European Commission, Horizon 2020

D4.4 Training Experience Framework and Structural Equation Model

SHOTPROS

Deliverable	D4.4
Deliverable Lead	Raimund Schatz
Related work package	WP4
Author(s)	Raimund Schatz, Sebastian Egger, Lisanne Kleygrewe, Till Bieg
Dissemination level	PUBLIC
Due submission date	31.10.2020
Actual submission	30.12.2020
Project number	833672
Instrument	RIA
Start date of project	01.05.2019
Duration	42 months
Version log	V1.1

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Versions

Vers.	Date	Author	Description
V0.1	05/09/20	Raimund Schatz (AIT)	First draft
V0.2	01/10/20	Raimund Schatz (AIT), Sebastian Egger (AIT), Lisanne Kleygrewe (VUA), Till Bieg (AIT)	Methodology and study description added
V0.5	05/11/20	Raimund Schatz (AIT), Sebastian Egger (AIT), Lisanne Kleygrewe (VUA), Till Bieg (AIT)	Results analysis added
V0.9	07/12/20	Raimund Schatz (AIT), Sebastian Egger (AIT), Lisanne Kleygrewe (VUA), Till Bieg (AIT)	Draft for Review
V0.95	21/12/20	Raimund Schatz (AIT), Till Bieg (AIT)	Final Version after Review
V1.0	28/12/20	Valerie Schlagenhaufen (USE), Gerhard Helletzgruber (USE)	Finalization and upload
V1.1	27/10/21	Olivia Zechner (AIT)	Additional reference in Executive Summary based on review feedback

List of Acronyms and Abbreviations

Acronym / Abbreviation	
VR	Virtual Reality
VirTra	Virtual Training Shooting Simulator
XR	Extended Reality
HMD	Head Mounted Device

UX	User Experience
ANOVA	Analysis of Variance
SOPI	Sense of Presence Inventory
QoE	Quality of Experience
TAC	Technology Acceptance
VAS	Visual Analogue Scales
QoLE	Quality of Learning Experience
Eud	Eudaimonic aspects
LMM	Linear Mixed Model
EFA	Explanatory Factor Analysis
RMSEA	Root Mean Square Error of Approximation
LEA	Law Enforcement Agency

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Table of Contents

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Table of Figures

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

1 Executive Summary

Deliverable D4.4 describes the outcomes and the process of the development of a **framework for measurement and modeling of VR training experience** in the context of training decision-making and acting of police force members. To this end, a range of different qualitative and quantitative experience assessment instruments from different research fields were adopted and customized to achieve the following goals:

a) comprehensively **assess end-user perception and training experience** of different training systems,

b) quantify the **relevance of the different experience dimensions** assessed and **model their impact** on training experience, and

c) empirically ground **recommendations on essential user needs and requirements** as well as which **measurement instruments** should be used to assess VR-based police training technologies and systems in future tasks and activities.

At the center of the empirical data collection is a **large VR training experience study** (ZüriVR) that was conducted in Q2+Q3 2020 with the City Police of Zurich, Switzerland. The study was designed to address five research questions related to the impact and trainee experience of two virtual training systems (VirtTra and Refense VR). Systematic results analysis explores the impact of and relationships between different experience-related constructs (quality of experience, acceptance, presence and immersion, etc.), compares the two virtual training systems tested, and clusters feedback from trainers and trainees. Analysis results and findings are then used to develop a) a framework of recommend measurement instruments and b) a model for relating relevant experience-factors assessed via different instruments to overall Quality of Learning Experience (QoLE).

Key results and findings of D4.4 are:

- Both virtual training systems were **highly positively received** by trainers and trainees alike, as reflected in feedback and ratings. VirTra received higher acceptance, enjoyment and ease of use ratings, while the VR (Refense) system resulted in higher immersion and presence.
- The **measurement** instruments that have turned out to be the most useful ones in terms of insight and information value, are: SOPI (presence & immersion), TAC (technology acceptance), VAS (stress and mental exertion), and QoLE (Quality of Learning Experience).
- Concerning experience **modeling**, we found that linear mixed models (LMM) deliver the most satisfying results. We developed a framework of three models that can be used to

predict the central construct of Quality of Learning Experience (QoLE) using instruments like SOPI, TAC or VRQoE that cover key dimensions of the VR experience.

- Regarding **implications for the design of future VR-based training systems**: realistic weapon handling, realistic 3D content and realistic, high-quality audio rendering are essential requirements that need to be fulfilled. Strong support of trainee interaction with objects, non-playable characters and team members is essential. An additional artificial pain stimulus (shocker) has only minor impact on experience and stress levels.

These results directly inform the different project activities of SHOTPROS work packages WP3 to WP8. For example, knowledge of the relevance of different experience dimensions will aid the development of training concepts and curriculum propositions (WP3). In addition, they provide measurement tools for later studies (WP6) and the field trials (WP7). These measurement tools allow to assess the different experience facets that are triggered by the system under evaluation, but also allows to compare different systems with regard to training experience delivered, which will also become an important part of the envisaged policymaker toolkit (WP8). Furthermore, the results can equally be used in projects other than SHOTPROS for purposes of training evaluation for continuous quality control as well as virtual training technology benchmarking.

Recommendations and specific guidelines on best practice scenarios and the experiences covered above will be provided in deliverable **D7.6 SHOTPROS Final Guidelines for VR Training** (M41), with the aim to provide VR guidelines for an ideal but still flexible training environment to support a variety of training needs of LEAs.

2 Assessment of VR Training Experience Dimensions

This section describes the theoretical and practical project background of the experience measurement framework development in WP4.

2.1 Introduction and Background

One key goal of SHOTPROS work package WP4 is to develop a framework of virtual training experience measurement methods and models that can be applied to human-factors studies, system evaluations as well as assessment of training experience after the training for continuous training evaluation for quality control in the context of VR-based training for police forces that are part of subsequent activities inside (WP3-8) and outside the project.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

As first step towards achieving this goal, the consortium originally planned to identify the most relevant experience dimensions/factors in the context of virtual training for DMA and investigate suitable means and instruments to reliably measure them. However, due to the COVID-19 pandemic situation in 2020 (which triggered an array of lockdown and social distancing measures in the different participant countries) as well as the emerging collaboration with the city police of Zurich (Switzerland), the project team chose to adapt its methodological approach: instead of conducting a series of smaller lab user experiments for exploring key experience dimensions of VR training (WP6), the project took advantage of the opportunity to conduct a large user experiment (involving more than 700 police officers) in the context of an extensive training campaign by the city police in Zurich that featured virtual police training on behalf two different operational systems (cf. deliverable D3.3 for details).

In this context, WP4 adopted and developed a range of different qualitative and quantitative experience assessment instruments from different research fields (User Experience, Quality of Experience, Technology Acceptance, VR & Presence Research, etc., see upcoming Sections 2.2 and 2.3) with the purpose to

a) **comprehensively assess end-user perception and training experience** of the two concrete systems at hand,

b) **quantify the relevance of the different experience dimensions** assessed and model their impact on training experience, and

c) **empirically ground recommendations which measurement instruments should be used** to assess VR-based police training technologies and systems in future tasks and activities.

Note that the finalization of this deliverable was delayed for several COVID-19 related reasons. First of all, relevant user studies (like the Berlin human factors study) had to be cancelled due to travel and safety restrictions. Ultimately, the ZüriVR study became the central source of empirical data for D4.4. AIT received the final debugged ZüriVR dataset by beginning of September. Before that, AIT was active in identifying potential bugs in the dataset and had intensive exchange with VUA on that matter. These issues and the multi-factorial nature of the investigated research questions led to a longer, more complex process of data analysis with respect to the data acquired.

2.2 Quantitative assessment

Quantitative assessment of end-user experience of immersive VR systems (and resulting opinions) has a long tradition in user experience, quality of experience, VR and technology

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

acceptance research (cf. Perkis et al., 2020; Cipresso, 2018; Bayerl et al., 2019). In quantitative assessment setups, defined stimuli are presented to participants who either intentionally (e.g., by answering rating scales) or non-intentionally (e.g., through behavioral actions or physiological responses) provide qualitative and/or quantitative measures of different aspects of their experience with the system or technology.

In general, the definition of proper stimuli and measurements is not trivial when conducting assessments under realistic conditions of technology use. Therefore, the main goal of the assessment activities described in this deliverable was to develop measurement instruments and experience metrics that remain ecologically valid when being utilized in the context of the project's VR application domain (police training), ensure sufficient reliability (e.g., in terms of errors and noise), and deliver diagnostic value. For this reason, we decided to focus on subjective experience and acceptance testing based on explicitly inquiring participants' feedback to specific aspects of interest of their virtual training experience on behalf of questionnaires that trigger necessary introspection processes and capture participant opinions.

Furthermore, to leverage the large number of participants available in this study, we decided to work with a range of different instruments (Table 1) in order to

a) cover a **broad range of experience dimensions** and perspectives, and to
b) obtain evidence on **which subset of instruments should be used for future assessment** of the experience of VR-based training for police officers.

We also adapted and extended the chosen instruments to better address the specifics of VR-based police training. For example, we extended the QoE and UX related questionnaires with questions on eudaimonic aspects (self-actualization, fulfilment) in order to complement hedonic and pragmatic aspects of experience (Hammer et al., 2018)), which were tested using the short version of the AttrakDiff semantic differential by Hassenzahl et al. (2008). For technology acceptance, the questionnaire of Huang et al. (2013) was slightly modified since it originally targeted VR training in a medical context. Similarly, the questionnaire on Quality of Learning Experience is based on the framework of Kirkpatrick & Kirkpatrick (2006), but uses a wording adapted to the context of police training (see Appendix A for a description of the questionnaires used).

Table 1: Instruments used for Virtual Training Experience Assessment.

Instrument / Construct	Experience Dimensions	References

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Sense of Presence Inventory (SOIP)	Presence Immersion Realism	Lessiter et al. (2001)
Quality of Experience (QoE) & User Experience (UX)	Perceived quality Pragmatic aspects Hedonic aspects Eudaimonic aspects	Möller & Raake (2014), Hammer et al. (2018), Hassenzahl et al. (2008)
Technology Acceptance (TAC)	Ease of Use Usefulness Intention to use Imagination Immersion Interactivity Enjoyment	Vekantesh & Bala (2008), Huang et al. (2013)
Visual Analogue Scales (VAC)	Experienced stress Mental exertion	Houtman & Bakker (1989), Zijlstra (1993)
Quality of Learning Experience (QoLE)	Self-Efficacy Assessment	Kirkpatrick & Kirkpatrick (2006)

2.3 Qualitative assessment

Police instructors' and trainees' experiences, perceptions and resulting opinions play an essential role when it comes to the deployment of a new training technology. In this context, it is important to identify potential expectations of both stakeholder groups, perceptions of usefulness and the added value of virtual training systems to existing training plans. This information can support informed decisions about suitable areas of application and possible improvements. Moreover, it is also important to also consider individual perceptions of the participants taking part in the different trainings as they can provide deep insight into the perceived usefulness of these systems and implications for the improvement. However, quantitative methods are limited in this regard which suggests the application of qualitative methods. Qualitative methods are a powerful tool to shed light on individual perceptions and experiences and to identify reasons as to why a training system is positively assessed or not. In this way, they can also help to understand reasons behind the quantitative assessments.

Thus, qualitative interviews (e.g., Adams, 2015) with both participant groups in the training (trainees), and police instructors operating the training systems (trainers) are employed in addition to the quantitative assessment. Data obtained from the interviews is analyzed by grouping statements from the interviewees into comprehensive themes using thematic analysis (Braun & Clarke, 2006).

3 Virtual Training Study (ZüriVR)

This section describes the main study that underlies the WP4 training experience framework and model development. After the description of the study and research questions, its qualitative and quantitative results are presented and discussed.

3.1 Study Background and Introduction

As an outcome of the collaboration with the City Police of Zurich (Switzerland), a training study was conducted in the summer of 2020 in the context of a large virtual training campaign. Two commercial systems for virtual were used: a 2D cinema-based system (vendor: ViTra[1]), as well as a full-fledged HMD-based VR system offered by the company Refense AG[2]. Since the Refense VR system uses a similar technology and device setup as the envisaged SHOTPROS system (VR HMDs, multiple participants move freely around in a dedicated area), the ZüriVR study results and findings are highly relevant and directly transferable to the project.

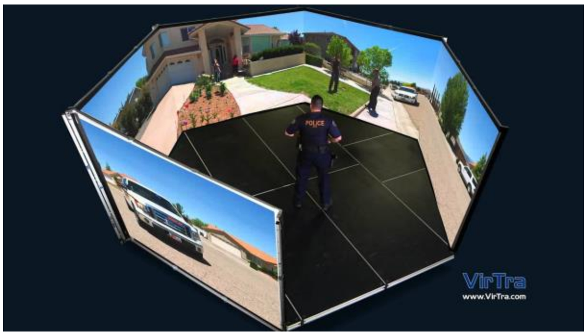
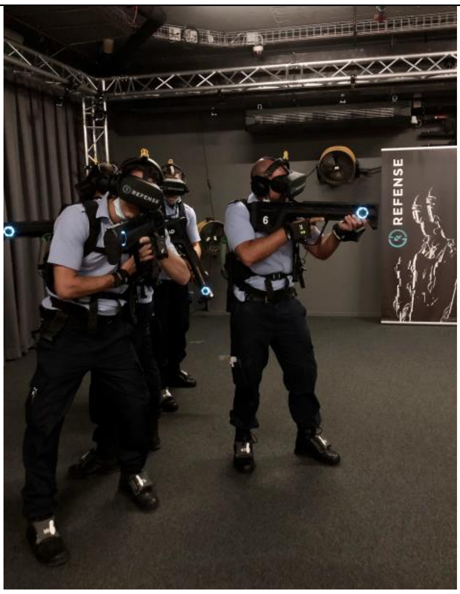
Table 2: Description of the two systems used in the ZüriVR study.

Technology 1: VirTra	**Technology 2: VR (Refense)**

[1] https://www.virtra.com/

[2] https://www.refense.com/

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

 	
Decision-making simulation and firearms training simulator	Decision-making simulation and firearms training simulator
5 multi-screen 2D technology 300-degree immersive training environment	HMD-based stereo display with audio Training environment for max. 10 participants on >200 m²
Training Scenario: 2 participants at the same time Multiple smaller scenarios Active threat De-escalation 10-14 minutes of active scenario	Training Scenario: 4 participants at the same time 1 large scenario Consisting of multiple “layers” 12-15 minutes of active scenario

In the remainder of the document, the two systems are referred to as “Virtra” and “VR” for brevity.

3.2 Study Goals and Research Questions

As a result of extensive consultations with LEAs in six workshops (Amsterdam, Selm, Brussels, Berlin, etc., total n= 60) and a review of the state of the art in training and VR (cf. Deliverables

D2.2 and D2.3 for more details), we derived the following five research questions related to assessment and modeling of VR-based training from the research agenda of D3.2:

- RQ1: How well do trainees differentiate between the **key experience dimensions** of virtual training?
- RQ2: What is the overall level of **acceptance** of the different experienced training systems from the trainee perspective? What are the **key factors** that influence acceptance?
- RQ3: How do trainees assess the **training effect and utility** of the technology?
- RQ4: How do **trainers assess the deployment** of such a system in their training routines?
- RQ5: Does the presence of a **pain stimulus** influence the experience?

These questions have practical as well as scientific relevance as they address a) general aspects related to which factors and qualities are relevant for VR-based police training, b) which aspects drive training success and acceptance of VR technology, and c) which measures should be taken to improve VR based training for police officers. In this respect, answering these research questions is also relevant for the development of training concepts and training curriculum in WP3.

3.2.1 Background and rationale behind the research questions:

RQ1: For a strong learning experience and possible successful implementation of a training technology, we assume that it is important that different **experience dimensions** (like presence, immersion, perception, quality perception, and acceptance) are positive. However, the question is to which extent these different experience dimensions are relevant and orthogonal to each other in the context of virtual police training. This knowledge is relevant for effectively designing, benchmarking and improving police training support technology.

For example, part of a positive experience with a training technology is the experience of a high sense of presence. For police training using the VR and the VirTra systems, this means that the officers find the virtual environment realistic, can be immersed in it, and hardly experience any adverse effects (such as cybersickness). These aspects are important, since for participants to develop salient and useful embodied decisions and motor heuristics, both sensory input and cognitive input is required simultaneously. This would require both sense of presence and the ability to move naturally and thus have proper propriocepsis and action affordances.

Furthermore, human experience of technical systems can be characterized along different key dimensions which, together, constitute quality of experience (QoE, i.e. degree of annoyance

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

or delight when using a system) and thus determine success or failure from the end-user's perspective. The key dimensions of the quality of experience refer to hedonic (pleasure, fun, etc.), pragmatic (helpful in achieving goals/completing tasks), but also eudaimonic (personal growth, self-actualization) aspects are particularly relevant in training contexts (Hammer et al. 2018).

RQ2: In addition to sense of presence and quality of experience, high levels of **acceptance of a training technology** are important as in practice they increase utilization, compliance, and ultimately, positive effects of learning. Assessing the acceptance of a given technology helps to identify the extent and likelihood of its actual adoption as a training tool in terms of voluntary everyday usage. In addition, we want to know the underlying drivers of acceptance as well as how acceptance relates to the overall training experience.

RQ3: in order to measure and optimize usefulness and impact of virtual training technologies, it is essential to assess how they affect the overall **quality of learning experience of trainees**, including pragmatic aspects in terms of perceived usefulness. The quality of learning experience can be assessed with a short questionnaire that asks trainees how effective they regard the training to be for their own learning process. Since this such perceived effectiveness is such a central topic, we want to relate all other experience dimensions to it in terms of training experience models (see Section 4.2). In addition, with explanations from respondents, answering this question yields qualitative and quantitative input in terms of requirements as well as critical aspects of virtual police training technology. These answers (along with RQ2 insights in acceptance-related factors) will also aid the development of training concepts and curricula in WP3.

RQ4: in addition to trainees, **trainers** (police instructors) are a critical stakeholder group, too, because they have the responsibility of providing the most efficient and effective training to prepare their trainees for their tasks and missions. Thus, trainers' opinions play an essential role when it comes to procurement and actual deployment of new training tools and technologies. Thus, it is important to explicitly gather trainer opinion, in terms of requirements as well as critical aspects of virtual police training technology. Doing so also ensures that demands, requirements and expertise of LEAs is properly included in the project.

RQ5: when training police officers with reality-replicating technologies under conditions of stress, adding the presence of a **pain stimulus** might enhance the perception of realism by simulating the threat of getting hurt or injured. In addition, the pain stimulus might also strengthen the learning experience of police officer by providing more immediate feedback on their performance during the training. To determine whether the presence of a pain stimulus truly adds value by enhancing realism and immersion of police officers in reality-

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

replicating training technology, it is important to compare the stress/anxiety, the quality of learning experience and the immersion police officers experience when performing training scenarios with a pain stimulus and without a pain stimulus present.

3.3 ZüriVR Study Description

This section provides an overview of the study design and assessment methods used. For more details, please refer to deliverable D3.3.

3.3.1 Test Protocol

To interfere as little as possible with the organization and execution of the training campaign, maintain data quality and avoid delays wherever possible, data needed to be gathered in such a way that little time or effort is required from participants and instructors. For a visual representation, the table below shows the overall structure of the training day as scheduled by the Zurich City Police. The yellow shaded parts of the table show the moments during the training days at which the measurements were performed.

Table 3: Plan for Daily Training Schedule of ZüriVR Training. Study-relevant slots are shaded in yellow. Questionnaires were filled out directly after execution of the respective training.

Time	Variant A	Variant B	Variant C
07:00 - 07:10	Welcome		
07:15 - 07:45		Theory contact communication	
08:15 - 09:15	VR	VirTra	Contact communication
09:15 - 09:45	Break		
09:45 - 10:45	VirTra	Contact communication	VR
11:15 - 12:15	Contact Communication	VR	VirTra
12:30	Recap		

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

3.3.2 Quantitative Assessment

After each virtual training (either performed using the VirTra or the Refense VR system), all participants filled in a short questionnaire (using iPads) that assessed the overall quality of the learning experience from the VR training. Furthermore, participants completed two visual analogue scales for experienced stress as well as mental effort.

This short questionnaire was followed by one of the three specific, more detailed questionnaires below, depending on which group (1,2,3) the participant was randomly allocated to:

a) A questionnaire that measures the sense of presence of the virtually displayed environment (ITC-SOPI), or
b) A questionnaire that measures the quality of experience (QoE), or
c) A questionnaire that measures the acceptance of technology (TAC).

Each participant completed only one of these specific questionnaires. The questionnaires were distributed in such a way that an equal number of responses were gathered for each questionnaire group. Note, that it would have been the ideal case if every participant would have answered all three detailed questionnaires. However, due to time limits in the given training schedule as well as the risk of low data quality by overwhelming respondents, the number of detailed questionnaires per participant was deliberately limited to one.

Table 4: Questionnaires used for quantitative assessment (see Appendix A for details).

Questionnaire / Scale	Acronym	Sub-Scales	Items	Group
Sense of Presence Inventory	**SOPI**		**43**	1
		Spatial Presence	19	
		Engagement	13	
		Ecological Validity	5	
		Negative Effects	6	
Quality of Experience	**QoE**		**18**	2
		VRQoE ACR-5 MOS scale	5	
		AttrakDiff-short questionnaire	10	
		Eudaimonic aspects	3	
Technology Acceptance	**TAC**		**45**	3
		Perceived Ease of Use	3	
		Perceived Usefulness	4	
		Intention to use	4	
		Imagination	4	
		Immersion	3	

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

		Interaction Perceived Enjoyment Technology Curiosity	4 3 5	
Visual Analogue Scales	**VAS**	Experienced stress Mental exertion	**1** **1**	All
Quality of Learning Experience	**QoLE**	Self-Efficacy Assessment	**4**	All

3.3.3 Qualitative Assessment

Qualitative interviews were conducted on selected training days with 22 participants (trainees) and 4 trainers in total. Interview guidelines for both trainees and trainers are provided in “Appendix B: Guidelines for Qualitative Interviews”.

Trainees were interviewed after completing a training session (VR or VirTra). Accordingly, selected participants (trainees) answered open questions to reflect on their training experience. Questions targeted general positive and negative aspects of the trainings regarding their effectiveness and usefulness (RQ3).

Furthermore, at the end of selected training days, individual semi-structured interviews with the trainers or a joint focus group meeting with multiple trainers were conducted to reflect on their experience with the training systems as an additional training modality (RQ4). As for the trainees, trainers were asked about positive and negative aspects of the trainings from their individual point of view. In addition, the use of gamification elements and the suitability of the training systems for specific training areas (e.g., training of tactical aspects) was discussed with the trainers.

3.4 Results Dataset Description

The statistical analysis of the data was carried out by AIT using R, version 4.0.2 (R Core Team, 2020). The raw study results dataset was first cleaned and preprocessed to be then further analyzed in order to answer the different research questions RQ 1-5.

The original raw dataset contained 1120 observations in total. Three groups of participants were divided by the type of questionnaires they filled in: 1) SOPI, 2) QoE, Attrakdiff and Eudaimonic aspects, and 3) TAC. In addition, all participant groups filled in the QoLE questionnaire and Visual Analogue Scales (VAS) for Stress and Mental Effort.

With regard to cleaning and preprocessing, 68 observations (or: cases) were excluded from the original dataset. These are suspicious cases or cases with missed crucial data: duplicates, typos in participant ID's, missing training type, repeated training type, or participants who were marked in different groups. Additionally, we applied filtering by the standard deviation (SD) - extra 47 cases were excluded. That is, cases that had almost the same score throughout the whole questionnaire (*SD*[3] <= 0.4).

The final analysis dataset consists of 1005 cases (Group 1 (SOPI): n = 384, Group 2 (QoE, Attrakdiff and Eudaimonic): n= 319, Group 3 (TAC): n=302) from 596 participants. Since we deliberately did not exclude cases that had missing data within their questionnaires, we applied additional filtering of cases because missing data only when focusing on a particular questionnaire/experience dimension, when completeness of the specific data was required. In all three user groups of the analysis dataset, participants were trained by two training systems (VirTra and VR). The majority of participants that took part in VR study also took part in VirTra study. However, VirTra sample size is bigger than VR (n= 560 vs. 445 respectively). Figure 1 shows the proportion of responses regarding groups and training systems.

Our analysis sample contains 487 male participants (82%), 101 female participants (17%) and 8 people did not specify sex (1%). Participants' age ranges between 21 and 65 years old (*M*[4] = 38.56; *SD* = 9.29; *Mdn*￼ = 37; 11 missing values), and their experience varies between 2 and 42 years [5]*M* = 12.74; *SD* = 9.14; *Mdn* = 10; 14 missing values). Figure 2: Age distribution between male and female participants. Participants with gender information missing (n=8) have been excluded. shows the distribution of age between males and females in the analysis dataset.

[3] SD = Standard Deviation
[4] M = Mean
[5] Mdn = Median

Figure 1: Distribution of data by groups/questionnaires and training systems (VirTra and VR).

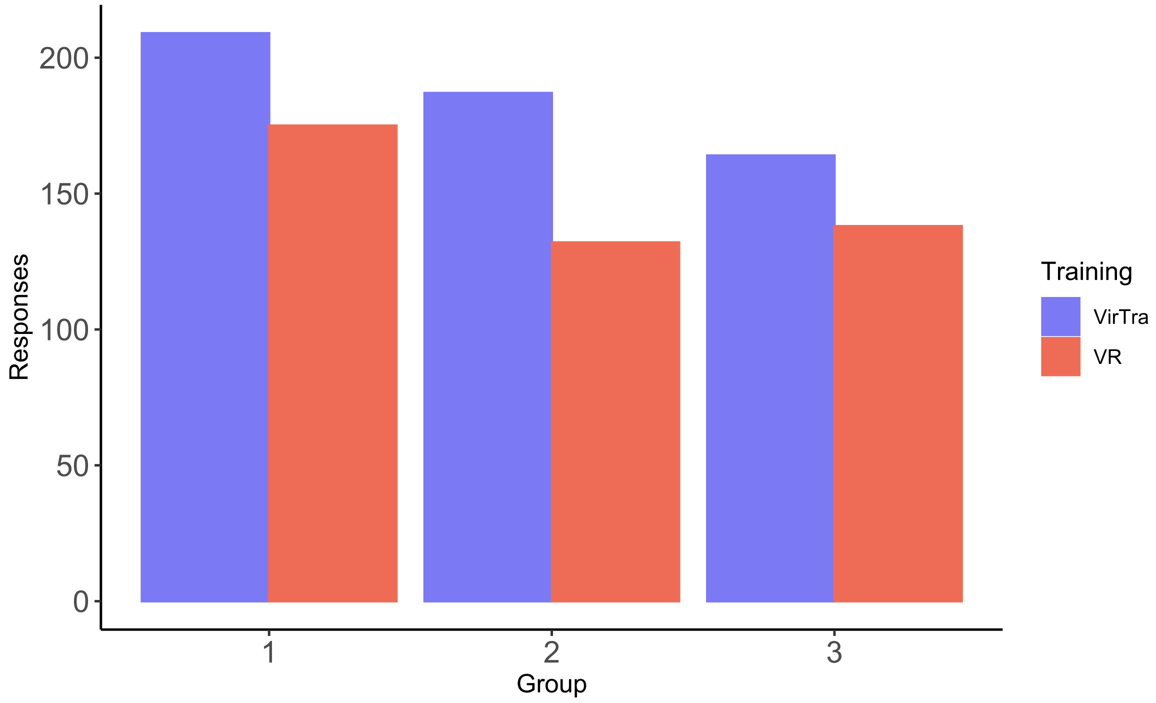
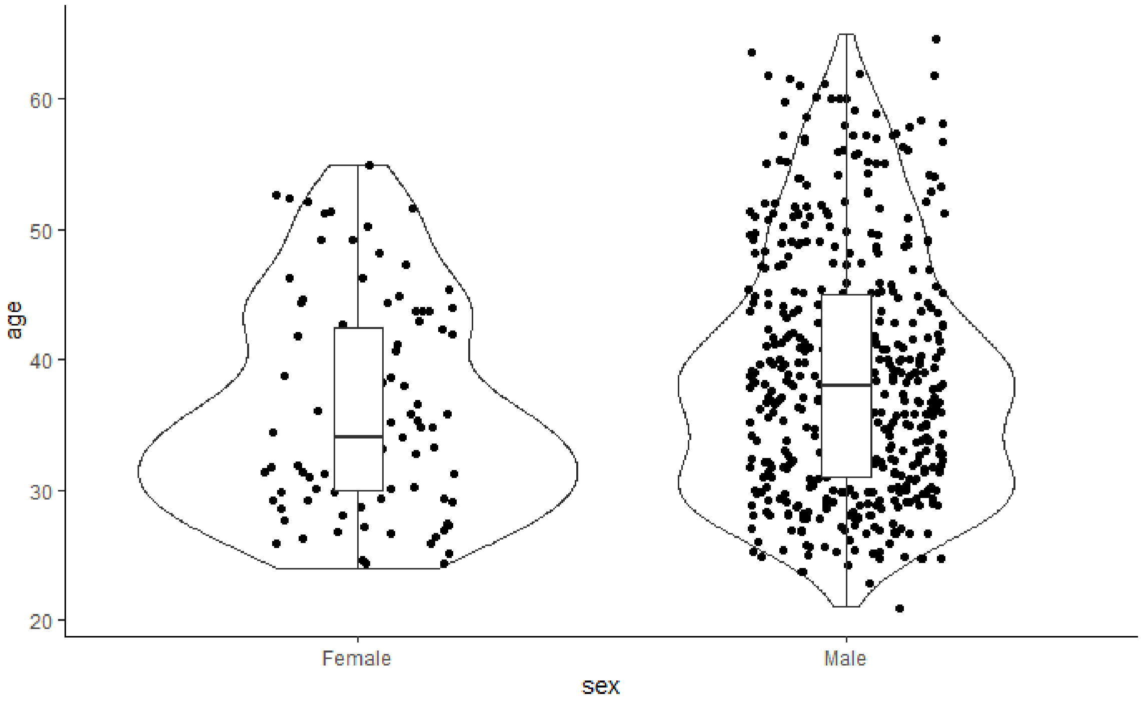


Figure 2: Age distribution between male and female participants. Participants with gender information missing (n=8) have been excluded.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

3.5 Quantitative Results

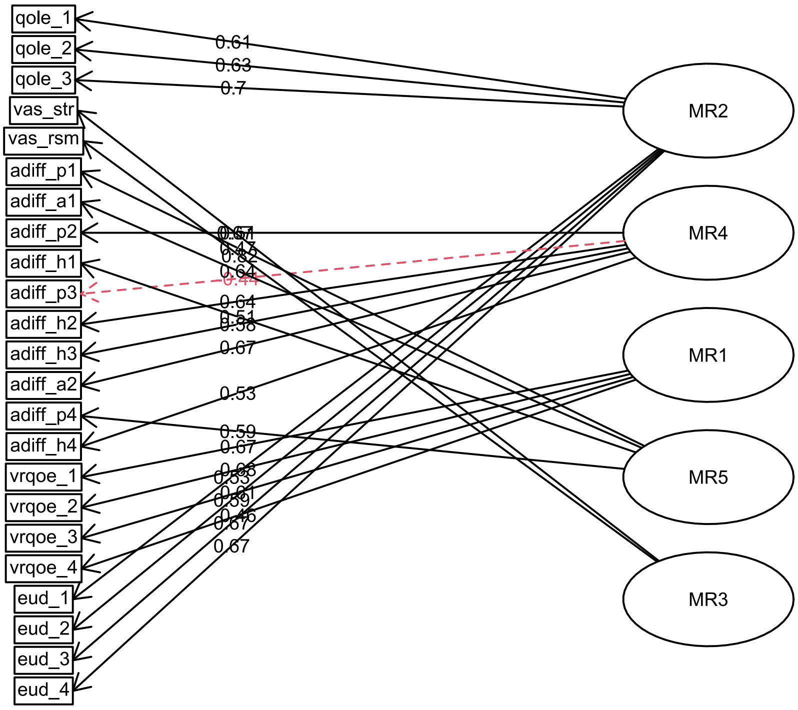
3.5.1 Research Question RQ1: How well do trainees differentiate between the key experience dimensions of virtual training?

In the context of factor exploration, common factor analysis methods were used. The fundamental suitability of the available data for factor structure was verified based on the correlation matrix (Spearman) and the Kaiser-Mayer-Olkin criterion. It's worth to mention that Kaiser Criterion Scree plot and Horn's Parallel Analysis Scree plot defined less factors for Technology Acceptance (TAC) and short Attrakdiff questionnaires that were expected. Exploratory Factor analysis (EFA) also shows mixed combination of items inside factors for these two questionnaires, i.e., factors according to EFA had different from original papers structure, particularly in the case of Attrakdiff. Figure 3 illustrates the result of EFA (varimax rotation) for Group 2 (Attrakdiff, Eudaimonic and VRQoE questionnaires). It shows that most questionnaire items in that group cluster according to the intended constructs, except for AttrakDiff, where items related to hedonic, pragmatic aspects and attractiveness visibly are mixed up between those factors[6]. This confirms that in the context of the given study, the AttrakDiff short questionnaire utilized did not work as intended and that the items represent different constructs than originally intended. Apart from this, most questionnaires proved to be suitable for the intended factor structure even though factors could not always be perfectly discerned in the data.

[6] The item adiff_p3 which gauges the predictability of the system has a reversed meaning in the context VR systems and therefore was excluded from further analysis.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

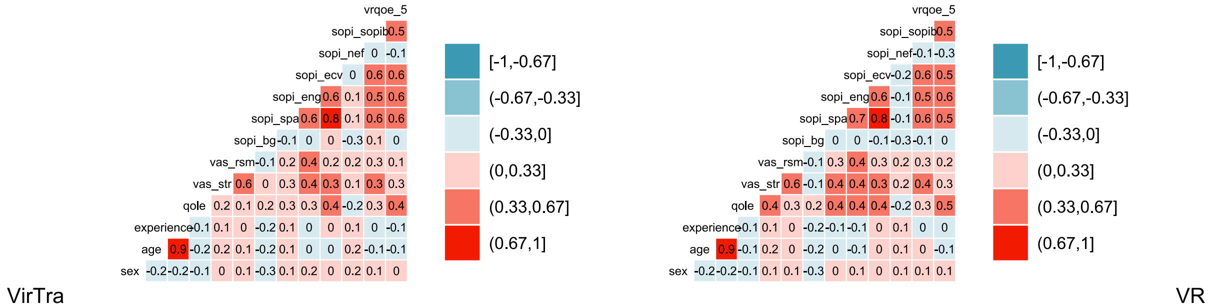
Figure 3: Exploratory Factor Analysis diagram for Group 2 constructs (including Attrakdiff, Eudaimonic and VRQoE questionnaires). Ellipses on the right are factors that influence the different item-variables on the left. Numbers on arrows are factor loadings.



To define key experience dimensions (factors) for both training systems (VirTra and VR), we applied Spearman-rank correlation matrices on factor level. Spearman correlation was used since the initial data are represented by sets of ordinal ranked scores rather than interval. However, Pearson correlation coefficient shows very similar results.

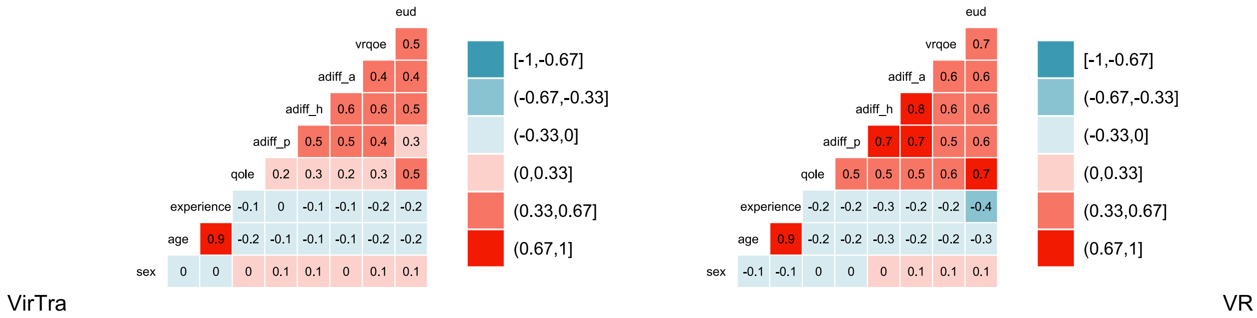
Group 1 (SOPI questionnaire) shows high positive correlation among SOPI factors (r_s=0.6-0.8) - besides "Negative Effects" item - and with "Stress" parameter (r_s=0.6) for both training system. For the VR training system we additionally found fairly high correlation among SOPI and QoLE (r_s=0.4).

Figure 4: Spearman Correlation Matrix for VirTra (n= 209, left) and VR (n=175, right) in Group 1 (SOPI).



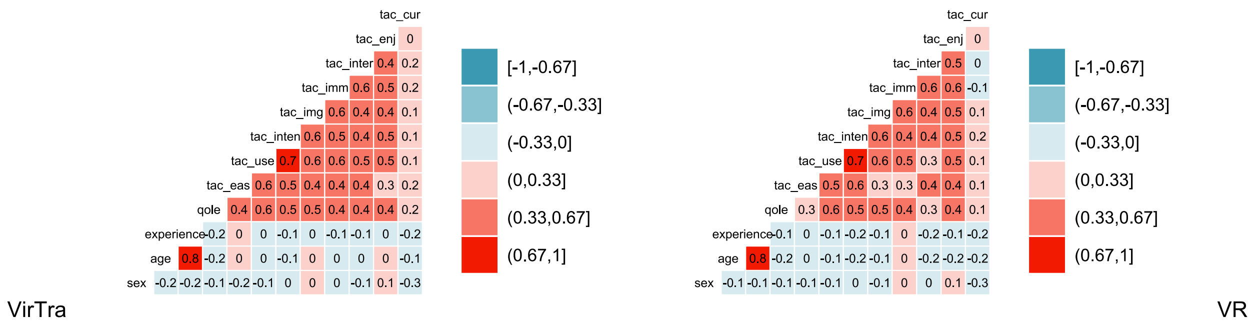
In Group 2 (Attrakdiff, Eudaimonic aspects, QoE) we observe medium to high positive correlations among Attrakdiff factors in both training systems, however in VR training system we see stronger correlations among factors (r_s=<0.7-0.8). We can also report positive correlation among QoLE and Eudaimonic aspects questionnaires, and again, this correlation is stronger for the VR training system (r_s=0.7).

Figure 5: Spearman Correlation Matrix for VirTra (n=186, left) and VR (n=131, right) in Group 2 (Attrakdiff, Eudaimonic aspects, QoE).



In Group 3 (TAC) there is fairly high positive correlation between TAC factors (except “Curiosity”) and QoLE (r_s=0.3-0.7).

Figure 6: Spearman Correlation Matrix for VirTra (n= 164) and VR (n=138) in Group 3 (TAC).



Answer to research question RQ1:

Our findings suggest that users do differentiate between the different experience dimensions, but only to a limited extent (correlations between factors are approx. r_s~0.6 (Spearman rank) within vs. r_s~0.4 across experience dimensions). Maximum correlation between two experience factors found was r_s=0.8 (Spearman rank correlation). Together with the factor analysis this result suggests that the questionnaires used in the study are valid. The results also suggest that the different experience dimensions tested are independent from each other to a certain extent and thus dimensions can only be reduced with great care.

In general, the correlation matrices show no pronounced correlation between experience factors and user attributes (e.g., sex, age, years of experience, or user background with computer/VR systems).

3.5.2 Research Question RQ2.1: What is the overall level of acceptance of the two experienced training systems from the trainee perspective?

To understand overall level of acceptance of the two training systems we analyzed the score means of each factor at Technology Acceptance questionnaire. The data shows that the overall level of technology acceptance (score means) is fairly high (3.3-4.4 score on 5-Likert scale). Means comparison by system shows slightly higher preference of VirTra training system compared to VR. We found are significant different for “Imagination", "Immersion", "Perceived Enjoyment" factors according to t-test and Wilcoxon tests, and additionally "Intention to Use" factor (t-test only). We excluded “Curiosity” factor from the analysis since this factor measures general users’ preferences (e.g., “I love owning new electronic devices.”) and not system preferences, so “Curiosity” related questions were answered only once.

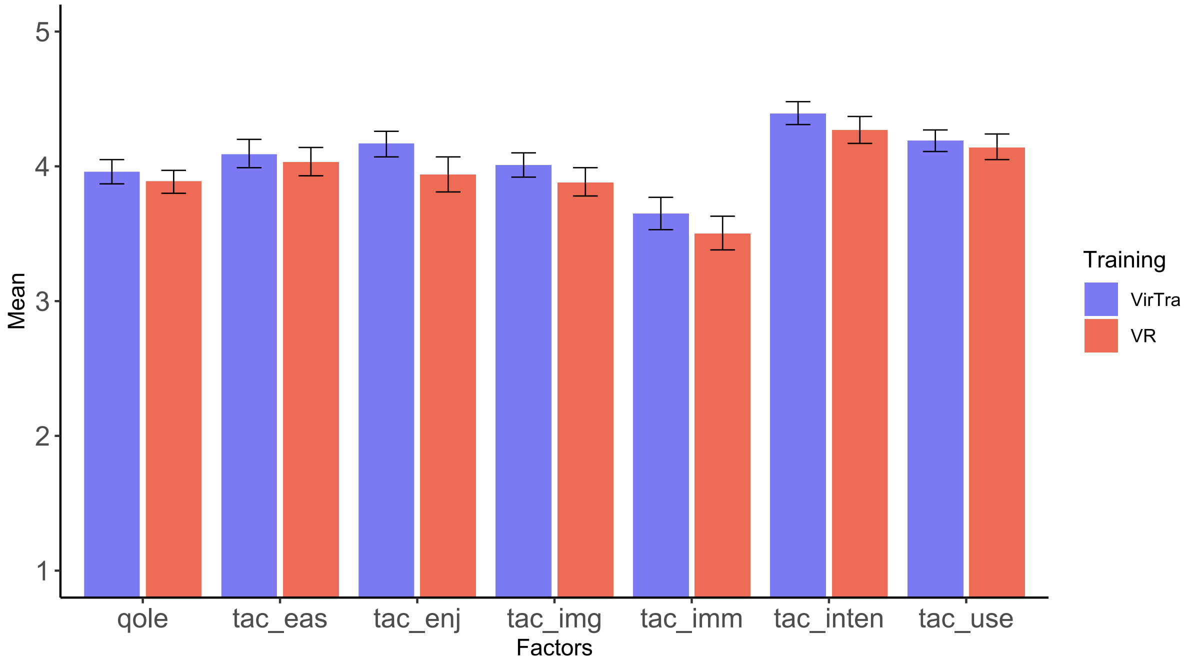
Moreover, there is significant preference for VirTra in “Quality of Leaning Experience” (QoLE) in Group 3. Respectively, we see higher “Mental Effort” scores in VR, which also may be interpreted as preference for VirTra. Table 5 and Figure 7 below show the comparison of score means per each factor.

Table 5: Factor-level comparison of the score means for Group 3 (TAC): paired t-test and Wilcoxon test.

Factor	VirTra	VR	t-test	Significance	Wilcoxon test	Significance
qole	4.019	3.903	0.036	*	0.040	*
vas_str	56.542	60.133	0.063		0.104	
vas_rsm	63.000	68.192	0.022	*	0.006	***
tac_eas	4.136	4.069	0.369		0.181	
tac_use	4.198	4.158	0.510		0.958	
tac_inten	4.419	4.302	0.032	*	0.065	
tac_img	4.021	3.869	0.020	*	0.037	*
tac_imm	3.672	3.500	0.025	*	0.035	*
tac_inter	3.402	3.300	0.185		0.079	
tac_enj	4.189	3.961	0.000	***	0.002	***

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Figure 7: Means of the different training experience and acceptance related factors (Group 3). All error bars in this deliverable represent 95% confidence intervals.



Besides Technology Acceptance (Group 3), we have also looked at score means of the factors from the other groups. We wanted to see did participants assess training systems in SOPI, Attrakdiff, Eudaimonic and Quality of Experience questionnaires correspondently to TAC. The results show that this is not the case. Thus, all SOPI factors in Group 1 show significant differences between two training systems. Most of them (“Spatial Presence”, “Engagement”, and “Ecological Validity”) report significantly higher score for VR training system, which can be explained that VR system provides more immersive/natural experience. “Negative Effects” SOPI factor also has significantly higher score in VR than VirTra, what in this dimension signifies better assessment of VirTra. However, overall level of negative effects still stays low (M=1.82 for VR vs M=1.51 for VirTra). At the same time, QoLE and Visual Analogue Scale do not reveal significant differences between systems. Table 6 and Figure 8 below show the comparison of score means per each factor for Group 1 (SOPI).

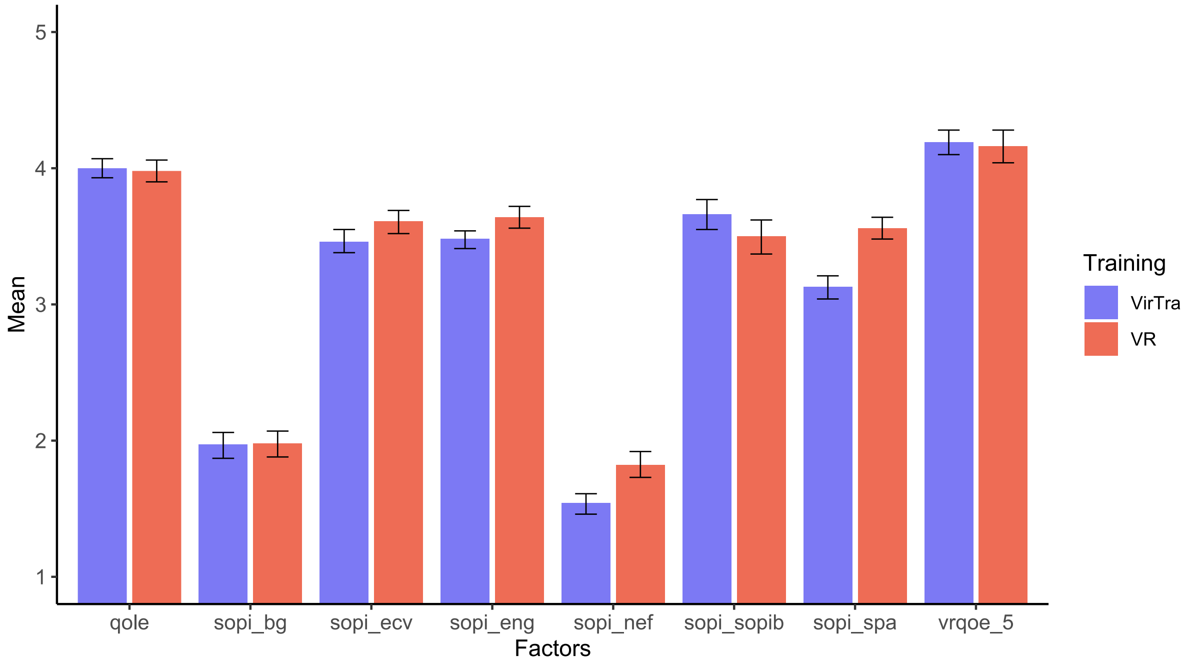
Table 6: Factor-level comparison of the score means for Group 1 (SOPI) using paired t-test and Wilcoxon test.

Factor	VirTra	VR	t-test	Significance	Wilcoxon test	Significance
qole	4.022	3.982	0.357		0.286	

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

vas_str	59.357	58.571	0.659		0.587	
vas_rsm	68.911	65.821	0.121		0.120	
sopi_spa	3.081	3.558	0.000	***	0.000	***
sopi_eng	3.468	3.655	0.000	***	0.000	***
sopi_ecv	3.456	3.610	0.003	***	0.001	***
sopi_nef	1.512	1.820	0.000	***	0.000	***
sopi_sopib	3.625	3.494	0.133		0.115	
vrqoe_5	4.173	4.185	0.859		0.781	

Figure 8: Means of the different sense of presence related factors in Group 1 (SOPI). All error bars in this deliverable represent 95% confidence intervals.



As regards Group 2 (QoE), we find the results in this group to be the least consistent for interpretation, even though some factors in this group indicate significant distinctions between training systems. Participants reported significantly more stress and mental effort (Visual Analogue Scale) for VR than for the VirTra training system, but on the other hand, higher levels of hedonic qualities (t-test and Wilcoxon test) and attractiveness (Wilcoxon test) for VR in Attrakdiff (see Table 7 below).

Table 7: Factor-level comparison of the score means for Group 2 (Attrakdiff) using paired t-test and Wilcoxon test.

Factor	VirTra	VR	t-test	Significance	Wilcoxon test	Significance
qole	3.988	3.980	0.890		0.990	
vas_str	63.096	67.509	0.021	*	0.011	*
vas_rsm	71.588	81.991	0.000	***	0.000	***
adiff_p	5.170	5.108	0.567		0.896	
adiff_h	5.621	5.796	0.032	*	0.002	***
adiff_a	5.434	5.548	0.272		0.028	*
vrqoe	4.101	4.022	0.260		0.317	
eud	3.860	3.914	0.409		0.107	

Answer to RQ2.1:

Overall levels of acceptance (TAC) are fairly high for both training systems. Even though participants slightly prefer VirTra (significant differences for 3 out of 7 factors), VR acceptance levels are also generally high and the differences between both systems are small in practice.

Note, that comparison of the score means of the SOPI results indicates significantly better immersive experience for VR. However, negative effects from VR are also significantly higher than for VirTra, which should be taken into consideration.

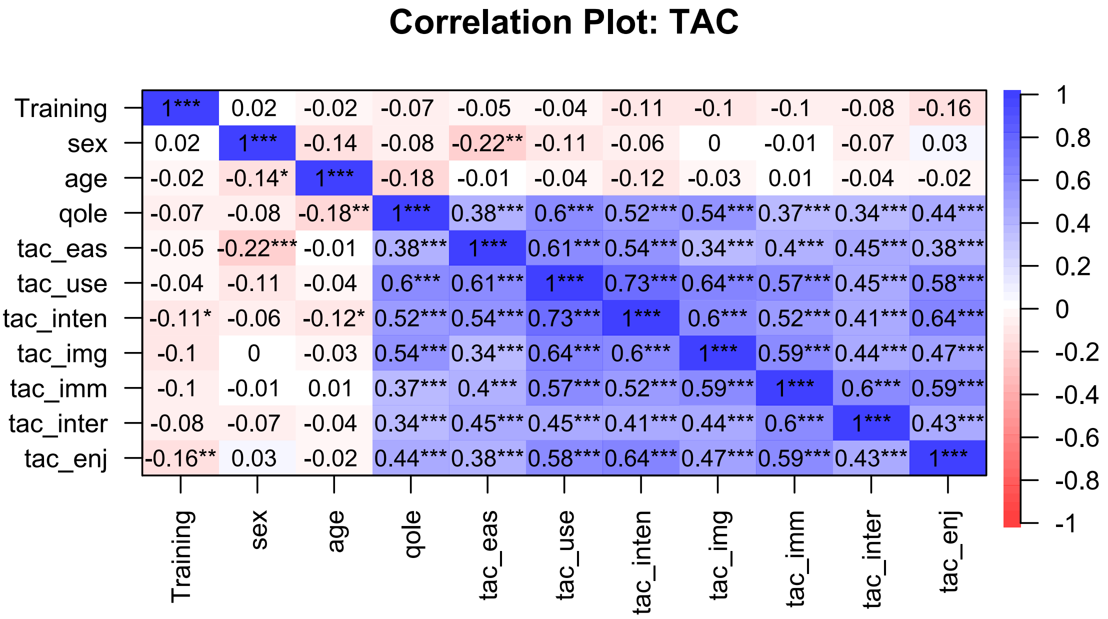
3.5.3 Research Question RQ2.2: What are the key factors that influence acceptance?

3.5.3.1 Correlation Scatter Plot Matrix

To identify the key factors that influence acceptance (i.e., “Intention to use”), we looked at the pairwise correlation scatter plot below and applied linear mixed model on the TAC questionnaire and demographic data. The correlation scatter plot’s significance levels reveal significant relationships between “Intention to use” (tac_inten, the factor that directly represents a person’s propensity to accept a given technology) with all other TAC factors (except “Curiosity”), “Quality of Learning” factor (qole) and user age.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Figure 9: Spearman-Rank Correlation Scatter plot matrix for TAC and QoLE factors, and demographic data. Stars indicate levels of significance.



	Training	sex	age	qole	tac_eas	tac_use	tac_inten	tac_img	tac_imm	tac_inter	tac_enj
Training	1***	0.02	-0.02	-0.07	-0.05	-0.04	-0.11	-0.1	-0.1	-0.08	-0.16
sex	0.02	1***	-0.14	-0.08	-0.22**	-0.11	-0.06	0	-0.01	-0.07	0.03
age	-0.02	-0.14*	1***	-0.18	-0.01	-0.04	-0.12	-0.03	0.01	-0.04	-0.02
qole	-0.07	-0.08	-0.18**	1***	0.38***	0.6***	0.52***	0.54***	0.37***	0.34***	0.44***
tac_eas	-0.05	-0.22***	-0.01	0.38***	1***	0.61***	0.54***	0.34***	0.4***	0.45***	0.38***
tac_use	-0.04	-0.11	-0.04	0.6***	0.61***	1***	0.73***	0.64***	0.57***	0.45***	0.58***
tac_inten	-0.11*	-0.06	-0.12*	0.52***	0.54***	0.73***	1***	0.6***	0.52***	0.41***	0.64***
tac_img	-0.1	0	-0.03	0.54***	0.34***	0.64***	0.6***	1***	0.59***	0.44***	0.47***
tac_imm	-0.1	-0.01	0.01	0.37***	0.4***	0.57***	0.52***	0.59***	1***	0.6***	0.59***
tac_inter	-0.08	-0.07	-0.04	0.34***	0.45***	0.45***	0.41***	0.44***	0.6***	1***	0.43***
tac_enj	-0.16**	0.03	-0.02	0.44***	0.38***	0.58***	0.64***	0.47***	0.59***	0.43***	1***

3.5.3.2 Linear Mixed Model

A linear mixed model was fitted with “Intention to use” as outcome variable and all remaining TAC factors, “Quality of Learning” and demographic data as fixed effects. By-subject random intercepts were also included in the model to account for multiple measures for each participant: after VirTra and after VR training. Non-significant parameters in the model were excluded by using a backward-selection approach. Results of the linear mixed modelling show a significant positive impact of the factors “Perceived Usefulness”, “Enjoyment”, “Imagination”, and “Ease of Use”. Significant negative impact on “Intention to use” was found for age. All other factors (sex, experience, training system, “Immersion” (TAC), “Interaction” (TAC)) did not show any significant impact on acceptance and were thus excluded via the backward-selection procedure. Parameter estimates of significant model parameters (fixed effects) are shown in Table 8 and Table 9 below. Additional statistics for the model are reported in. In total, 74% of variance in “Intention to use” can be explained by this model (where 11% via is explained by the random effect).

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Table 8: Parameter estimates for fixed effects in the Linear Mixed Model for "Intention to Use" with other TAC and demographic factors as fixed effects.

Predictors	Estimates	CI	P
(Intercept)	0.98	0.60 – 1.35	<0.001
tac_eas	0.12	0.04 – 0.19	0.002
tac_use	0.35	0.24 – 0.46	<0.001
tac_img	0.18	0.10 – 0.26	<0.001
tac_enj	0.24	0.16 – 0.31	<0.001
age	-0.01	-0.01 – -0.00	0.006

Table 9: Additional statistics for the Linear Mixed Model for "Intention to Use".

Statistic	Value
σ^2 of random effect	0.08
$N_{\text{Participant}}$	178
Observations	297
Marginal R^2 / Conditional R^2	0.634 / 0.741

Answer to Research Question RQ2.2:

Our results reveal significant positive impact of the factors "Perceived Usefulness" (most relevant), "Enjoyment", "Imagination", and "Ease of Use" (less relevant) on participants' acceptance of virtual training technology. Furthermore, there is a negative relationship between user age and acceptance i.e., propensity to accept the virtual training technologies declines slightly with age. In contrast, other factors like sex, experience, training system, "Immersion" (TAC), or "Interaction" (TAC) do not show any significant impact on acceptance.

3.5.4 Research Question RQ3: How do trainees assess the training effect and utility of the technology?

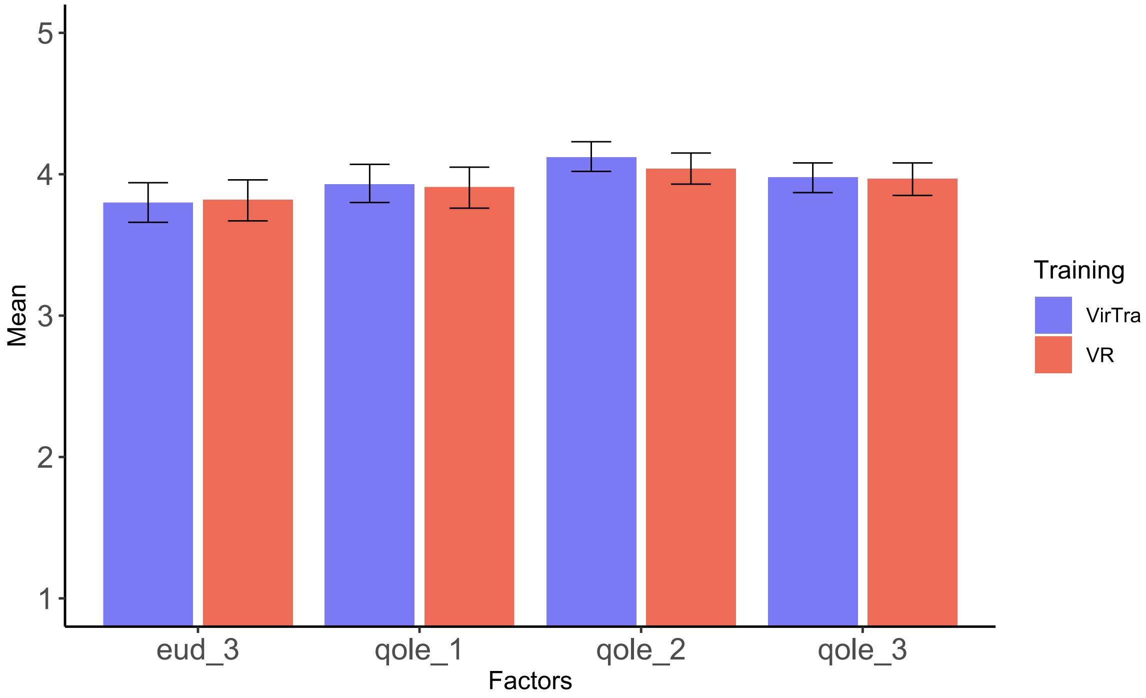
3.5.4.1 Training Effect

Perceived training effect was measured in Group 2 (QoE, n=319) by items in "Quality of Learning" (QoLE) questionnaire and questions of the "Eudaimonic aspects" (Eud)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

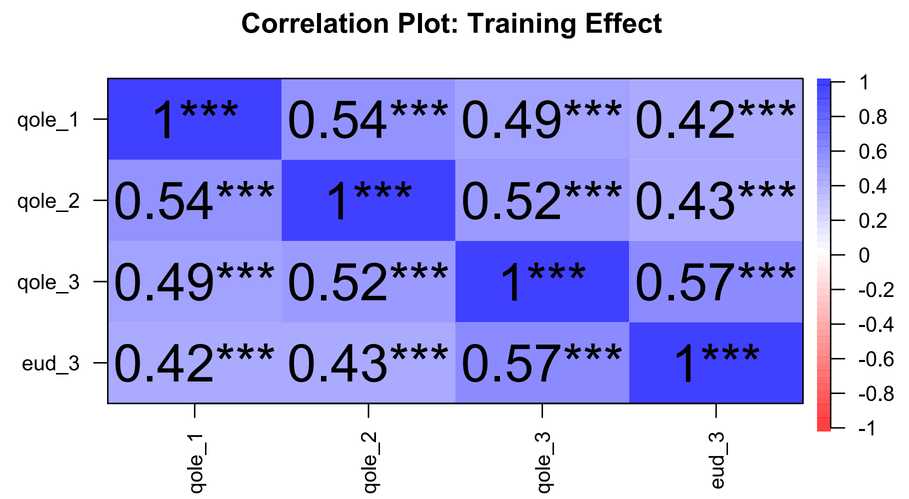
questionnaire ("Training with the system will make me a better police officer."). All items were evaluated fairly high (3.8-4.1 on 5-Likert scale) on average and exhibit relatively high correlation between each other (0.40-0.56). Interestingly, we have found no significant difference between two training systems (VirTra and VR). Figure 10 and Figure 11 show the distribution of the answers as well as factor correlations.

Figure 10: Means of perceived training effect related items (Quality of Learning, Eudaimonic Impact) for each training system with confidence intervals.



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Figure 11: Spearman-Rank Correlation scatter plot matrix of training effect related items. Stars indicate levels of significance.



3.5.4.2 Training Utility

Utility was measured in Group 2 (n=319) and Group 3 (n=302) by items in the Attrakdiff questionnaire (Pragmatic part) and TAC questionnaire (Usefulness part). All items were evaluated fairly high (4.1-4.4 in 5-Likert scale and 4.5-5.7 in 7-Likert scale).

Nonetheless, some differences of the means were found in Attrakdiff questionnaire (pragmatic part). VirTra system is significantly simpler (adiff_p1) than VR system (5.1 vs. 4.6 on 7-Likert scale, p<0.001) and slightly more clearly structured (adiff_p4) (4.7 vs. 4.9 on 7-Likert scale, p=0.056).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Figure 12: Attrakdiff Pragmatic items related to training utility by training system.

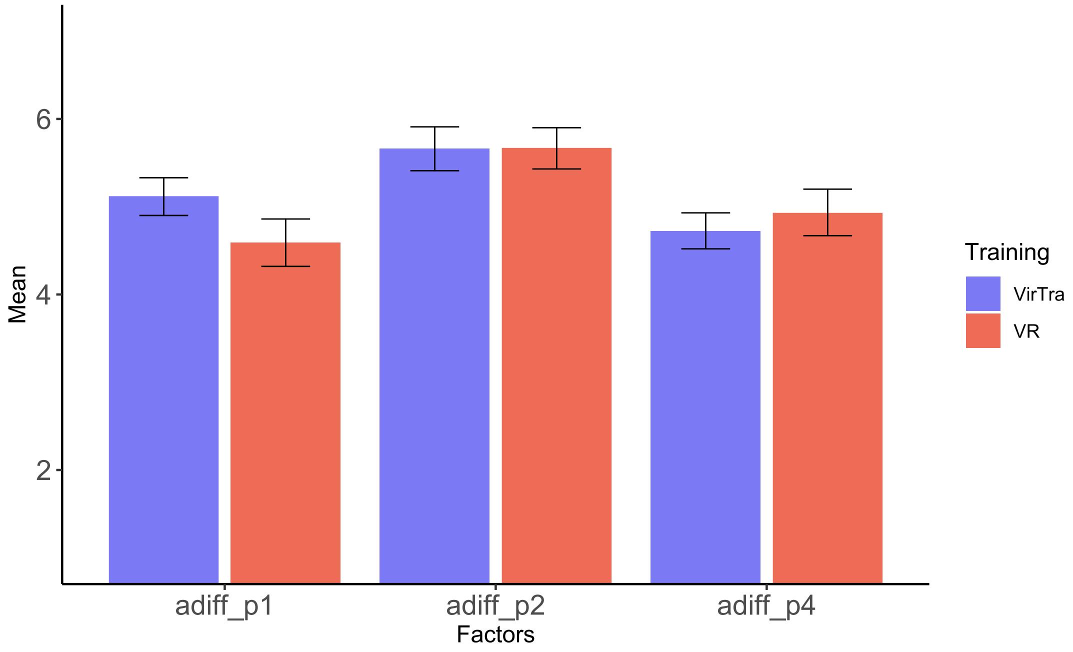
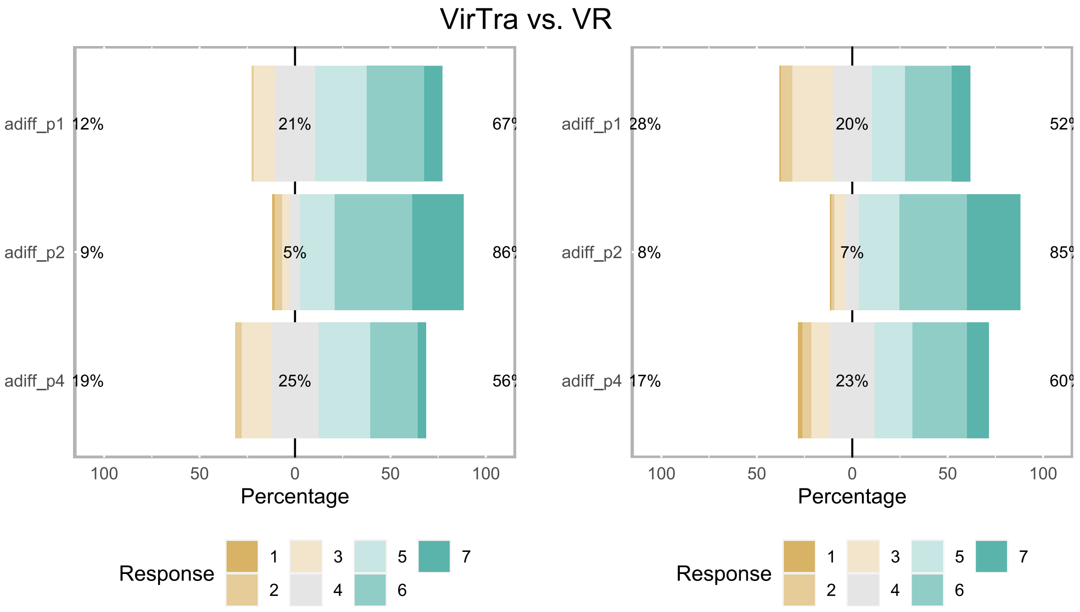


Figure 13: Distribution of Attrakdiff Pragmatic responses related to training utility per item by training system (7-Likert scale).



This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Answer to Research Question RQ3 (quantitative part):

Trainees rated both aspects, Training Effect as well as Training Utility as fairly positive, with scores of the two systems being very close to each other. For further details and explanations, please refer to the qualitative answers to this question in Section 3.6.1.

3.5.5 Research Question RQ5: Which experience dimensions are affected by the presence of a pain stimulus?

To explore the viability and impact of an additional pain stimulus, a sub-group of participants was trained using an artificial pain stimulus (electroshock). The analysis data was analyzed to find factors that were significantly influenced by the presence of pain stimulus. We have found that "Stress" (vas_str) and "Mental effort" (vas_rsm) factors received significantly higher scores in conditions with pain stimulus comparing to the absence of pain stimulus. "Stress" with pain stimulus was assessed on a 100-point Visual Analogue Scale (VAS) with the mean score 63.17 comparing to 59.15 without (t-test p=0.002, Wilcoxon test p=0.002). "Mental effort" with pain stimulus was assessed on a 150-point VAS with the mean score 71.96 comparing to 67.59 without (t-test p=0.009, Wilcoxon test p=0.026), see Figure 14 and Figure 15 below.

Figure 14: Violin and box plots with distribution of "Stress" score between groups depending on pain stimulus presence. Colored points indicate distribution of responses. Black dots indicate arithmetic means of scores.

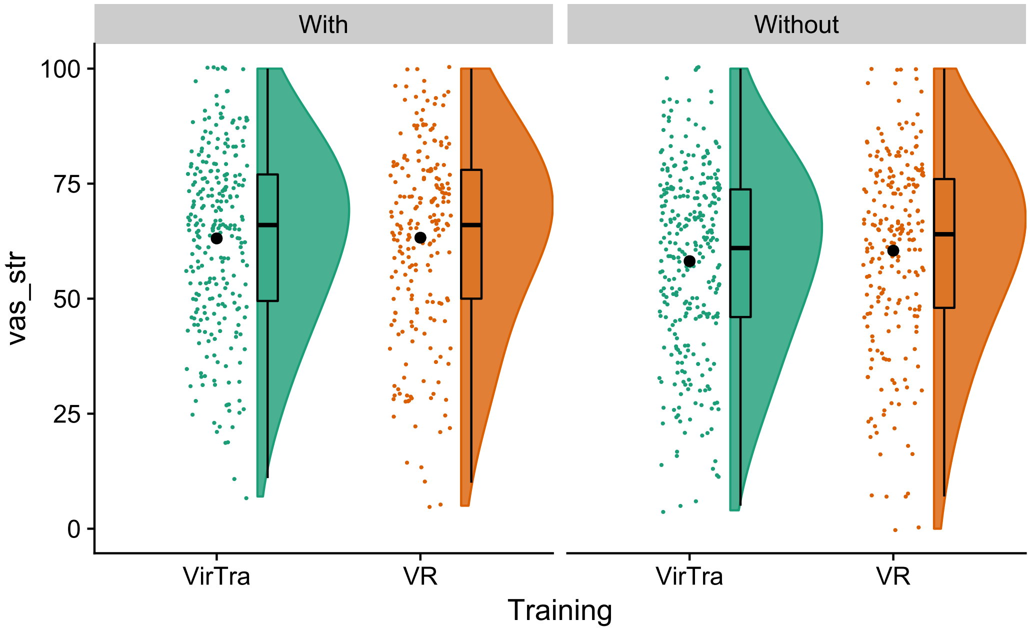
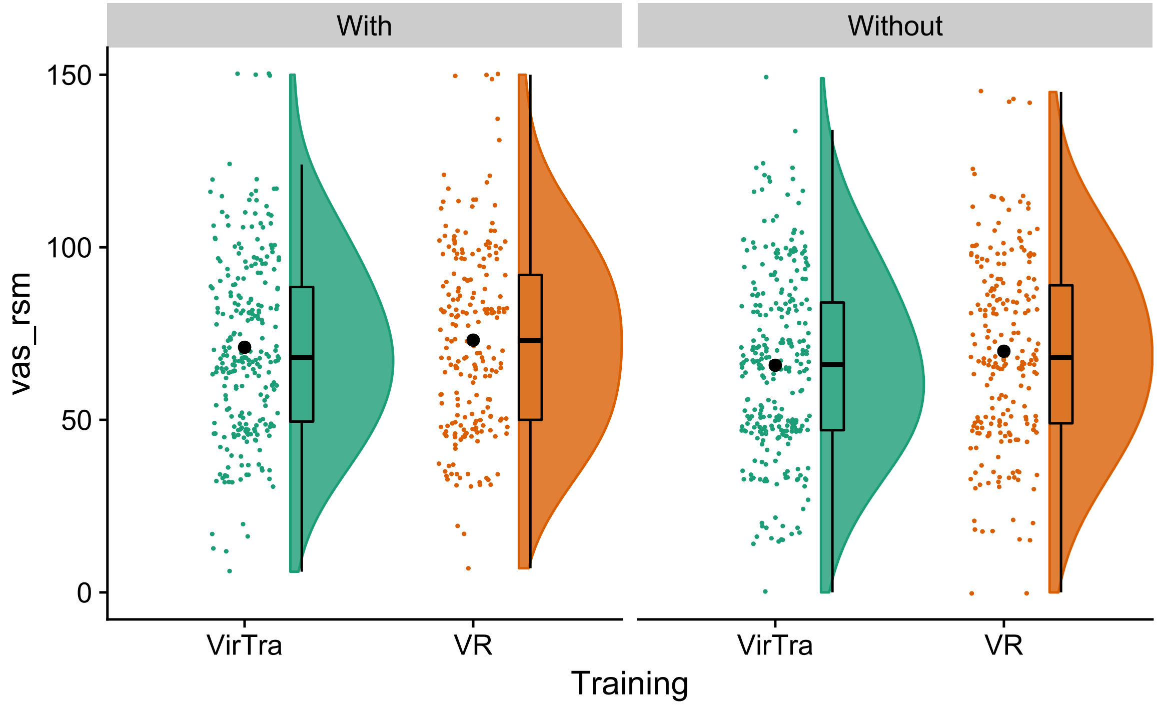


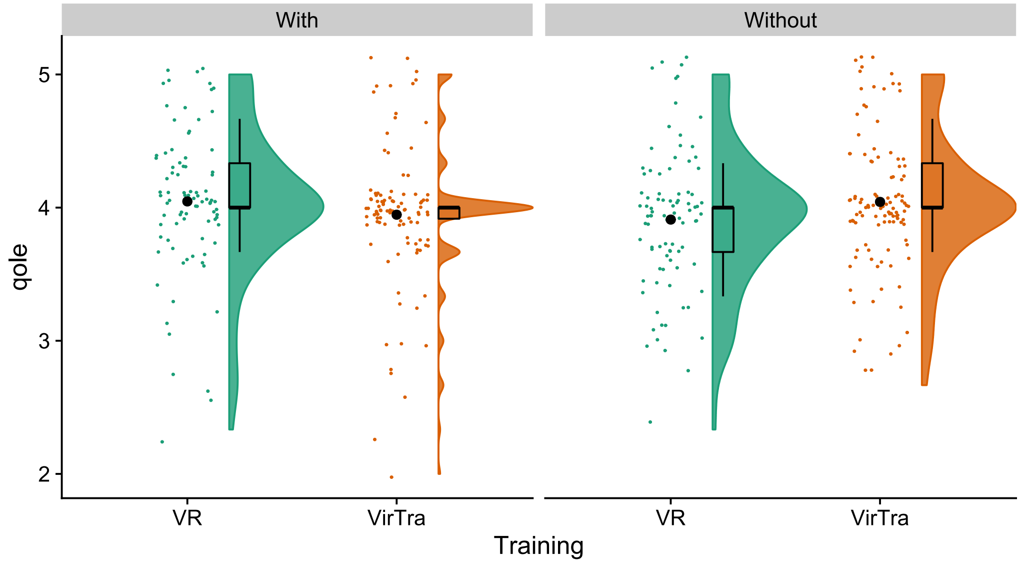
Figure 15: Violin and box plots with distribution of "Mental effort" score between groups depending on pain stimulus presence. Colored points indicate distribution of responses. Black dots indicate arithmetic means of scores.



Additionally, we compared score means of other factors than stress and mental between training systems. However, only the VR system in Group 1 (SOPI), "Quality of Learning" (QoLE) was assessed significantly higher *with* pain stimulus comparing to the absence of pain stimulus – 4.04 vs 3.90 on 5-Likert scale (ART-test, p=0.042), which is a small yet statistically significant difference.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Figure 16: Violin and box plots with distribution of "Quality of Learning" score between training systems depending on pain stimulus presence (Group 1 SOPI). Colored points indicate distribution of responses. Black dots indicate arithmetic means.



<u>Answer to Research Question RQ5:</u>

The presence of a pain stimulus only affects reported levels of stress and mental effort. All other factors (acceptance, presence, etc.) were not affected (except for QoLE in the case of VR system in Group 1). This is an important finding, since according to the conceptual DMA model (see D3.2) stress and mental effort represent essential components in training action (and salient motor heuristics and embodied choices). Being able to increase these components' levels while leaving other experience factors intact (including acceptance and presence) confirms the viability of artificial pain stimuli in virtual police training.

3.6 Qualitative Results

In the following, results from the qualitative interviews are summarized from the perspective of trainees (referring to RQ3) and trainers (referring RQ4). The section describing results for the trainees is structured as follows: First, results the VR system (Refense) are discussed. Second, we report findings regarding the VirTra system. Third, results from the interviews are interpreted in the context of the quantitative findings to explore possible explanations of quantitative differences and indications for causal relationships. For trainers and trainees, findings are reported by training system (VR, VirTra).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

3.6.1 Trainee Perspective (RQ3)

3.6.1.1 Experience with VR

Overall feedback on the VR training was positive while it was mentioned that the Refense system in its current state can only be used to train certain characteristics of police operations (for example tactical training). This relates to the fact that many details of real world scenarios were not yet implemented in the VR training system at the time of the training campaign.

There were individual differences regarding the appraisal of general **realism** of the training. Many trainees reported a **strong feeling of immersion inside the VR environment**. However, some participants felt like being in a video game which might reduce the positive impact of the training. The perception of the VR environment is likely related to certain factors that might limit the realism and perceived added value of the training from the perspective of the participants. Based on the qualitative interviews, realism of VR was reduced mainly due to unrealistic weapon handling, audio limitations and interaction with objects and team members in the VR environment. These three sources of issues are discussed in the following.

Weapon handling is not very realistic according to participants and should be improved. The physical weapon model used for the Refense system is quite different compared to the MP5 the participants use in real scenarios (e.g., regarding shape, weight, reloading). It would be preferable to have a weapon model that resembles the original MP5 as closely as possible. Trainees also asked for the possibility to switch to their regular service weapon (hand gun) and use it as well.

The **audio experience** within the VR environment was also critically discussed by trainees. Trainees could not localize the source of a particular sound due to the fact that spatial audio output was not implemented in the Refense system at that time. Participants mentioned that many of the different types of sound effects including voice output from Non-Playable Characters (NPCs), shooting sounds and communication with teammates (especially when being in with each other’s proximity) which participants require to be spatialized. Participants mentioned that in real-life police operations, continuous localization, separation and interpretation of sounds are integral part of their decision-making and acting. Thus, any limitations of spatial audio rendering reduce the perceived realism and effectiveness of police force training and thus clearly represent a point for further improvement of the system.

Moreover, interaction with objects, NPCs and team members needs to be improved. For example, NPCs can’t react to questions from the trainees which would be an important source of information in a real scenario. Thus, a more intelligent, reactive behavior of the NPCs should be implemented, for example, NPCs respond to questions and follow instructions from the

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

officers (note: trainers can partly control the behavior of the NPCs, however a more nuanced control or automatized behavior is not provided, so far). Furthermore, some problems became apparent when interacting with virtual objects. These problems might be associated with participants' little prior experience with VR environments. For example, one participant attempted to lean against a "virtual wall" and subsequently fell to the floor. Interaction with team members is sometimes difficult because the character models all look the same and can only be distinguished by arbitrarily assigned numbers. Also, the physical distance between trainees appears to be slightly distorted in the VR environment and hand signs are not possible, thus hindering non-verbal communication between team members. In addition, trainees pointed out that they did not recognize when a colleague got shot inside the simulation: "I did not recognize that one of my colleagues got shot, I would not have missed this in real life." The reason is that shot team members in the current version of the system just disappear from the simulation without any interaction and the team simply continues its mission. This is obviously not realistic according to participants and should be improved.

Some trainees pointed out that their (subjective) **stress levels during VR training were rather low** compared to other types of trainings (e.g., the VirTra system). This might be related to the general perception of the VR training as a videogame as mentioned before which in turn might be connected to current shortcomings of the VR training (weapon handling, audio experience, interaction with the environment).

Moreover, some issues regarding the **accessibility** of the system were highlighted in the interviews. Firstly, these issues relate to the lack of familiarity with VR systems so that some participants needed a lot more time to get used to the VR equipment. In this context, it is important to note that all participants trained in the VR system for the first time while most participants already had experience with the VirTra system. This lack of familiarity might also cause reduced acceptance for the VR system. Secondly, a few interviewed participants (3 out of 22) reported physical symptoms related to VR, such as dizziness or headaches.

To summarize, the aforementioned shortcomings currently limit the realism of the VR training and should be further improved. However, trainees also reported on specific aspects that they specifically liked about the VR system (specific advantage of the VR system over alternatives): One key advantage relates to the scenario **content**: Trainees liked the possibility to train a scenario that resembles a real pace in Zurich. They pointed out that in other types of trainings (non-VR) this would not possible since these trainings usually take place in dedicated remote training areas. Another important aspect that was mentioned by the trainees was the **learning effect** associated with the Refense system. The after-action review which is conducted directly after the training session was perceived as one of the most beneficial features of the training

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

system by many participants. Being able see one's individual line of sight and position was considered highly useful especially for training of tactical aspects. Moreover, one's own behavior in a given situation is tracible allowing trainees to better comprehend possible mistakes. The perceived immersion was considered as a unique feature of the VR system that can provide great opportunities for virtual training. Accordingly, participants recognized the **high potential** of the VR system especially for future applications. As trainees also pointed out, this potential can be realized by improving the design and implementation (e.g., in terms audio rendering, weapon handling and interaction) of the VR system used.

3.6.1.2 Experience with VirTra

General feedback on the VirTra system was also positive. A lot of trainees praised the **realism** of the training. One trainee stated: "This is the most realistic training I have ever done". This perceived realism is related to specific factors (e.g., weapon handling and sound). The **weapon handling** was rated as very authentic as the weapon used in the VirTra system is very similar to the actual service weapon of the officers. Thus, handling and aiming was described as very convenient and "natural". Regarding **sound**, trainees said that the VirTra system offers a realistic volume level and sound output can be better localized compared to the Refense system.

The **communication** between team members was assessed as positive as it is very similar to a real police operation. This includes both verbal and non-verbal communication due to physical proximity and non-verbal cues. Thus, the VirTra system is useful for training team work (in a team of two). Specifically, it would be nice to do the VirTra training with colleagues you usually work with as one trainee mentioned.

Moreover, participants pointed out that the quickly changing surroundings and distractions on the 270° canvas create a **challenging, stressful training situation.** The relatively high stress levels are a product of two factors according to participants: First, the situations themselves presented via VirTra, and secondly, the spatial orientation that is specific to the system (e.g., sudden changes in perspective on the 270° canvas). While the spatial orientation was described as "awkward" at times by some participants, the high stress levels inside the VirTra system were generally perceived as beneficial (in terms of a learning effect).

These high stress levels are also associated with the combination of **realistic scenarios** with a shooting exercise (with moving targets). These characteristics support a **strong learning effect**: Trainees emphasized that the VirTra training has improved the awareness for certain scenarios. However, the sudden changes in the environment can sometimes be confusing and trainees often need to reorient themselves. This could be further improved. Moreover,

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

scenarios are based on a US-American cultural context, so trainees wished for scenarios taking place in Switzerland. It was also mentioned that the after-action review offer by the VirTra system is worse compared to the Refense VR system as it provides less features and details about the respective training trials.

Obviously, VirTra is also associated with certain **limitations** that were also discussed with the participants. One limitation is due to the restricted movement space. Thus, while it may not be practically feasible, trainees suggested to provide more space to walk. In general, participants pointed out that one can only practice certain aforementioned aspects within the current VirTra system. However, for the training those aspects, the system is very good.

3.6.1.3 Qualitative results in context of quantitative results

The interviews shed light on apparent contradictions in the quantitative findings (e.g., quantitatively higher feelings of sense of presence for VR while ratings for acceptance-related items are tendentially higher for VirTra). The qualitative interviews help to interpret and solve such contradictions as distinct advantages for each system were qualitatively identified that can be interpreted sensibly in line with the quantitative differences.

For example, the significantly higher ratings regarding sense of presence for the VR system are in line with the emphasis of immersion by trainees in the interviews. These experiences (such as “feeling involved in an environment” or “feeling that all senses were stimulated at the same time”) might be more meaningful and tangible in a 3D environment (VR system) compared to a 2D environment (VirTra system). However, higher levels of sense of presence might not automatically imply a “better” or satisfying experience overall. While trainees might have perceived a higher feeling of immersion in the three-dimensional VR environment, they might value certain advantages of the VirTra system more resulting in a higher rating for the VirTra systems in terms of overall experience. In this context, characteristics such as weapon handling, sound output and levels of stress come to mind which were positively mentioned in the interviews. These characteristics are not directly related to sense of presence, which helps to explain the rating differences between the two systems that vary according to the experience dimension assessed.

It should also be mentioned that regarding many aspects, the quantitative measurement differences between systems are statistically significant, yet not of high magnitude. This can be explained by the fact that the VR (Refense) system used holds specific advantages/disadvantages over the VirTra system (e.g., real 3D environment, movement space) which cause positive and negative rating differences between the two systems, depending on the experience dimension/aspect being assessed. Still, the VR system offers a

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

higher potential for future virtual trainings as it is more easily extensible compared to the VirTra system. In conclusion, to build on the great foundation of the tested VR system, its shortcomings (e.g., weapon handling, sound, interaction) should be improved to provide a more effective and authentic training experience.

3.6.2 Trainer Perspective (RQ4)

3.6.2.1 Experience with VR

In general, trainers recognized several **distinct advantages** and a **high potential** of the Refense system. From an educational standpoint, they assessed many characteristics of the training as positive and helpful. They also felt that most trainees experience a strong feeling of **immersion** in the VR environment. Regarding graphics and visuals, trainers mentioned that graphics do not need to be "hyper-realistic" to provide an authentic training experience, however improvement would be a nice-to-have. However, they mentioned that some participants tend to perceive the VR training as a videogame which obviously reduces the learning effect. Importantly, they pointed out that the system is still in an early stage and there is a lot of room for improvement. In this context, several characteristics of the Refense VR system were considered that could be further improved.

In general, feedback from trainers was similar to the impressions described by the trainees in many ways. These congruent themes relate for example to weapon handling, audio experience, interaction with objects the learning effect for the after action. A full description of the trainers' perspectives can be found in the interview transcripts. Additional aspects that were discussed specifically with trainers (e.g., usability of the training systems, suitability of systems for different aspects of training, use of gamification elements) are described in the following.

Regarding **maintenance and usability** of the system, trainers liked that little effort for preparation is required compared to other types of trainings since fewer logistical factors need to be considered. These efforts could yet be minimized as trainers are still depended on external staff to conduct trainings (contrary to the VirTra system). Furthermore, trainers pointed out that the Refense system is also rather easy to use. For example, it is quite easy to vary a given scenario. Suggestions for improvement regarding the system usability were discussed, as well. First, although it is quite easy to vary scenarios, trainers wished for faster editing and different default options for scenarios (e.g., per default there are too many characters in the scenario that have to be removed by the trainers which takes some time). Second, an option to save changed scenarios and reload them should be implemented. Third, there is currently no push-to-talk functionality possible for trainers when they operate the

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

radio (speaking as the perpetrator): The operator must turn on the radio, talk and then press the button once again to switch off the radio communication. This can lead to errors, as trainers pointed out. Fourth, the control of NPCs is not very convenient: The walking routes of NPCs cannot be set with one single click, instead trainers must specify the path via multiple subsequent clicks. Fifth, one trainer mentioned that "shooting" a trainee inside the VR environment could be further simplified since it requires coordination between the trainer and the Refense staff. These refinements can help to avoid unnecessary multi-tasking and cognitive load for the trainers. One trainer who was interviewed twice (when first using the Refense system as a trainer and four weeks after) mentioned that while the Refense system is easy to use, increasing experience with the system over time definitely helps to create more realistic, challenging scenarios as a trainer (e.g., the performance of the trainer as the perpetrator is crucial for an educative training experience). Lastly, trainers mentioned general technical issues (e.g., system breakdowns) which should obviously be minimized as they might mess up training schedule and decrease the motivation of trainees.

Moreover, trainers were explicitly asked about **suitability of the Refense system for different aspects of training**. These categories were derived from D3.1 and include: Tactical training, personal safety, shooting and weapon handling, fitness training, combat training, law and regulations training, communications training, perception and action, situation training, and psychological training. Trainers saw the highest potential in the training of tactical aspects, law and regulations, communication, situational and psychological competency training. Opinions were divergent for the training of personal safety (as this type of training comprises a lot of different facets) and perception and action (as subtleties in behavior / in the environment might be hard to implement realistically). Finally, trainers pointed out that the Refense system is not particularly well suited for the training of fitness, combat, and weapon handling. In conclusion, the Refense system offers educational value and high potential for specific types of training and hence constitutes a great supplementation for trainings in the real world according to the trainers. One should be aware, however, that there are also areas that should be practiced using other training approaches. In this context, one trainer emphasized that the VR training generally should not be used "too early" (i.e., not in basic training for new trainees). Instead, newer participants should practice in real scenarios, first.

While the **training of communication outside of the team** works well (the trainer can speak as the perpetrator), trainers also proposed improvements in this context. It was suggested to add the possibility for the trainer / supporting staff to enter the scenario as an avatar (freely moving around in the VR environment) supporting a more authentic interaction between perpetrator and officers. Trainers also criticized that the voice of the perpetrator is always distorted (resembling a "computer voice") which reduces the realism of communication. One

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

trainer also argued that the addition of facial expressions and gestures (more expressive body language) would be helpful to support a more natural interaction.

Most trainers expressed critical opinions regarding **gamification elements**. Most of them were skeptical about the implementation of positive rewards like a point system, badges achievements, or competition. While the use of these elements generally might make sense from a didactic standpoint (i.e., higher motivation to improve performance), it bears the risk of the VR training being perceived as just a videogame. In general, these elements might distract trainees from the actual scenario so that trainees rather focus on receiving certain rewards. One trainer pointed out that these kinds of positive rewards do not reflect incentives in a real operation. However, trainers do not necessarily disqualify gamification for the VR training: Instead, they describe the use of such elements as a difficult balancing act that requires great caution (i.e., gamification elements should be used in an unobtrusive way not distracting participants and resembling incentives that reflect characteristics of a real operation). In the case of pain stimuli, trainers emphasized the relevance of this specific tool. By increasing physical arousal of the trainees, it creates a more realistic and tense training situation and can help to prevent participants perceiving the VR training as a game. Thus, it can support a stronger learning effect. Pain stimuli are also closer to the real situation than most positive rewards since you don’t want to collect points in the real situation but rather avoid physical damage. Importantly, these pain stimuli should not be used mindlessly: They should only be applied in cases of obvious misbehavior.

Another topic that was explicitly discussed with the trainers revolved around ideas for the **measurement of training progress and effectiveness**. In this context, trainers expressed the idea of creating digital equivalents of existing (real) training environments. This would make it possible to compare different kinds of trainings (e.g., VR training, training in real scenario, no training at all) and assess the learning transfer. Another idea proposed a combination of performance self-assessments and assessments by the trainer (e.g., using a standardized questionnaire). By comparing self-assessment and trainer assessments, deviations become apparent and trainees might be able to rate their own performance more accurately over time. However, trainers mentioned that it would be hard to define categories for assessment that cover all important aspects. Moreover, trainers were indecisive about rating the team as whole or individual members, only.

3.6.2.2 Experience with VirTra

Although interviews with trainers put an emphasis on the VR system, some feedback on the VirTra was collected during the interviews, as well. Feedback from trainers was similar to feedback provided by trainees (e.g., realism of weapon handling, request for scenarios based in Switzerland). Additional aspects specifically mentioned by the trainers are described in the following.

As for the Refense system, trainers liked that less effort is needed concerning the preparation of a training. They also gave positive feedback on the ability to control the scenarios. This control allows to replicate a certain situation and provoke specific behaviors / errors from the trainees. However, from the trainers' perspective, the **usability** of the VirTra system is not very good as one "wrong" click might already result in an error or breakdown of the system.

Moreover, the VirTra is well suited to provide **feedback** (while not as good as the VR system as trainers pointed out). For example, you can show trainees in which formations they were standing / point out errors via video recordings which is hardly possible for trainings in the real world.

Finally, the **suitability for several aspects of training** was discussed with the trainers. VirTra can be used to practice certain characteristics of weapon handling (as the weapon within the VirTra system closely resembles the officers' actual service weapon). Furthermore, VirTra is suited for situational training (specifically perception and action, i.e., observing suspicious and reacting appropriately) and - potentially - training of law and regulations. Yet, contrary to the VR system, training for communication outside of the team (e.g., appropriate communication with the offender) is hardy feasible in the VirTra system.

3.7 Conclusion

The ZüriVR study provided a broad range of qualitative and quantitative results. Overall, both virtual training systems (VirTra and VR/Refense) were highly positively received by trainers and trainees. In terms of differences, VirTra received higher acceptance, enjoyment and ease of use ratings, while the VR (Refense) system resulted in higher immersion and presence.

The study results revealed the strengths and weaknesses of the (current versions of the) two systems and the underlying technologies used. The VirTra system with its 2D cinema favors joint decision-making and realistic weapon handling. In contrast, the Refense VR system with its 3D HMDs currently favors procedural learning and tactical training. However, with careful design and feature additions (e.g. spatial audio), the VR system's training experience and application range can be significantly increased.

4 VR Training Experience Framework and Model

This section summarizes the results and learnings from the WP4 activities related to the development of a VR experience measurement framework and model for decision and acting.

4.1 VR Training Experience Measurement Framework

The **quantitative measurements** conducted in the ZüriVR study provided clear evidence on the fact, that in the context VR training for decision-making and acting, a number of experience dimensions are relevant and thus should be addressed a future evaluation setup. The instruments that have turned out to be most useful ones in terms of insight and information value, are: SOPI (presence & immersion), TAC (technology acceptance), VAS (stress, mental effort), and QoLE (Quality of Learning Experience). These inventories can be also considered as fairly orthogonal to each other in terms of correlation and coverage of aspects (RQ1). In this respect, as the modeling in Section 4.2 also shows, it is important to mention that Quality of Learning Experience (QoLE) should be treated as construct on its own, since it can only partially be predicted by the other inventories (like SOPI or acceptance). Thus, direct measurement using the QoLE questionnaire (see Annex A) is recommended. Furthermore, it is important to notice that the SOPI inventory only addresses the levels of presence and immersion provided by a system. Thus, high SOPI scores do not automatically imply that a system is better suited or more effective for DMA training purposes. The technology acceptance (TAC) questionnaire addresses critical aspects related to acceptance of a virtual training technology very well and showed high discriminatory power (RQ2). Furthermore, the two visual analogue scales (stress, mental exertion), albeit they do not represent experience dimensions per se, provide relevant information regarding (intended) mental and emotional impact of the experienced training setting.

These conclusions do not automatically imply that all other measurement instruments used in our tests do not work or lack value. According to our factor analysis, our questionnaires on eudaimonic aspects (eud) as well as VR quality of experience (VRQoE) function very well in that the different items map to the intended factors. And taken together, both questionnaires are able to predict QoLE as good as technology acceptance (see Section 4.2.2) – which can be also explained by the fact that eudaimonic aspects of self-actualization and personal growth are strongly intertwined with perceived training effectiveness. Still, we regard the information value of SOPI and TAC as higher when it comes to a holistic evaluation of a VR training experience, due to their more complementary nature and discriminative power. However, we found that in our case, the AttrakDiff semantic differentials did not provide conclusive results and AttrakDiff was also identified as problematic by the factor analysis. This might be rooted

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

in the fact, that only the short version of the instrument was used in the ZüriVR Study, and thus the long version might be trialed in a future study.

Surprisingly, we found little influence of user-related variables (like age, gender, technology curiosity, etc.), with few exceptions like age (→ QoLE) and gender (→ negative effects in the context of SOPI, which have been also reported in prior work). This might be explained by (unproven) thesis that participants' background (police officers) and the given context (police training) might mitigate the impact of user diversity on the ratings issued.

Recommendation: assuming a general evaluation of VR training experience without additional special requirements, we recommend using a combination of TAC, SOPI, QoLE and VAS. In the case of restrictions (e.g. SOPI licensing, survey maximum time), SOPI can be replaced by the (shorter) VRQoE questionnaire, if still the full TAC questionnaire is used. In case that using SOPI and TAC together results in too many items to be answered, reducing the TAC questionnaire to the three top factors (ease of use, usefulness, intention to use) or even intention to use only, represents an alternative, albeit at the expense of diagnostic power. Furthermore, although it is generally recommendable to inquire user demographics and background (like age, gender, experience), one should not expect pronounced impact of these variables on experience ratings in the context of police officer training.

The **qualitative interviews** conducted in the ZüriVR study were found to be highly valuable in addition to the quantitative measurements. The added value of the qualitative interviews relates to two themes: First, they provide deep insight into *why* training systems are perceived as useful. Second, they are a suited to identify barriers and shortcomings of current trainings. In this way, they encourage trainees and trainers to reflect on suggestions for improvement and future directions of development. This reflection from a user-centered perspective is crucial to create training systems that meet the requirements of its users as closely as possible.

Recommendation: the use of qualitative interviews as applied during the ZüriVR study is highly recommended for future research in the context of police training evaluation. Even if collected only from a subset of participants, the interview results provided a plethora of findings and insights that inform requirements elicitation and prioritization, identification of problem points and options for improvement of virtual police training technology.

However, beyond interviews, alternative qualitative data collection methods can be considered as well, particularly when interviews might not be feasible. These methods include the use of open questions inside questionnaires or stronger emphasis of methods relying on group interaction (e.g., in the form of focus groups or workshops). They can provide certain advantages that might be useful given specific conditions. For example, open questions inside a questionnaire can be used to collect qualitative feedback from a larger number of people with less effort compared to qualitative interviews. Drawbacks associated with this approach are that answers from participants are likely to be less detailed and there is no possibility for the researcher to ask about additional details based on the given answers. Focus groups and workshops on the other hand might stimulate exchange between participants that reveals further insights into the experience with the training systems but are associated with greater effort in their execution.

Table 10: Recommended VR training experience measurement framework. See Appendix A for further details on the different instruments suggested.

Instrument Type	**Recommended Instruments**
Quantitative Instruments	Quality of Learning Experience (QoLE) Sense of Presence (SOPI) Technology Acceptance (TAC) Visual Analogue Scales (VAS)
Qualitative Instruments	Interviews with a subset of participants / stakeholders

4.2 Training Experience Model

This section outlines the result and process of the WP4 modeling activities. Model development took place in two iterations. Initially, it was planned to develop a structural equation model (SEM) for VR training experience (Iteration 1). However, due to insufficient model convergence already on the level technology acceptance modeling as well as due the partitioned nature of the dataset, the modeling approach was changed to linear mixed modeling (Iteration 2).

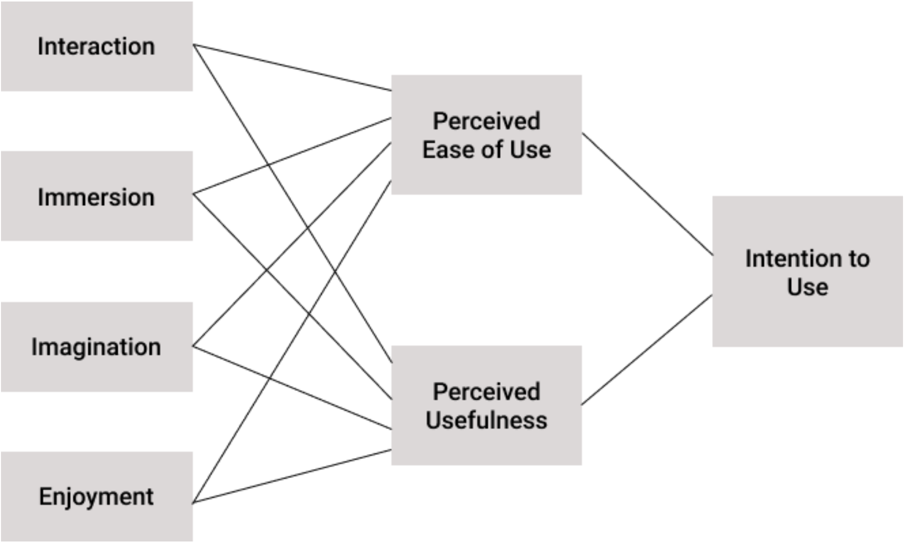
4.2.1 Iteration 1: Structural Equation Model

Our initial model approach is best exemplified by our work on acceptance modeling. According to Vekantesh & Bala (2008), a conceptual model of users' attitudes toward technology (Intention to Use) is nested. The multilevel TAC technology acceptance model is represented in Figure 17 (adapted from the work of Huang et al. on a VR-based training system). We performed Confirmatory Factor Analysis to prove the plausibility of applying Structural Equation Modeling (SEM) in the next step. However, the data shows that the model fits not

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

that good: Confirmatory Factor Index (0.86) and Tucker Lewis Index (0.84) do not reach the recommended threshold of 0.9, and RMSEA (0.085, 90% CI [0.079; 0.092]) is higher than recommended. Thus, applying Structural Equation Modeling with an even deeper nested hierarchy of factors (as would be required for an experience model using SOPI, TAC, QoE factors) was not a recommendable approach in the light of the given data.

Figure 17: Technology Acceptance Model (adapted from Huang et al., 2016).



4.2.2 Iteration 2: Linear Mixed Model

As alternative to the structural equation model, we performed linear mixed modeling of the training experience, again piloting the approach with the technology acceptance results data first. A linear model assumes a linear combination of different influence factors (here: factors underlying acceptance) on the target variable (here: intention to use). The Linear Mixed Model (LMM) is an extension of the traditional linear model that includes both fixed and random effects. LMMs can be useful when there is non-independence in the data. In our case the data was not independent on the participant level (most of participants filled in questionnaires twice – for VirTra and for VR training). Instead of analyzing data independently (which assumes two individual linear models for VirTra and VR) we combine it into linear mixed model by implementing random effect variable ("Participant"). This approach avoids some of the noise of linear regression and take advantage of all the data (comparing to simple aggregation on participant/subject level).

In the Section 3.5.3.2, mixed modeling was successfully performed to analyze technology acceptance (target "Intention to use"), with all underlying TAC factors and demographic data as fixed effects, and with subject ("Participant") as random effect. It showed positive impact of the factors "Perceived Usefulness", "Enjoyment", "Imagination", and "Ease of Use" and negative impact of age. This model explains 74% of variance (thereof 11% via random subject effects, see Section 3.5.3.2 for details).

Given these results, we decided to generally use linear mixed modeling not only to model acceptance but also to model the whole training experience, targeting the central construct "Quality of Learning Experience" (QoLE). The ZüriVR study data is partitioned in three different groups (SOPI, QoE, TAC), but it includes Quality of Learning (QoLE) results in all of them. Consequently, we have performed LMM for these three groups. QoLE is the target variable and all other factors from the questionnaires in the group are fixed effects with "Participant" being a random effect (i.e. a source of random variation in the observed data).

As Table 11 shows the final model for **Group 1 (SOPI)** that contains significant parameters only (after backward elimination): even though most SOPI factors (besides "Spatial Presence") are significant, the random effect influence is very high (34% out of 49% of variance is explained by subject random effect, only 15% by fixed effects i.e., the three SOPI factors). Fairly wide confidence intervals (CI) in predictors also indicate a certain level of uncertainty regarding the exact influence of the different predictor variables on QoLE. This matches the results of the comparison of score means in section 3.5.2: even though participants in SOPI questionnaire assessed the VR system significantly higher, we could not find significant QoLE preference (higher QoLE) in this group due to score variances. According to the model, training type and user parameters (e.g., sex, age, working experience and user's computer background, i.e., sopi_bg) do not show significant effect on QoLE.

Table 11: Linear Mixed Model for "Quality of Learning Experience" (QoLE) with SOPI and demographic factors as fixed effects and participant as a random effect (Group 1). Significant p values are marked as bold.

Predictors	Estimates	CI	p
(Intercept)	3.09	2.68 – 3.51	**<0.001**
sopi_eng	0.22	0.10 – 0.35	**0.001**
sopi_ecv	0.12	0.01 – 0.22	**0.036**
sopi_nef	-0.18	-0.27 – -0.09	**<0.001**
Random Effects			
σ^2	0.14		

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

τ_{00} Participant	0.09
ICC	0.40
N Participant	169
Observations	332
Marginal R^2 / Conditional R^2	0.150 / 0.491

In **Group 2 (QoE)**, the LMM shows that Attrakdiff does not significantly influence QoLE, in contrast to “Quality of Experience” (vrqoe) and Eudaimonic aspects (eud). This model explains 56% of variance (~13% by subject random effect). Like in Group 1, training system used and demographic variables are not significant for predicting QoLE. Detailed output of the final model with significant parameters (after backward elimination) is shown in Table 12 below.

Table 12: Linear Mixed Model for “Quality of Learning Experience” with VRQoE, Eudaimonic and demographic factors as fixed effects and participant as a random effect (Group 2).

Predictors	**Estimates**	**CI**	**p**
(Intercept)	1.35	0.99 – 1.71	**<0.001**
vrqoe	0.15	0.05 – 0.24	**0.002**
eud	0.52	0.43 – 0.62	**<0.001**
Random Effects			
σ^2	0.12		
τ_{00} Participant	0.04		
ICC	0.23		
N Participant	192		
Observations	305		
Marginal R^2 / Conditional R^2	0.429 / 0.562		

For **Group 3 (TAC)**, the final model explains 68% of variance (thereof 26% explained by subject random effects). Interestingly, in the Technology Acceptance group we see significant negative influence (*p* = .006) of age factor on QoLE with estimate -0.01, which means that for each additional 10 years of participant age, QoLE scores are reduced by 0.1 on average. More details of the final model with significant parameters (after backward elimination) are shown in Table 13 below. It is important to notice that in contradiction to the modeling results shown in Table 13, the influence of the main acceptance variable "intention to use" (tac_inten) on QoLE actually is significant. The reason is its collinearity with other TAC factors and resulting masking effects in the modeling process. To prove this, we performed LMM without "Perceived Usefulness" (tac_use) and "Perceived Ease of Use" (tac_eas) and obtained a model with a significant (*p* = .021) "Intention to Use" factor with estimate ~0.15) with a slightly lower Marginal R^2 of 0.373 (Conditional R^2: 0.666).

Table 13: Linear Mixed Model for "Quality of Learning Experience" with TAC and demographic factors as fixed effects and participant as a random effect (Group 3). Note, that tac_inten is not significant due to collinearity with other factors.

Predictors	**Estimates**	**CI**	**p**
(Intercept)	1.41	0.92 – 1.90	**<0.001**
tac_use	0.33	0.20 – 0.46	**<0.001**
tac_inten	0.00	-0.14 – 0.14	0.964
tac_img	0.25	0.14 – 0.35	**<0.001**
tac_enj	0.12	0.02 – 0.21	**0.016**
age	-0.01	-0.02 – -0.00	**0.006**
Random Effects			
σ^2	0.10		
τ_{00} Participant	0.08		
ICC	0.45		
N Participant	178		
Observations	297		
Marginal R^2 / Conditional R^2	0.418 / 0.679		

4.2.3 Modeling Conclusion

Table 14 below shows the resulting framework of linear models that can be used for predicting the central concept of Quality of Training Experience (QoLE) on behalf of the different factors studied in this deliverable. (For detailed model coefficients, please consult the different tables in the previous subsections.) In this context, it is important to notice, that according to our results, QoLE represents an experience dimension on its own that can only be partially predicted on behalf of the experience dimensions surveyed and analyzed. And, the different inventories that test for the dimensions and factors provide considerable information value and insights, not only prediction of QoLE. For example, the SOPI result clearly show how the two systems tested differ in conveyed sense of presence, independent of the fact that SOPI factors can only explain 15% of QoLE variance in our results data. Furthermore, a noteworthy limitation of the current version of the framework is that it does not provide a single model featuring all predictors in one equation. This is a consequence of constraints imposed on the design of the ZüriVR study and the resulting partitioning of the dataset. To this end, a dedicated study would be required that uses e.g. SOPI and VRQoE in the same test conditions.

Table 14: Framework of models for Predicting Quality of Training Experience (QoLE).

Group	**Model Structure**	**Performance** Marginal R^2 / Conditional R^2
1	qole ~ sopi_eng + sopi_ecv + sopi_nef + (1 | Participant)	0.150 / 0.491
2	qole ~ vrqoe + eud + (1 | Participant)	0.429 / 0.562
3	qole ~ tac_use + tac_img + tac_enj + age + (1 | Participant)	0.418 / 0.679

5 Conclusion and Implications for the Project

In this deliverable, a framework for measurement and modeling of VR training experience in the context of DMA training for police forces has been developed. To this end, a range of different qualitative and quantitative experience assessment instruments from different research fields were adopted and customized with the following goals in mind:

a) comprehensively **assess end-user perception and training experience** of different training systems,

b) **quantify the relevance of the different experience dimensions** assessed and model their impact on training experience, and

c) empirically ground **recommendations which measurement instruments should be used** to assess VR-based police training technologies and systems in future tasks and activities.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

At the core of the empirical data collection was a large VR training experience study (ZüriVR) which was conducted in Q2+Q3 2020 with the City Police of Zurich, Switzerland. The study was designed to answer five research questions related to the impact and experience of two virtual training systems (VirtTra and Refense VR).

Regarding **measurement and evaluation results** of VR training experience, the results show that both virtual training systems were highly positively received by trainers and trainees alike. As regards differences, VirTra received higher acceptance, enjoyment and ease of use ratings, while the VR (Refense) system resulted in higher immersion and presence. The study results reveal the strengths and weaknesses of the (current versions of the) two tested systems and the underlying technologies used. The VirTra system with its 2D-cinema based setup favors team decision-making and realistic weapon handling. In contrast, the Refense VR system with its 3D stereo-vision HMDs currently favors procedural learning and tactical training. However, with **careful design** and **feature additions** (e.g. **spatial audio**, see below), the **VR system's training experience** and **application range can be significantly increased**.

Regarding **VR training experience measurement and evaluation methods**, our results show that almost all instruments used yield plausible, valid results, with the exception of AttrakDiff (short version). Furthermore, the technology acceptance questionnaire used could be slightly improved, since convergence metrics do not reach the desired levels (albeit they come very close to the required thresholds). The instruments that have turned out to be most useful ones in terms of insight and information value, are: SOPI (presence & immersion), TAC (technology acceptance), and QoLE (Quality of Learning Experience). Furthermore, the two visual analogue scales (VAS for stress and mental exertion), albeit they do not represent experience dimensions per se, provide relevant information regarding (intended) mental and emotional impact of the experienced training setting.

Concerning **VR training experience modeling**, our results suggest that the acquired VR training dataset, albeit featuring a high number of observations (n>1000), does not support development of a structural equation model with sufficient convergence. This was compensated by switching to linear mixed modeling (LMM), which results in a framework of three models that can be used to predict the central construct of quality of learning experience on behalf of instruments like SOPI, TAC or VRQoE that cover specific dimensions of the VR experience.

Regarding **implications for the design of future VR-based training systems** (including the SHOTPROS VR Simulator-Toolkit), the qualitative and quantitative results yield a number of insights and recommendations. In particular, we found that:

- VR truly excels at delivering **high levels of presence and immersion** for DMA training. However, this property **does not automatically guarantee high acceptance or a better training experience** of trainees. In particular, the VR training system must **avoid the look and feel of a video game**, since otherwise wrong perceptions and associations are triggered, which might even result in much lower stress levels than required for police training.
- Another confirmed strength of VR is the possibility to perform **comprehensive, detailed after-action reviews**. The ability to jointly replay and reflect on what has happened directly after executing a training scenario is a feature that is highly regarded by trainers and trainees alike and a strong differentiator of VR training systems compared to other approaches.
- **Realistic weapon handling** is a frequently articulated requirement for virtual training. Participants notice if a physical weapon dummy does not match the simulated one, particularly if it does not resemble their regular service weapon.
- **Realistic 3D content** that matches the look and feel of potential real-world sites of operation represents a strong differentiator and strongly increases the perceived presence and credibility of the virtual training.
- **Realistic audio rendering** is another essential requirement because localization of audio sources and natural communication with others nearby are critical for dealing with almost any police training scenario. Thus, spatial audio rendering represents a necessary feature and in addition, special attention needs to be paid to making face to face communication within the VR as natural as in reality in order to avoid a walkie-talkie like experience.
- In general, **interaction with objects, non-playable characters (NPCs) and team members** need strong support by the VR system. Critical objects (like certain walls) should have well-aligned physical counterparts, NPCs should be able to exhibit plausibly intelligent and responsive behavior, team members' avatars should look distinct (allowing for peer identification), and non-verbal communication with them should be supported to enable natural interaction and coordination.
- Adding an **artificial pain stimulus** via an electro-shocking device increases reported stress and mental exertion levels which is positive in the context of training decision-making and action under stress. The feature received positive mentions in the interviews and did not negatively affect other experience factors like acceptance or sense of presence.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Above results described in this deliverable impact and influence the SHOTPROS project in the following ways:

- The qualitative and quantitative results of the measurement activities inform and confirm the requirements analysis in **WP2** as regards **essential user needs and critical features** required for VR-based DMA training as elicited from trainers and trainers who have been exposed to two virtual training systems.
- Knowledge of the relevance of different experience dimensions as well as the impact of different influencing factors (e.g. stimuli) will guide the **development of training concepts and curriculum propositions** in **T3.3**.
- In turn, these results in essential requirements and critical features that matter also influences the **development agenda** of the contextual VR Simulator-Toolkit in **WP5**. Furthermore, the results on measurement and modeling directly inform the **VR results dashboard development in T5.4**, since QoLE-related KPIs and feedback can serve as indicators for properly working training scenarios and training sessions.
- Furthermore, the developed training experience framework will inform subsequent **human-factors studies** in **WP6** with regard to measurement instruments to be used in the different envisaged experiments. In the same way, the **field trials in the evaluation phase** of **WP7** will follow the measurement and modeling recommendations stated in this deliverable.
- Finally, the developed measurement approaches will translate to **benchmarking methods** (for comparing different systems) that become part of the policymaker toolkit in **WP8.**

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

6 References

Adams, W. C. (2015). Conducting Semi-Structured Interviews. In Handbook of Practical Program Evaluation (S. 492–505). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119171386.ch19

Bayerl, P. S., Davey, S., Lohrmann, P., & Saunders, J. (2019). Evaluating Serious Game Trainings. In B. Akhgar (Hrsg.), Serious Games for Enhancing Law Enforcement Agencies: From Virtual Reality to Augmented Reality (S. 149–169). Springer International Publishing. https://doi.org/10.1007/978-3-030-29926-2_9

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cipresso, P., Chicchi Giglioli, I., Alcañiz Raya, M., & Riva, G. (2018). The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature. Frontiers in Psychology, 9, 2086. https://doi.org/10.3389/fpsyg.2018.02086

Hammer, F., Egger-Lampl, S., & Möller, S. (2018). Quality-of-user-experience: A position paper. Quality and User Experience, Quality and User Experience, 3(1), Springer, https://doi.org/10.1007/s41233-018-0022-0

Hassenzahl, M., Burmester, M. und Koller, F. (2008). Der User Experience auf der Spur: Zum Einsatz von www.attrakdiff.de. In: Brau, H. et al. (Hrsg.), Usability Professionals 08, German Chapter der Usability Professionals' Association. (S. 78–82). Stuttgart: Fraunhofer IRB Verlag.

Houtman, I. L. D., & Bakker, F. C. (1989). The Anxiety Thermometer: A Validation Study. Journal of Personality Assessment, 53(3), 575–582. https://doi.org/10.1207/s15327752jpa5303_14

Huang, H.-M., Liaw, S.-S. & Lai, C.-M.. (2013). Exploring learner acceptance of the use of virtual reality in medical education: a case study of desktop and projection-based display systems, Interactive Learning Environments, 24:1, 3-19, DOI: 10.1080/10494820.2013.817436

Kirkpatrick, D. L., Kirkpatrick , J. D. (2006). Evaluating Training Programs – The four Levels. 3. Ausgabe. 2006, ISBN 1-57675-348-4.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Le Callet, P., Möller, S. and Perkis, A., Eds. (2013). Qualinet White Paper on Definitions of Quality of Experience (2012). European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.2, March 2013.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J.D. (2001). A Cross-Media Presence Questionnaire: The ITC Sense of Presence Inventory. Presence: Teleoperators and Virtual Environments, 10(3), pp 282-297.

Möller, S., & Raake, A. (Eds.). (2014). Quality of Experience. Springer International Publishing. http://link.springer.com/10.1007/978-3-319-02681-7

Perkis, A., Timmerer, C., et al. (2020). QUALINET White Paper on Definitions of Immersive Media Experience (IMEx), European Network on Quality of Experience in Multimedia Systems and Services, 14th QUALINET meeting (online), May 25

Schatz, R., & Talypova, D. (2020). ZüriVR Quantitative Analysis of Experience Assessment Results. SHOTPROS Internal Report, Nov 15.

Venkatesh, V. & Bala, H. (2008), 'Technology acceptance model 3 and a research agenda on interventions', Decision Science 39(2), 273–315.

7 Appendices

7.1 Appendix A: Questions for Quantitative Assessment

Questionnaire	Item Code	Label	Question	Response	Response Code
Training Information	Q1.1	Group	To which training group do you belong?	1	A
				2	B
				3	C
	Q1.2_1	Participant	Participant number		
	Q1.3	Training	Which training did you do?	1	VirTra Shooting Simulator
				2	VR Training
Quality of Learning	Q2.2	qol_1	How sure are you that you can put what you learned in this training into practice?	1	Not at all sure
				2	Only partially sure
				3	Neutral
				4	Sure
				5	Very safe
	Q2.3	qol_2	If any of the situations trained with this system occur in practice, I will be able to master them better.	1	Doesn't apply at all
				2	Doesn't apply
				3	Neutral
				4	Applies
				5	Applies completely
	Q2.4	qol_3	Thanks to the training, I will be able to deal with demanding operational situations more safely in the future.	1	Doesn't apply at all
				2	Doesn't apply
				3	Neutral
				4	Applies
				5	Applies completely
	Q2.5	qol_4	After their experience with the VR training system, how do they assess the usefulness of complementary VR training in police training?	1	Not at all meaningful
				2	Not meaningful
				3	Neutral
				4	Meaningful
				5	Extremely meaningful
VAS	Q3.1_1	vas_str	0-100		Visual Analogue Scale - Stress Thermometer
	Q3.2_1	vas_rsm	0-150		Visual Analogue Scale - Rating Scale for Mental Effort

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

Quality of Experience	Q5.2_1	adiff_p1	Simple - Complicated	1= left 7= right	Items: 10 semantic differential pairs Scale: 7-point radio button scale between the semantic differential pairs
	Q5.2_2	adiff_a1	Ugly - Attractive		
	Q5.2_3	adiff_p2	Practical - Impractical		
	Q5.2_4	adiff_h1	Stylish - Tacky		
	Q5.2_5	adiff_p3	Predictable - Unpredictable		
	Q5.2_6	adiff_h2	Cheap - Premium		
	Q5.2_7	adiff_h3	Unimaginative - Creative		
	Q5.2_8	adiff_a2	Good - Bad		
	Q5.2_9	adiff_p4	Confusing - Clearly structured		
	Q5.2_10	adiff_h4	Dull - Captivating		
	Q5.4	vrqoe_1	How would you rate the overall quality of your experience with the system?	1	Bad
				2	Poor
				3	Fair
				4	Good
				5	Excellent
	Q5.5	vrqoe_2	How would you rate the visual quality of your experience with the system?	1	Bad
				2	Poor
				3	Fair
				4	Good
				5	Excellent
	Q5.6	vrqoe_3	How would you rate the audio quality of your experience with the system?	1	Bad
				2	Poor
				3	Fair
				4	Good
				5	Excellent
	Q5.7	vrqoe_4	How would you rate the quality of the interaction (responsiveness, controllability, freedom to move and act) with the system?	1	Bad
				2	Poor
				3	Fair
				4	Good
				5	Excellent
	Q5.9	eud_1	Training with the system makes me feel fulfilled.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q5.10	eud_2	Training with the system provides me with a sense of purpose.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree

				4	Agree
				5	Strongly agree
	Q5.11	eud_3	Training with the system will make me a better police officer.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q5.12	eud_4	Training with such a system will help me in developing my personal potential.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
Technology Acceptance Questionnaire (for Police VR use)	Q6.2_1	tac_eas_1	I think the virtual environment is easy to use	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.2_2	tac_eas_2	I think the virtual environment is comfortable to use	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.2_3	tac_eas_3	It was easy for me to learn how to use the virtual environment	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.2_4	tac_use_1	The virtual environment allows me good training performances	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.2_5	tac_use_2	The virtual environment helps me to better understand the training content	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.2_6	tac_use_3	The virtual environment helps me gain knowledge about training content	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree

				4	Agree
				5	Strongly agree
	Q6.3_1	tac_use_4	I think the virtual environment is a good training tool	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.3_2	tac_inten_1	I think the virtual environment supports me in my training progress	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.3_3	tac_inten_2	I would be ready to use the virtual environment in my future workouts	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.3_4	tac_inten_3	I would be willing to share my knowledge of the virtual environment with other trainees	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.3_5	tac_inten_4	I would like other trainees to use the virtual environment for training	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.3_6	tac_img_1	I think the virtual environment helps me assess my position relative to my teammates	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.4_1	tac_img_2	I think the virtual environment helps me to experience my own danger realistically	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.4_2	tac_img_3	I think the virtual environment helps me	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree

			better understand critical processes	4	Agree
				5	Strongly agree
	Q6.4_3	tac_img_4	Training in the virtual environment is more interesting than without a virtual environment	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.4_4	tac_imm_1	Training in the virtual environment is fun	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.4_5	tac_imm_2	I like to use the virtual environment for training	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.4_6	tac_imm_3	I inform myself about electronic devices, even if I have no intention of buying.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.5_1	tac_inter_1	I love owning new electronic devices.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.5_2	tac_inter_2	I'm thrilled when a new electronic device comes on the market.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.5_3	tac_inter_3	I like to go to the specialist trade for electronic devices.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.5_4	tac_inter_4	I enjoy trying out an electronic device	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

				4	Agree
				5	Strongly agree
	Q6.5_5	tac_enj_1	Training in the virtual environment is more interesting than without a virtual environment	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.5_6	tac_enj_2	Training in the virtual environment is fun	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.5_7	tac_enj_3	I like to use the virtual environment for training	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.6_1	tac_cur_1	I inform myself about electronic devices, even if I have no intention of buying.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.6_2	tac_cur_2	I love owning new electronic devices.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.6_3	tac_cur_3	I'm thrilled when a new electronic device comes on the market.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.6_4	tac_cur_4	I like to go to the specialist trade for electronic devices.	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree
				4	Agree
				5	Strongly agree
	Q6.6_5	tac_cur_5	I enjoy trying out an electronic device	1	Strongly Disagree
				2	Disagree
				3	Neither agree, nor disagree

				4	Agree
				5	Strongly agree
SOPI[7]					
User Background	Q4.3_1	BG1	How do you rate your level of computer experience?	1	New
				2	Beginner
				3	Advanced
				4	Expert
	Q4.3_2	BG2	How do you rate your level of knowledge about how 3D images are produced?	1	New
				2	Beginner
				3	Advanced
				4	Expert
	Q4.3_3	BG3	How do you rate your level of knowledge about virtual reality (for example how it works)?	1	New
				2	Beginner
				3	Advanced
				4	Expert
	Q4.4	BG4	How often do you play computer games?	1	Never
				2	Occasionally (once or twice a month)
				3	Often, but less than half the days
				4	Half or more of the days
				5	Every day
	Q4.5	BG5	Have you experienced virtual reality before? (multiple answers possible)	1	No
				2	Yes, with a consumer system
				3	Yes, with a professional system in an arcade
				4	Yes, in a training environment
				5	Yes, in a research setting
				6	Yes, otherwise
	Q4.6	BG6		TEXT	If 6, specify
Spacial Presence	SPA01 – SPA19			1-5	Strongly Disagree – Strongly agree
Engagement	ENG01 – ENG13			1-5	Strongly Disagree – Strongly agree
Ecological Validity/ Naturalness	ECV01 – ECV05			1-5	Strongly Disagree – Strongly agree
Negative Effects	NEF01 – NEF06			1-5	Strongly Disagree – Strongly agree

Additional Item	SOPIB6		1-5	Strongly Disagree – Strongly agree
AIT Extra Items	SOPIX1	How would you rate the overall quality of your experience with the system?	1	Bad
			2	Poor
			3	Fair
			4	Good
			5	Excellent
	SOPIX2	Did you experience problems?	1	Yes
			2	No
	SOPIX3	If yes, which ones?	Text	

7.2 Appendix B: Guidelines for Qualitative Interviews

7.2.1 Questions to both trainees and trainers

1) What was positive, what worked well?
2) What was negative, what did not work well?
3) Which ideas/proposals do you have for improving the training?

7.2.2 Questions to trainers only

1) Which training objectives can be trained well with the system from your point of view?
    a) Tactical training: *tactical procedures such as entering a spacing, scanning a room, car procedures*
    b) Personal safety ("Eigensicherung"): *distance to suspect, protection within a team (e.g. 360 degree protection)*
    c) Shooting and weapon handling training: *correct handling and precise shooting of the various service weapons*?
    d) Fitness training: *physical components such as endurance and strength*
    e) Combat training: *various close combat skills for self-protection and to handcuff suspects (also includes training with the baton)*
    f) Law and regulations training: *theoretical lessons and scenario training in which laws and regulations need to be considered*

[7] Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J.D. (2001). A Cross-Media Presence Questionnaire: The ITC Sense of Presence Inventory. Presence: Teleoperators and Virtual Environments, 10(3), pp 282-297.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.

   g) Communication training: *de-escalation tactics, contact communication, regular interactions with civilians*
   h) Perception and action ("Wahrnehmung und Verhalten") *perceiving suspicious behaviour/threats instantly and reacting/behaving fast and correctly (training of quick reaction time with minimal error in perception, also decision-making training)*
   i) Situation training: *exposure to various scenarios to combine skills and competencies a nd familiarize officers with different levels of stress*
   j) Psychological competency training: *enhancing mental capabilities and techniques to r educe stress (e.g. breathing techniques), exerting situational control, etc.*

2) How should training progress be measured best in the VR system from your point of view?
3) What is your overall opinion about integrating gamification elements into training? (e.g. symbolic rewards, medals, achievements points, pain stimuli)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No 833672. The content reflects only the SHOTPROS consortium's view. Research Executive Agency and European Commission is not liable for any use that may be made of the information contained herein.