

Cusveller, J., Gerritsen, C., & De Man, J. (2014). Evoking and measuring arousal in game settings. In S. Göbel & J. Wiemeyer (Eds.), *Lecture notes in computer science: Vol. 8395. Games for training, education, health and sports* (pp. 165–174). Springer. https://doi.org/10.1007/978-3-319-05972-3_17

Delfino, A. P. (2019). Student engagement and academic performance of students of Partido State University. *Asian Journal of University Education*, *15*(3), 42–55. https://doi.org/10.24191/ajue.v15i3.05

DEKRA. (n.d.). *Baumaschinen bedienen mit dem Baumaschinen-Bedienausweis.* https://www.dekra-akademie.de/weiterbildung/baumaschinenfuehrerschein

Deutscher Asphaltverband e.V. (2016). *Asphalt Leitfaden. Ratschläge für den Einbau von Walzasphalt* (2nd ed.) [Brochure]. https://www.asphalt.de/fileadmin/user_upload/downloads/dav/ratschlaege-2016.pdf

Deutsche Handwerks Zeitung (2019, January 16). *Kosten-Nutzen-Rechnung. Was kostet ein Azubi?* https://www.deutsche-handwerks-zeitung.de/was-kostet-ein-azubi-145741/

Dewe, P. J., O'Driscoll, M. P., Cooper, C. L. (2012). Theories of psychological stress at work. In R. Gatchel & I. Schultz (Eds.), *Handbook of Occupational Health and Wellness. Handbooks in Health, Work, and Disability.* Springer. https://doi.org/10.1007/978-1-4614-4839-6_2

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends*, *66*, 810–824. https://doi.org/10.1007/s11528-022-00717-w

Di Natale, A. F., Repetto, Riva, G., & Villani, D. (2020). Immersive virtual reality in K-12 and higher education: A 10-year systematic review of empirical research. *British Journal of Educational Technology*, *51*(6), 2006–2033. https://doi.org/10.1111/bjet.13030

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. Pr*oceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), Article 103. https://doi.org/10.1145/3264913

Drill, T. (2013, November 20). *Risiko in der Bauwirtschaft* [PowerPoint slides]. https://www.fh-muenster.de/bau/downloads/personen/dellen/Risikomanagement_Thomas_Drill.pdf

Edwards, J. R., & Cooper, C. L. (1990). The person-environment fit approach to stress: Recurring problems and some suggested solutions. *Journal of Organizational Behavior*, *11*(4), 293–307. http://www.jstor.org/stable/2488277

Empatica. (2020a). Empatica E4 wristband. User's manual. *Empatica S.R.L.* https://empatica.app.box.com/v/E4-User-Manual

Empatica. (2021, July 6). What should I know to use EDA data in my experiment? https://support.empatica.com/hc/en-us/articles/203621955-What-should-I-know-to-use-EDA-data-in-my-experiment

Empatica. (2022, November 2). What is electrodermal activity? https://support.empatica.com/hc/en-us/articles/115000198283-What-is-electrodermal-activity-

Endsley, M. R. (1988). Situation Awareness Global Assessment Technique (SAGAT). *Proceedings of the National Aerospace and Electronics Conference* (pp. 789–795). New York: IEEE. https://doi.org/10.1109/NAECON.1988.195097

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R. (1999). Situation awareness and human error: Designing to support human performance. *Proceedings of the High Consequence systems Surety Conference, Albuquerque, NM.* http://www.kbs.twi.tudelft.nl/Research/Projects/ICE/Local/_related%20documents/situation%20awareness/Endsley(1999)%20-%20SA%20and%20human%20error%3B%20designing%20to%20support%20human%20performance.pdf

Epel, E. S., Crosswell, A. D., Mayer, S. E., Prather, A. A., Slavich, G. M., Puterman, E., & Mendes, W. B. (2018). More than a feeling: A unified view of stress measurement for population science. *Frontiers in Neuroendocrinology*, *49*, 146–169. https://doi.org/10.1016/j.yfrne.2018.03.001

Eppler, M. J., & Mengis, J. (2002). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. *The Information Society*, *20*(5), 325–344. https://doi.org/10.1080/01972240490507974

Eurostat. (2019). *One in three EU workers experience TP.* https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20201006-2

Feng, C., Liu, S., Wanyan, X., Chen, H., Min, Y., Ma, Y. (2022). EEG feature analysis related to situation awareness assessment and discrimination. *Aerospace*, *9*(10), Article 546. https://doi.org/10.3390/aerospace9100546

Fliege, H., Rose, M., Arck, P., Levenstein, S., & Klapp, B. F. (2009). PSQ. Perceived Stress Questionnaire [Verfahrensdokumentation, PSQ20-Skalenberechnung, PSQ20-Fragebogen Englisch, Deutsch, Deutsch (letzte 2 Jahre), PSQ30-Skalenberechnung, PSQ30-Fragebogen Englisch, Französisch, Deutsch, Italienisch und Spanisch]. In Leibniz-Institut für Psychologie (ZPID) (Ed.), *Open Test Archive. ZPID.* http://dx.doi.org/10.23668/psycharchives.5138

Föhr, T., Tolvanen, A., Myllymäki, T., Järvelä-Reijonen, E., Rantala, S., Korpela, R., Peuhkuri, K., Kolehmainen, M., Puttonen, S., Lappalainen, R., Rusko, H., & Kujala, U. M. (2015). Subjective stress, objective heart rate variability-based stress, and recovery on workdays among overweight and psychologically distressed individuals: A cross-sectional study. *Journal of Occupational Medicine and Toxicology*, *10*, Article 39. https://doi.org/10.1186/s12995-015-0081-6

Gallego-Durán, F. J., Molina-Carmona, R., & Llorens-Largo, F. (2016). An approach to measuring the difficulty of learning activities. In P. Zaphiris & A. Ioannou (Eds*.), Lecture notes in computer science: Vol. 9753. Learning and collaboration technologies* (pp. 417–428). Springer. https://doi.org/10.1007/978-3-319-39483-1_38

George, D., & Mallery, P. (2003). *SPSS for windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn and Bacon.

Giannakakis, G., Grigoriadis, D., Giannakaki, K., Simantiraki, O., Roniotis, A., & Tsiknakis, M. (2022). Review on psychological stress detection using biosignals. *IEEE Transactions on Affective Computing*, *13*(1), 440–460. https://doi.org/10.1109/TAFFC.2019.2927337

Grantcharov, T. P., Kristiansen, V. B., Bendix, J., Bardram, L., Rosenberg, J., & Funch-Jensen, P. (2004). Randomized clinical trial of virtual reality simulation laparoscopic skills training. *British Journal of Surgery*, *91*(2), 146–150. https://doi.org/10.1002/bjs.4407

Greco, A., Valenza, G, Lazaro, J., Garzon-Rey, J. M., Aguilo, J., De-la-Camera, C., Bailon, R., & Scilingo, E. P. (2012). Acute stress state classification based on electrodermal activity modelling. *IEEE Transactions on Affective Computing.* https://doi.org/10.1109/TAFFC.2021.3055294

Green, B., Parry, D., Oeppen, R. S., Plint, S., Dale, T., & Brennan, P. A. (2017). Situational awareness – what it means for clinicians, its recognition and importance in patient safety. *Oral Diseases*, *23*(6), 721–725. https://doi.org/10.1111/odi.12547

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, *54*, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

Hasanzadeh, S., Esmaeili, B., & Dodd, M. D. (2018). Examining the relationship between construction workers' visual attention and situation awareness under fall and tripping hazard conditions: Using mobile eye tracking. *Journal of Construction Engineering and Management*, *144*(7), Article 04018060. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001516

Health and Safety Executive (n.d.). *Traffic management on site.* https://www.hse.gov.uk/construction/safetytopics/vehiclestrafficmanagement.htm

Hiel Van, A., & Mervielde, I. (2016). The search for complex problem-solving strategies in the presence of stressors. *The Journal of the Human Factors and Ergonomics Society*, *49*(6), 1072–1082. https://doi.org/10.1518/001872007X249938

Horan, K. A., Nakahara, W. H., DiStaso, M. J., Jex, S. M. (2020). A review of the challenge-hindrance stress model: Recent advances, expanded paradigms, and recommendations for future research. *Frontiers in Psychology*, *11*, Article 560346. https://doi.org/10.3389/fpsyg.2020.560346

Huang, W., Roscoe, R. D., Johnson-Glenberg, M. C., & Craig, S. D. (2020). Motivation, engagement, and performance across multiple virtual reality sessions and levels of immersion (2020). *Journal of Computer Assisted Learning*, *37*(3), 745–758. https://doi.org/10.1111/jcal.12520

IFO Institut (2021, October 6). *Fachkräftemangel auf deutschen Baustellen verschärft sich.* https://www.ifo.de/pressemitteilung/2021-10-06/fachkraeftemangel-auf-deutschen-baustellen-verschaerft-sich

Jensen, L., & Konradsen, F. (2018). A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies*, *23*, 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

Jokela, M. (2022). Why is cognitive ability associated with psychological distress and wellbeing? Exploring psychological, biological, and social mechanisms. *Personality and Individual Differences*, *192*, Article 111592. https://doi.org/10.1016/j.paid.2022.111592

Jonsson, K., Brulin, C., Härgestam, M., Lindkvist, M., & Hultin, M. (2021). Do team and task performance improve after training situation awareness? A randomized controlled study of interprofessional intensive care teams. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, *29*, Article 73. https://doi.org/10.1186/s13049-021-00878-2

Kim, H. G., Cheon, E. J., Bai, D. S., Lee, Y. H., & Koo, B. H. (2018). Stress and heart rate variability: A meta-analysis and review of the literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doig.org/10.30773/pi.2017.08.17

Kim, S., Lee, H., Hwang, S., Yi, J.-S., & Son, J. W. (2022). Construction workers' awareness of safety information depending on physical and mental load. *Journal of Asian Architecture and Building Engineering*, *21*(3), 1067–1077. https://doi.org/10.1080/13467581.2021.1908899

Knaust, T., Felnhofer, A., Kothgassner, O. D., Höllmer, H., Gorzka, R.-J., & Schulz, H. (2022). Exposure to virtual nature: The impact of different immersion levels on skin conductance level, heart rate, and perceived relaxation. *Virtual Reality*, *26*, 925–938. https://doi.org/10.1007/s10055-021-00595-2

Kolfschoten, G. L., & Brazier, F. M. T. (2013). Cognitive load in collaboration: Convergence. *Group Decision and Negotiation*, *22*, 975–996. https://doi.org/10.1007/s10726-012-9322-6

Kompier, M. (2005). Assessing the psychosocial work environment – "subjective" versus "objective" measurement. *Scandinavian Journal of Work, Environment & Health*, *31*(6), 405–408. https://doi.org/10.5271/sjweh.946

Koutitas, G., Smith, S., & Lawrence, G. (2021). Performance evaluation of AR/VR training technologies for EMS first responders. *Virtual Reality*, *25*, 83–94. https://doi.org/10.1007/s10055-020-00436-8

Kultusministerkonferenz (2021). *Handreichung für die Erarbeitung von Rahmenlehrpläne der Kultusministerkonferenz für den berufsbezogenen Unterricht in der Berufsschule und ihre Abstimmung mir Ausbildungsordnungen des Bundes für anerkannte Ausbildungsberufe* [Brochure].

https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2021/2021_06_17-GEP-Handreichung.pdf

Lang, Y., Wei, L., Xu, F., Zhao, Y., and Yu, L. (2018). Synthesizing personalized training programs for improving driving habits via virtual reality. *IEEE Conference on Virtual Reality and 3D User Interfaces*, 297–304. https://doi.org/10.1109/VR.2018.8448290

Langevin, P., & Mendoza, C. (2021). Procedural fairness of objective and subjective performance evaluations: Exploring the combined effects of uncertainty and trust. *Comptabilité Contrôle Audit*, *27*, 111–154. https://doi.org/10.3917/cca.272.0111

LeBlanc, V. R. (2009). The effects of acute stress on performance: implications for health professions education. *Academic Medicine*, *84*(10), 25–33. https://doi.org/10.1097/ACM.0b013e3181b37b8f

Le Fevre, M., Matheny, J., & Kolt, G. S. (2003). Eustress, distress, and interpretation in occupational stress. *Journal of Managerial Psychology*, *18*(7), 726–744. https://doi.org/10.1108/02683940310502412

Lenggo Putri, D., Sumihardi, S., Irfan, A., & Djaja, I. M. (2019). Relationship between unsafe action and condition with work accident among production unit workers at the Jaya Sentrikon Indonesia Company, Pandang, West Sumatra. *The 6th International Conference on Public Health*, Indonesia, 22–30. https://doi.org/10.26911/the6thicph-FP.01.04

Lepine, M. A. (2022). The challenge-hindrance stressor framework: An integrative conceptual review and path forward. *Group & Organization Management*, *47*(2), 223–254. https://doi.org/10.1177/10596011221079970

Lepine, J. A., Podsakoff, N. P., & Lepine, M. A. (2005). A meta-analytic test of the challenge stressor-hindrance stressor framework. An explanation for inconsistent relationships among stressors and performance. *The Academy of Management Journal*, *48*(5), 764–775. https://doi.org/10.5465/amj.2005.18803921

Leung, M.-Y., Lian, Q., & Olomolaiye, P. (2016). Impact of job stressors and stress on the safety behaviour and accidents of construction workers. *Journal of Management in Engineering*, *32*(1). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000373

Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., & Andreoli, A. (1993). Development of the Perceived Stress Questionnaire: A new tool for psychosomatic research. *Journal for Psychosomatic Research*, *37*(1), 19–32. https://doi.org/10.1016/0022-3999(93)90120-5

Li, H., Chan, G., & Skitmore, M. (2012). Multiuser virtual safety training system for tower crane dismantlement. *Journal of Computing in Civil Engineering*, *26*(5), 638–647. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000170

Liu, Y., & Du, S. (2018). Psychological stress-level detection based on electrodermal activity. *Behavioural Brain Research*, *341*, 50–53. https://doi.org/10.1016/j.bbr.2017.12.021

Marquardt, N. (2019). Situation awareness, human error, and organizational learning in sociotechnical systems. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *29*(4), 327–339. https://doi.org/10.1002/hfm.20790

Marques, A. H., Silverman, M. N., & Sternberg, E. M. (2021). Evaluation of stress systems by applying noninvasive methodologies: Measurements of neuroimmune biomarkers in the sweat, heart rate variability and salivary cortisol. *NeuroImmunoModulation*, *17*(3), 205–208. https://doi.org/10.1159%2F000258725

McGuire, K., & Lorenz, R. (2018). Effect of simulation on learner stress as measured by cortisol. An integrative review. *Nurse Educator*, *43*(1), 45–49. https://doi.org/10.1097/NNE.0000000000000393

Martín-Gutiérrez, J., Efrén Mora, C., Añorbe-Díaz, B., & González-Marrero, A. (2017). Virtual technologies trends in education. *EUROSIA Journal of Mathematics Science and Technology Education*, *13*(2), 469–486. https://doi.org/10.12973/eurasia.2017.00626a

McCorry, L. K. (2007). Physiology of the autonomic nervous system. *American Journal of Pharmaceutical Education*, *71*(4), Article 78. https://doi.org/10.5688/aj710478

McNeal, K. S., Zhong, M., Soltis, N. A., Doukopoulos, L., Johnson, E. T., Courtney, S., Alwan, A., & Porch, M. (2020). Biosensors show promise as a measure of student engagement in a large introductory biology course. *CBE Life Sciences Education*, *19*(4), Article 50. https://doi.org/10.1187%2Fcbe.19-08-0158

Mefford, J. M., Kahle, S., Gupta, S., Tancredi, D., Danielson, A. R., & Clarke, S. O. (2019). Heart rate variability and acute stress among novice airway managers. *AEM Education and Training*, *3*(3), 291–294. https://doi.org/10.1002/aet2.10335

Molen Van der, H. F., & Hoonakker, P. L. T. (2000). Work stress in the construction industry: Causes and measures. *Proceedings of the IEA 2000/HFES 2000 Congress*, *44*(31), Article 5-651–5-654. https://doi.org/10.1177/154193120004403

Mohammadfam, I., Mahdinia, M., Soltanzadeh, A., Aliabadi, M. M., & Soltanian, A. R. (2021). A path analysis model of individual variables predicting safety behavior and human error: The mediating effect of situation awareness. *International Journal of Industrial Ergonomics*, *84*, Article 103144. https://doi.org/10.1016/j.ergon.2021.103144

Mohammadi, A., Tavakolan, M., & Khosravi, Y. (2018). Factors influencing safety performance on construction projects: A review. *Safety Science*, *109*, 382–397. https://doi.org/10.1016/j.ssci.2018.06.017

Moridi, G., Khaledi, S., & Valiee, S. (2014). Clinical training stress-inducing factors from the students' viewpoint: A questionnaire-based study. *Nurse Education in Practice*, *14*(2), 160–163. https://doi.org/10.1016/j.nepr.2013.08.001

Nepal, M. P., Park, M., & Son, B. (2006). Effects of schedule pressure on construction performance. *Journal of Construction Engineering and Management*, *132*(2), 182–188. https://doi.org/10.1061/(ASCE)0733-9364(2006)132:2(182)

Noushad, S., Ahmed, S., Ansari, B., Mustafa, U. H., Saleem, Y., & Hazrat, H. (2021). Physiological biomarkers of chronic stress: A systematic review. *International Journal of Health Sciences*, *15*(5), 46–59. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8434839/

Orizon. (2019, July 24). *Wenn die To-Do-Liste zur Last wird: Orizon Arbeitsmarktbefragung legt größte Stressfaktoren im Job offen* [Press release]. https://www.orizon.de/de/presselounge/wenn-die-to-do-liste-zur-last-wird

Papoušek, J., Stanislav, V., & Pelánek, R. (2016). Impact of question difficulty on engagement and learning. In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *Lecture notes in computer science: Vol. 9684. Intelligent Tutoring Systems* (pp. 267–272). Springer. https://doi.org/10.1007/978-3-319-39583-8_28

Pasca, R. (2014). Person-environment fit theory. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 4776–4778). Springer. https://doi.org/10.1007/978-94-007-0753-5_2155

Pegler, W. (2021). When the going gets tough. *Build*, *183*, 90–91. https://www.buildmagazine.org.nz/assets/PDF/Build-183-90-Health-And-Safety-When-The-Going-Gets-Tough.pdf

Pluut, H., Curșeu, P. L., & Fodor, O. C. (2022). Development and validation of a short measure of emotional, physical, and behavioral markers of eustress and distress (MEDS). *Healthcare*, *10*(2), Article 339. https://doi.org/10.3390/healthcare10020339

Psotka, J. (1995). Immersive training systems: Virtual reality and education and training. *Instructional Science*, *23*(5), 405–431. https://www.jstor.org/stable/23370939

Razavi, T. B. (2001). *Self-report measures: An overview of concerns and limitations of questionnaire use in occupation stress research* [discussion paper]. University of Southampton. https://eprints.soton.ac.uk/35712/

Ringen, K., Seegal, J., & Englund, A. (1995). Safety and health in the construction industry. *Annual Review of Public Health*, *16*, 165–188. https://doi.org/10.1146/annurev.pu.16.050195.001121

Rubio-Tamayo, J. L., Gertrudix Barrio, M., & García García, F. (2017). Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation. *Multimodal Technologies and Interaction*, *1*(21), 1–20. https://doi.org/10.3390/mti1040021

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, *31*(9), 1005–1017. https://doi.org/10.1080/01446193.2013.828844

Salkind, N. J. (2010). *Encyclopedia of Research Design* (Vol. 2). SAGE.

Salmon, P. M., Walker, G. H., & Stanton, N. A. (2016). Pilot error versus sociotechnical systems failure: A distributed situation awareness analysis of Air France 447. *Theoretical Issues in Ergonomics Science*, *17*(1), 64–79. https://doi.org/10.1080/1463922X.2015.1106618

Sänger, J., Bechtold, L., Schoofs, D., Blaszkewicz, M., & Wascher, E. (2014). The influence of acute stress on attention mechanisms and its electrophysiological correlates. *Frontiers in Behavioral Neuroscience*, *8*(353), 1–13. https://doi.org/10.3389/fnbeh.2014.00353

Sapateiro, C., Antunes, P., Zurita, G., Baloian, N., & Vogt, R. (2009). Supporting unstructured activities in crisis management: A collaboration model and prototype to improve situation awareness. In J. Löffler & M. Klann (Eds.), *Lecture notes in computer science: Vol. 5424. Mobile response* (pp. 101–111). Springer. https://doi.org/10.1007/978-3-642-00440-7_10

Sattar, M. U., Palaniappan, S., Lokman, A., Shah, N., Khalid, U., & Hasan, R. (2020). Motivating medical students using virtual reality based education. *International Journal of Emerging Technologies in Learning*, *15*(2), 160–174. https://doi.org/10.3991/ijet.v15i02.11394

Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, *98*, 146–151. https://doi.org/10.1016/j.jclinepi.2017.12.006

Schmitt, A., Ohly, S, & Kleespies, N. (2015). Time pressure promotes work engagement. Test of illegitimate tasks as boundary condition. *Journal of Personnel Psychology*, *14*(1), 28–36. https://doi.org/10.1027/1866-5888/a000119

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., & Ehlert, U. (2010). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine*, *14*(2), 410–417. https://doi.org/10.1109/TITB.2009.2036164

Shehata, M. E., & El-Gohary, K. M. (2011). Towards improving construction labor productivity and projects' performance. *Alexandria Engineering Journal*, *50*, 321–330. https://doi.org/10.1016/j.aej.2012.02.001

Sneddon, A., Mearns, K., & Flin, R. (2013). Stress, fatigue, situation awareness and safety in offshore drilling crews. *Safety Science*, *56*, 80–88. https://doi.org/10.1016/j.ssci.2012.05.027

Sommerfeldt, S. L., Schaefer, S. M., Brauer, M., Ryff, C. D., & Davidson, R. J. (2019). Individual differences in the association between subjective stress and heart rate are related to psychological and physical well-being. *Psychological Science*, *30*(7), 1016–1029. https://doi.org/10.1177/0956797619849555

Stangl, W. (n.d.). Stressor. *Online Lexikon für Psychologie & Pädagogik. Online-Enzyklopädie aus den Wissenschaften Psychologie und Pädagogik.* https://lexikon.stangl.eu/4142/stressor

Statistisches Bundesamt. (2022a). *Produzierendes Gewerbe: Tätige Personen und Umsatz der Betriebe im Baugewerbe 2021* [Data file]. Statistisches Bundesamt (Destatis). https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Bauen/Publikationen/Downloads-Baugewerbe-Struktur/personen-umsatz-baugewerbe-2040510217004.pdf?__blob=publicationFile

Statistisches Bundesamt. (2022b). *Betriebe, Tätige Personen, Entgelte, Geleistete Arbeitsstunden, Umsatz im Bauhauptgewerbe (Betriebe mit 20 u.m. tätigen Personen): Deutschland, Monate, Wirtschaftszweige* [Data file]. Statistisches Bundesamt (Destatis). https://www-genesis.destatis.de/genesis/downloads/00/tables/44111-0004_00.csv

Stavroula, L., Griffiths, A., & Cox, T. (2003). Work organization & stress. *Protecting Workers' Health Series No. 3.* https://apps.who.int/iris/bitstream/handle/10665/42625/9241590475.pdf

Stone, E. R. (2010). T-test, paired samples. In N. J. Salkind (Ed.), *Encyclopedia of research design* (pp. 1560–1562. SAGE.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99–103. https://doi.org/10.1207/S15327752JPA8001_18

Szalma, J. L., Hancock, P. A., & Quinn, S. (2008). A meta-analysis of the effect of time pressure on human performance. *Proceedings of the HUMAN FACTORS AND ERCONOMICS SOCIETY 52nd Annual meeting 2008*, *52*, 1513–1516. https://doi.org/10.1518/107118108X353174

Tijani, B., Jin, X., & Osei-Kyei, R. (2021). A systematic review of mental stressors in the construction industry. *International Journal of Building Pathology and Adaptation*, *39*(2), 433–460. https://doi.org/10.1108/IJBPA-02-2020-0011

Turpin, G. & Grandfield, T. (2007). Electrodermal Activity. In G. Fink (Ed.), *Encyclopedia of Stress* (2nd ed., pp. 899–902). Academic Press. https://doi.org/10.1016/B978-012373947-6.00139-2

Van Veldhoven, M. (2014). Quantitative job demands. In M. C. W. Peeters, J. De Jonge, & T. W. Taris (Eds.), *An introduction to contemporary work psychology* (pp. 117–143). Wiley Blackwell.

Wang, P., Wu, P., Wang, J., Hung-Lin, C., & Wang, X. (2018). A critical review of the use of virtual reality in construction engineering education and training. *International Journal of Environmental Research and Public Health*, *15*(6), Article 1204. https://doi.org/10.3390/ijerph15061204

Webb, C., Gao, L., & Song, L. (2015). Schedule compression impact on construction project safety. *Frontiers of Engineering Management*, *2*(4), 344–350. https://doi.org/10.15302/J-FEM-2015059

Weckesser, L. J., Dietz, F., Schmidt, K., Grass, J., Kirschbaum, C., & Miller, R. (2019). The psychometric properties and temporal dynamics of subjective stress, retrospectively assessed by different informants and questionnaires, and hair cortisol concentrations. *Scientific Reports*, *9*(1), Article 1098. https://doi.org/10.1038/s41598-018-37526-2

Weigel, A., Sauter, L., & Niehaves, B. (2021). More than you know – An investigation of VR to support tacit knowledge transfer within organisation. *Proceedings of the Twenty-fifth Pacific Asia Conference on Information Systems*, Article 166. https://aisel.aisnet.org/pacis2021/166

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance* (5th ed.). Routledge. https://doi.org/10.4324/9781003177616

Wickramasuriya, D. S., Qi, C., & Faghih, R. T. (2018). A state-space approach for detecting stress from electrodermal activity. *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 3562–3567. https://doi.org/10.1109/EMBC.2018.8512928

Wilson VanVoorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50. https://doi.org/10.20982/tqmp.03.2.p043

Wolfartsberger, J., Riedl, R., Jodlbauer, H., Haslinger, N., Kirisits, A., & Schuh, S. (2022). Virtual Reality als Trainingsmethode: Eine Laborstudie aus dem Industriebereich. *HMD Praxis der Wirtschaftsinformatik*, *59*, 295–308. https://doi.org/10.1365/s40702-021-00819-8

Xie, B., Liu, H., Alghofaili, R., Zhang, Y., Jiang, Y., Lobo, F. D., Li, C., Li, W., Huang, H., Akdere, M., Mousas, C., & Yu, L.-F. (2021). A Review on Virtual Reality Skill Training Applications. *Frontiers in Virtual Reality*, *2*, Article 645153. https://doi.org/10.3389/frvir.2021.645153

Yang, Y. & Li, X. (2021). The impact of challenge and hindrance stressors on thriving at work double mediation based on affect and motivation. *Frontiers in Psychology*, *12*, Article 613871. https://doi.org/10.3389/fpsyg.2021.613871

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482.

Zentralverband des Deutschen Handwerks. (2022). *Lehrlingsbestand nach Ausbildungsberufen 2021* [Data file]. https://www.zdh-statistik.de/application/stat_det.php?LID=1&ID=MDUwNjg=&cID=00857

Zhang, T., Yang, J., Liang, N., Pitts, B. J., Prakah-Asante, K. O., Curry, R., Duerstock, B. S., Wachs, J. P., & Yu, D. (2020). Physiological measurements of situation awareness: A systematic review. *Human Factors*, *0*(0), 1–22. https://doi.org/10.1177/0018720820969071

Zhang, S., Sunindijo, R. Y., Frimpong, S., & Su, Z. (2023). Work stressors, coping strategies, and poor mental health in the Chinese construction industry. *Safety Science*, *159*, Article 106039. https://doi.org/10.1016/j.ssci.2022.106039

7. Appendix

Appendix A: Description of the Practical Context

Strengths and Limitations of the German VET System Exemplified by the Construction Sector

Construction is a diverse and complex industry (e.g., Albert et al., 2014) divided into building construction (e.g., residential, public and industrial buildings) and civil engineering (e.g., roads and bridges) (Bundeszentrale für politische Bildung, 2016). Like many other sectors in Germany, the construction industry suffers from a shortage of qualified workers. A survey by the IFO Institute has shown that in building construction 33.5% of companies had problems finding skilled workers in September 2021. In civil engineering the figure was as high as 37.9% (IFO Institut, 2021). Road construction accounts for the largest share within civil engineering regarding turnover (€1.563 million in June 2021) (Statistisches Bundesamt, 2022a) and employees (78.9 thousand in 2021) (Statistisches Bundesamt, 2022b). Accordingly, trainees in road construction make the highest share within civil engineering (2.912 in 2021) (Zentralverband des Deutschen Handwerks, 2022) and are, regarding the labour shortage, of high demand after they have finished their education.

A high-quality education forms the basis for a smooth and secure start into working life which is in the interest of all parties involved (apprentices, employers, customers, and the government). Vocational training in Germany usually lasts between two to three and a half years and follows a dual approach. This means that trainees acquire knowledge, know-how and skills both at vocational school and at a company (BMBF, n.d.). Two distinct but coordinated curricula, for vocational school and in-company training, form the basis for the dual system (Bundesinstitut für Berufsausbildung [BIBB], n.d.). Moreover, vocational training for some sectors is enriched by inter-company training centres, which complement practical training by focusing on activities that are not necessarily carried out in every company for reasons of size and specialty (BMBF, 2020a). Recently qualified workers therefore start with some practical experience into their job which is a quality attribute that is not only valued by companies, but also by other European countries that increasingly use the German VET system as a blueprint (Bundesregierung, 2019).

Generally, the curricula provide for a steady increase in the level of demands and complexity (Kultusministerkonferenz, 2021). Construction trainees, in particular, would benefit greatly from diverse on-the-job experiences and a supervised exposure to job-related stressors during their in-company training. Such experiences would promote their situation awareness, a tacit knowledge that is considered helpful in dynamic work environments (Endsley, 1995). Likewise, trainees would gain valuable know-how how to deal with job-related stressors safely and efficiently. However, for demanding and dangerous working environments inherent in the construction industry, any increase in the level of demands and complexity requires special caution and supervision since errors could result in severe consequences, such as injury (Lenggo Putri et al., 2019) and damage (e.g., BAUA, n.d.; Drill, 2013). These demanding working conditions and related risks should therefore always be considered in the transition from developing to challenging trainees. In other words: An increase in complexity and demands should not be rushed but done with caution. Inter-company training centres support in a steady and safe increase of demands. However, their training opportunities are also limited when it comes to training with complex machines and training procedures under stressful conditions due to safety, capacity and cost reasons (e.g., Wang et al., 2018).

Appendix B: Briefings

General Briefing

Beginnen Sie in beiden Fällen auf der rechten Seite der Straße und verdichten Sie nur die rechte Hälfte der Straße. Zwei Fahrspuren reichen aus, um die Mitte der Straße zu erreichen. Ziel ist es, die Straße in der vorgegebenen Zeit viermal zu verdichten.

English: In both scenarios, start on the right side of the road and compact only the right half of the road. Two lanes are enough to reach the middle of the road. In the given time, aim to compact the road four times.

Briefing for the Low Time Pressure Condition

Heute ist es warm, etwa 30 Grad, und es bleibt den ganzen Tag über trocken. Wegen der warmen Temperaturen kühlt der Asphalt relativ langsam ab, was bedeutet, dass Sie keinen Zeitdruck beim Ausführen Ihrer Aufgabe haben. Im Gegenteil, Sie können und sollten ausreichend Abstand zum Fertiger einhalten.

English: Today it is warm, around 30 degrees, and it stays dry all day. Because of the warm temperatures, the asphalt cools down relatively slowly, which means that you have no time pressure when carrying out your task. Precisely, you should maintain sufficient distance from the paving machine.

Briefing for the High Time Pressure Condition

Die Wetterbedingungen sind heute nicht optimal, da sie Zeitdruck bei Ihrer Arbeit erzeugen: Heute sind es gerade einmal 7 Grad und es ist stark bewölkt. Neben den sehr niedrigen Temperaturen, die bereits im Grenzbereich für Asphalteinbau liegen, besteht daher die Möglichkeit von Regen und damit eine zusätzliche Gefahr, dass die Arbeit komplett eingestellt werden muss. Es ist also wichtig, dass Sie während Ihrer Arbeit die Temperatur und die Wolkendecke aufmerksam beobachten. Die geringe Temperatur sorgt außerdem dafür, dass der Asphalt relativ schnell abkühlt und Sie daher in der Nähe der Fertiger bleiben müssen.

English: Today the weather conditions are not ideal as they create time pressure on your work: Today it is just 7 degrees and very cloudy. Therefore, in addition to the very low temperatures, which are already in the marginal range for asphalt compaction, there is a possibility of rain and thus an additional risk that the work will have to be stopped completely. It is therefore important to keep a close eye on the temperature and clouds during your work. The low temperature also causes the asphalt to cool down relatively quickly, so you have to stay close to the paving machine.

Appendix C: Evaluation of the Situation Awareness Questions

SA Item	**Low** Time Pressure Condition	Evaluation	Notes/Coding
1	Wie warm ist es gerade?	automatic EVA	
2	Welche Art von Asphalt rollen Sie?	automatic EVA	
3	Auf welcher Stärke sind Ihre Sprühdüsen eingestellt?	screen recording [a]	Assumption: The data of the last glance at the dashboard was used. If a participant did not look at the dashboard at all, "off" was used as the default answer, as it is assumed that participants would at least take a quick look at the dashboard when operating the sprinklers.
4	Wie ist Ihre Geschwindigkeit jetzt gerade?	screen recording	In addition to the speed display on the dashboard, the speed was recorded separately. Thus, the evaluation of item 4 was not dependent on the screen recording.
5	Welchen Einfluss hat die derzeitige Außentemperatur auf den Abstand zum Asphaltfertiger?	automatic EVA	
6	Worauf müssen Sie bei diesen Temperaturen beim Asphalt achten?	automatic EVA	
7	Was passiert mit dem Asphalt aufgrund der Temperatur?	automatic EVA	
8	Worauf müssen Sie bei dieser Temperatur bei der Walze achten?	automatic EVA	
9	Wie lange dauert es, bis der Wassertank leer ist?	screen recording (automatic EVA) [a b]	
10	Wie viele Walzüberfahrten müssen Sie insgesamt noch durchführen, bis Sie die richtige Verdichtung erreicht haben?	automatic EVA	Due to the limited time of 6 minutes in combination with the length of the street, only one of the three answer options was actually possible.
11	Erwarten Sie Veränderungen der Verkehrssituation?	automatic EVA	
12	Erwarten Sie eine Verzögerung beim Walzen?	automatic EVA	

[a] The use of screen recordings has its limitations as only a fraction of the 3D environment can be captured in a 2D video, i.e., only the part that the participants were directly looking at is shown and can be used for evaluation.
[b] Automatic EVA could be used since the user scores for Q9 all were in the very high range and therefore matched the default answer.

SA Item	**High** Time Pressure Condition	Evaluation	Notes/Coding
1	Wie viele Walzüberfahrten haben Sie im angezeigten Bereich bereits durchgeführt?	heatmap	Measuring point: Cell A1
2	Welche Seite des Asphalts ist die niedrige Seite? Wählen Sie eines der zwei blauen Kästchen aus.	automatic EVA	
3	Sind die Sprühdüsen eingeschaltet?	screen recording [a]	Assumption: The data of the last glance at the dashboard was used. If a participant did not look at the dashboard at all, "no" was used as the default answer, as it is assumed that participants would at least take a quick look at the dashboard when operating the sprinklers.
4	Ist die Vibrationsfunktion eingeschaltet?	screen recording [a]	Assumption: The data of the last glance at the dashboard was used. If a participant did not look at the dashboard at all, "no" was used as the default answer, as it is assumed that participants would at least take a quick look at the dashboard when operating the vibration.
5	An welcher Walzphase arbeiten Sie gerade?	heatmap & screen recording	Identification of the measuring point: Cell on which the roller was when the time was over (identified with screen recording); analysis of compaction degree: heatmap
6	Wie beeinflusst das Wetter Ihre Arbeitsweise?	automatic EVA	
7	Ist in Ihrem Arbeitsbereich derzeit mit anderem Verkehr zu rechnen?	automatic EVA	
8	Können Sie jetzt die Vibration einschalten?	automatic EVA	
9	Wie lange dauert es, bis der Wassertank leer ist?	screen recording (automatic EVA) [a b]	
10	Wie lange dauert es, bis der Kraftstofftank leer ist?	screen recording (automatic EVA) [a b]	
11	Wie viele Walzüberfahrten müssen Sie im angezeigten Bereich noch durchführen, um die richtige Verdichtung zu erreichen?	heatmap	
12	Rechnen Sie mit Wetterumschwüngen?	automatic EVA	

[a] The use of screen recordings has its limitations as only a fraction of the 3D environment can be captured in a 2D video, i.e., only the part that the participants were directly looking at is shown and can be used for evaluation.
[b] Automatic EVA could be used since the user scores for Q9 and Q10 all were in the very high range and therefore matched the default answer.

Appendix D: Original and Adapted Demands Scale of the PSQ

Item	German Version		English Version	
	Original	Adaption	Original	Adaption
1	Sie haben das Gefühl, dass zu viele Forderungen an Sie gestellt werden.	none	You feel that too many demands are being made on you.	none
2	Sie haben zu viel zu tun.	none	You have to many things to do.	none
3	Sie fühlen sich gehetzt.	none	You feel you're in a hurry.	none
4	Sie haben genug Zeit *für sich*.	Sie haben genug Zeit.	You have enough time *for yourself*.	You have enough time.
5	Sie fühlen sich unter *Termindruck*.	Sie fühlen sich unter Zeitdruck.	You feel under pressure from *deadlines*.	You feel under time pressure.

Note. All Items were set into past tense. Items 4 and 5 were slightly adapted. Items 1, 2 and 3 were not adapted.

Appendix E: Checklist for Study Conduction

Tick the Box	Activity	Duration (min)	Notes	Empatica timestamp
	Welcome student, explanation and signing of informed consent	3		
	Attachment and start of Empatica	1		
	Technical explanation of the HMD, steering wheel and joystick	2		
	General briefing	3		
O 1 O 2	**Low TPC**			
	Condition-specific briefing	1	Change order of scenarios	
	Simulation execution	6		Set timestamp for beginning and end
	SA questions	2		
	Stress questionnaire	1		
O 1 O 2	**High TPC**			
	Condition-specific briefing	1	Change order of scenarios	
	Simulation execution	6		Set timestamp for beginning and end
	SA questions	2		
	Stress questionnaire	1		
	Questions on personal data, experience and trainees' opinion on the integration of VR in vocational training	1		
	Total duration	**30**		

Appendix F: Complete Questionnaire

Fragebogen

Teilnehmer-Nr.: _______

Szenario: O A O B

Im Folgenden finden Sie fünf Feststellungen. Bitte lesen Sie jede durch und wählen Sie aus den vier Antworten diejenige aus, die angibt, wie häufig die Feststellung auf Sie **während der vergangen VR-Simulation** zutraf. Es gibt keine richtigen oder falschen Antworten. Überlegen Sie bitte nicht lange und lassen Sie keine Frage aus.

Sie haben das Gefühl, dass zu viele Forderungen an Sie gestellt wurden.	O fast nie	O manchmal	O häufig	O meistens
Sie hatten zu viel zu tun.	O fast nie	O manchmal	O häufig	O meistens
Sie fühlten sich gehetzt.	O fast nie	O manchmal	O häufig	O meistens
Sie hatten genug Zeit.	O fast nie	O manchmal	O häufig	O meistens
Sie fühlten sich unter Zeitdruck.	O fast nie	O manchmal	O häufig	O meistens

---

Szenario: O A O B

Im Folgenden finden Sie fünf Feststellungen. Bitte lesen Sie jede durch und wählen Sie aus den vier Antworten diejenige aus, die angibt, wie häufig die Feststellung auf Sie **während der vergangen VR-Simulation** zutraf. Es gibt keine richtigen oder falschen Antworten. Überlegen Sie bitte nicht lange und lassen Sie keine Frage aus.

Sie haben das Gefühl, dass zu viele Forderungen an Sie gestellt wurden.	O fast nie	O manchmal	O häufig	O meistens
Sie hatten zu viel zu tun.	O fast nie	O manchmal	O häufig	O meistens
Sie fühlten sich gehetzt.	O fast nie	O manchmal	O häufig	O meistens
Sie hatten genug Zeit.	O fast nie	O manchmal	O häufig	O meistens
Sie fühlten sich unter Zeitdruck.	O fast nie	O manchmal	O häufig	O meistens

Persönliche Daten:

Ausbildungsjahr: _______

Alter: _______

Praktische Erfahrung:

Haben Sie bereits praktische Erfahrung in der Asphaltverdichtung?	O keine oder 1-mal	O manchmal	O häufig
Haben Sie bereits praktische Erfahrung mit Virtual Reality (z.B. durch Computerspiele)?	O keine oder 1-mal	O manchmal	O häufig

VR in der Berufsausbildung:

Wie würden Sie die Einführung von VR-Training in Ihre Berufsausbildung bewerten?	O nicht sinnvoll	O teilweise sinnvoll	O sinnvoll
Wie motiviert wären Sie VR-Training in Ihrer Berufsausbildung zu nutzen?	O nicht motiviert	O teilweise motiviert	O motiviert

Vielen Dank für Ihre Teilnahme! Haben Sie noch Fragen?

Appendix G: Assumption Testing for Model 1

Since the first model investigated differences between two groups, the assumptions for paired samples t-tests were tested prior to hypotheses testing of H1, H2 and H3 (Stone, 2010). The design of the study affirmed the assumption of paired samples and independence of measurements between participants (e.g., the measurement for participant 1 was independent from the measurement for participant 2). Moreover, the independent variable was measured on a nominal scale and has two categories (low TPC and high TPC), whereas the dependent variables (substress, phystress, SA and performance) were measured on an interval scale. Next, the difference scores of all dependent variables were tested for outliers and normal distribution. A step-by-step approach was used to reveal the implications of including or excluding outliers for changes in normal distribution and the results of the t-tests. It should be noted here that the assumption of normal distribution can be neglected for samples n > 30 (e.g., Stone). Overall, all graphs were sufficiently normally distributed. The step-by-step approach revealed that excluding outliers would yield in (slightly) better scores for normal distribution, however excluding outliers did not have a major effect of the results of the paired samples t-tests. Since the sample of *N* = 53 was greater than 30, it was therefore decided to not exclude the outliers.

Appendix H: Assumption Testing for Model 2

The second model investigated whether differences in SA-levels and performance between conditions can be explained by differences in stress-levels. For this purpose, the assumptions for repeated measures ANCOVAs (rmANCOVAs), with stress as a covariate, were tested. First, the design of the study affirmed the assumption of paired samples. Second, the within-subject factor was measured on a nominal scale with two characteristics (low TPC and high TPC) and the depend variable (SA or performance) as well as the covariate (substress or phystress) were measured on an interval scale. Specifically, stress was induced as a covariate by using the substress variable of the low TPC as a baseline and the difference score of substress between conditions as an indicator of change. For reasons of multicollinearity, both stress variables were centred. The same procedure was used for phystress. Third, normal distribution of the residuals was tested. For this purpose, rmANCOVAs were conducted and residuals were saved. Subsequent explorative data analyses revealed sufficient normal distribution for performance residuals. SA residuals partly showed slight deviations from normal distribution. Since the assumption of normal distribution is considered robust against violation (e.g., Salkind, 2010), the slight deviations for SA residuals could be neglected. Forth, no outliers were identified for SA or performance in both conditions. Testing for outliers in the SA and performance residuals revealed three outliers from which one negatively affected the normal distribution and the error term for SA residuals across all conditions and was therefore excluded. Since analyses without the other two outliers did not result in major differences for normal distribution and error terms, it was decided to not exclude them from the dataset. Fifth, sphericity is present if the differences between all conditions of the independent variable are equal and is commonly tested with Mauchly Test (Field, 2013). This study used two conditions (low TPC and high TPC) which resulted in one difference only, so sphericity is automatically given. Sixth, homogeneity of regression slopes should be given, i.e., the conditions should not differ on the covariate, otherwise the inclusion of the covariate into the analysis would not control or equalise differences between conditions. It was found that some conditions differed on the covariate and therefore had an inadequate homogeneity. For this reason, the outliers in stress variables, which were already known from assumption testing for the first model, were excluded. As a result, homogeneity of regression slopes was assessed satisfactory. Therefore, the assumptions for rmANOVA were met.

Appendix I: Assumption Testing for Model 3

The third model examined the effect of stress on SA and performance within conditions. To investigate whether subjective and physiological stress are predictors for SA or performance, the assumptions for multiple linear regression analyses were tested. First, residual analyses and partial regression diagrams were used to assess the linear relationship between predictors and the outcome variable. Whereas residual analyses were assessed satisfactory with data points gathering around zero, partial regression diagrams were satisfactory for subjective stress, but unsatisfactory for physiological stress for both outcome variables and in both TP conditions. Precisely, data points gathered at the lower end of the physiological stress scale and thus hardly showed any relation between physiological stress and SA or performance in both conditions. Also, the assumption of homoscedasticity was violated for the effect of stress on SA (H6a and H6b). Therefore, it was decided to transform the physiological stress variable with logarithmic transformation to dissolve the accumulation of data points at the lower end of the physiological stress scale. As a result, both the assumption of linear relationship between predictors and outcome variable and the assumption of homoscedasticity could be confirmed. Second, no outliers were identified across leverage values and Cook-Distance. One extreme value was identified for studentized excluded residuals in H6a and H6b. Since the extreme values were both related to the same participant, it was decided to conduct H6a and H6b including and excluding the respective participant. Third, independence of residuals was confirmed with Durbin-Watson test. Forth, correlation analyses did not show multicollinearity. Fifth, normal distribution of residuals was tested both analytical with Shapiro-Wilk-Test and graphical with a histogram and a P-P-plot. Except for slight deviations for H6a and H6b, normal distribution was overall satisfactory and could be approved due to the high sample size (Schmidt & Finan, 2018). Thus, the assumptions for multiple linear regression analyses could be confirmed.