HOLISTIC EVALUATION OF AR/VR-TRAININGS IN THE ARSUL-PROJECT

Jennifer Beckmann, Katharina Menke, Peter Weber

South Westphalia University of Applied Sciences, Center for Business Education (GERMANY)

Abstract

With the appearance of consumer hardware for virtual and augmented reality (VR/AR), the range of available applications has continuously grown. Today, VR and AR are renowned technologies that also draw attention in education, where especially the immersive and contextual experience are considered promising with regard to authentic and scalable learning environments.

This paper introduces the ARSuL Project (Augmented Reality Based Support for Learning in the Heating, Ventilation and Air Conditioning Industry), which aims to develop a system that assists craftsmen on site and during their vocational and further education by means of E-Learning, AR, and VR. The project consortium developed prototypes and demonstrators, which need to be tested and evaluated to better understand their potential, their shortcomings, and especially the user experience they create. As literature does not provide a proven holistic and integrated evaluation framework for AR / VR trainings, we present the integrated evaluation concept that we developed for ARSuL, covering the interactive experience and the learning experience of the users, as well as learning outcomes and technology acceptance.

Keywords: Augmented Reality, Virtual Reality, Evaluation, HVAC Industry, Vocational Training.

1 INTRODUCTION

In the past years, virtual reality developed into a promising technology for military training and engineering; moreover, it is a great fit for the entertainment industry. The range of applications has continuously grown since then, as the necessary devices have improved in their performance and their user-friendliness, while at the same time decreasing in price. With the appearance of consumer hardware, today the technologies are no longer limited to well-funded academic and corporate research labs. [1] Not only virtual reality (VR), but also augmented reality (AR) has developed into a renowned technology that draws attention in the industrial and private sector, as well as in education. [2, 3] In education, especially the immersive and contextual experience that AR/VR can provide is considered promising with regard to authentic and scalable learning environments. [4]

This paper introduces the ARSuL-Project (Augmented Reality Based Support for Learning in the HVAC Industry), which focuses on the use of AR/VR in the Heating, Ventilation and Air Conditioning industry (HVAC). The project aims to develop a system that assists craftsmen on-site and during their vocational and further education by means of E-Learning, AR, and VR. Prototypes and demonstrators have been created based on an extensive requirements analysis and with the help of conceptual recommendations for learning environments focusing on specific learning objectives [5]. To prepare the testing and evaluation of these prototypes and demonstrators, an appropriate and holistic evaluation concept is required. A literature review revealed that there is no such integrated evaluation concept available, which could be adopted for the evaluation of the AR and VR demonstrators and to compare the effectiveness of the developed AR and VR trainings with that of a classical face-to-face training. Therefore, this paper aims to compose a holistic evaluation concept for the ARSuL-Project to address the following questions:

- How do the participants experience and evaluate the AR, VR, and the classical face-to-face trainings in comparison?
- How is the knowledge gain in the AR and VR trainings in comparison to the classical training?
- What problems do users face when working with the developed prototypes/demonstrators?

2 AR AND VR IN EDUCATION

In 1994, Milgram and Kishino [6] developed the reality-virtuality continuum, which postulates a continuous transition between real and virtual environments. On the continuum, the left side defines real-world environments (cf. Fig. 1). Virtual environments (VE), classified on the right side, are completely composed of virtual objects. In mixed reality environments, real and virtual objects are combined in various ways. While in augmented reality real objects predominate, in augmented virtuality, virtual objects prevail.

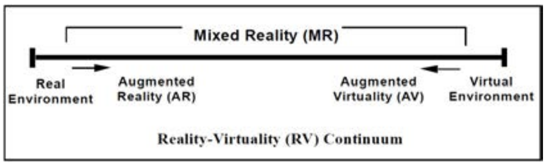

Figure 1. Reality-Virtuality Continuum [7].

Following Jerald (2015), “Virtual reality is defined to be a computer-generated digital environment that can be experienced and interacted with as if that environment were real” [1]. This realistic experience is enabled by the immersive, interactive, multisensory, user-oriented [8] nature of VR. In contrast, augmented reality is a direct, interactive enhancement [9] of the real world with virtual objects, allowing for real-time interaction through a three-dimensional link between virtual and real objects [10].

In 1969, Edgar Dale postulated that active usage of human senses and motor skills can increase learning and understanding [11]. His research is summarized in the ‘Cone of Experience’, in which eleven stages of abstraction and use of senses in learning are differentiated. Dale claims that direct purposeful experience provides the best basis for understanding, while on the other hand teachers should not misunderstand the direct purposeful experience as the only effective way of enabling learning and understanding. The stages can be mixed and are interrelated.

AR and VR then appear to be promising choices for creating engaging and motivating learning experiences, as they enable learners to actively participate and as concepts are made intuitively accessible. Embedding abstract information (text, symbols, multimedia, etc.) directly into the real or virtual world can foster understanding. [1] VR can create immersive experiences similar to the real world, while at the same time providing virtual flexibility, like e.g. the power of suspending physical laws, or making extreme situations approachable in a risk-free manner. In AR, computer-generated objects allow for perception enhancing additions. Broad and complex information can be visualized and thus taught in an intuitive manner, helping users to learn sophisticated and difficile tasks, like e.g. in medicine. [12]

3 THE ARSUL-PROJECT

The ARSuL-Project, funded by the Federal Ministry of Education and Research of the Federal Republic of Germany, focuses on lifelong learning with the help of AR/VR trainings for the HVAC industry. The three-year project aims to develop a system that assists craftsmen at their workplace and during vocational training and further education with the help of E-Learning, AR, and VR. Besides South Westphalia University of Applied Sciences, Ruhr West University of Applied Sciences, Hochschule Niederrhein (University of Applied Sciences), the heating manufacturer Vaillant, and the Central Agency for Continuing Vocational Education and Training in the Skilled Crafts (ZWH) belong to the project consortium. Additionally, four craft businesses with more than 15 employees each, support the project as associated partners.

The project aims at a modern, flexible, technology-enhanced learning and teaching approach for the HVAC industry, focusing on exemplary content and the development of prototypes and demonstrators. The ARSuL project follows the ADDIE instructional design model [13], to systematically develop these learning objects and according trainings, including the project phases depicted in fig. 1. The project’s approach includes two iterations of development, implementation and evaluation, to allow for an evaluation-based improvement of the developed prototypes (cf. Fig. 2).

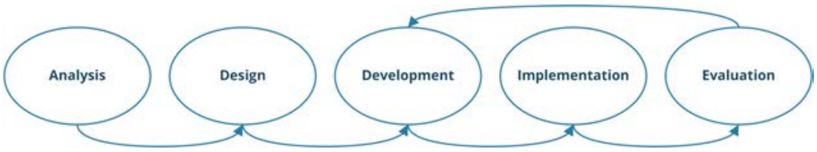

Figure 2. Instructional Design Model used in ARSuL (based on the ADDIE Model [13]).

In the analysis phase, general conditions, the needs, requirements and existing problems of the target group were assessed [5]. The results from the analysis were used to derive design elements for the didactical design and the conception of the AR and VR prototypes, as well as for the composition of learning scenarios for AR and VR trainings [5]. The construction and implementation of the prototypes followed in development phase. The prototypes and demonstrators developed in the first iteration were then implemented in a test setting and evaluated in several test runs for a systematic evaluation, to understand their potential, their shortcomings, and especially the user experience they create. This paper addresses the evaluation concept used to assess the AR and VR demonstrators, prototypes, and learning scenarios.

4 RELATED WORK

To identify the current state of research on the evaluation of AR/VR trainings and to find a suitable evaluation concept for the ARSuL-Project, a structured literature review according to Webster and Watson (2002) was carried out [14]. Title-and-Abstract searches in journals and conference contributions (only full papers) in the online databases of ScienceDirect, IEEE, LearnTechLib, SpringerLink, JSTOR, SAGEPub, Wiley, and Taylor & Francis were conducted with the time constraint '2012–now'. This time frame was chosen, as the announcement of the first consumer Head-Mounted-Display (HMD) Oculus Rift for VR on Kick-Starter in 2012 and the selling start of Google Glasses in 2013 are considered as milestones in the development of AR and VR technology [15].

During the search process, the keywords "virtual reality" and "augmented reality" were each combined with» (learning OR training OR education) AND ("evaluation concept" OR "concept for evaluation") NOT children NOT "high school" NOT cave NOT "second life" NOT simulator«. In this way, we aimed to eliminate learning applications for children or school purposes, to limit our search to HMD-VR systems, and to exclude medical simulators. Further, we excluded computer-based virtual environments displayed on computer screens (like second life), to focus on immersive applications for VR. The search process resulted in a corpus of 66 papers. After analyzing the titles and the abstracts of these papers, only three were assessed in depth, leading to zero matching papers (cf. Table 1).

Table 1: Literature Review Process.

	Total	*ScienceDirect*	*IEEE*	*LearnTechLib*	*SpringerLink*	*JSTOR*	*SAGEPub*	*Wiley*	*Taylor & Francis*
Total Papers	66	0	53	0	13*	0	0	0	0
After analyzing title and abstract	3	0	3	0	0	0	0	0	0
After detailed analysis	0	0	0	0	0	0	0	0	0

* *Restriction* to search only in title and abstract not possible

In summary, we found no papers proposing an evaluation concept neither for AR trainings nor for VR trainings. As a second step, we therefore examined textbooks on augmented and virtual reality [1, 16, 17], to find clues on how to evaluate AR/VR applications. Table 2 summarizes the identified aspects and the evaluation categories that we derived from the findings: *interactive experience*, *learning experience*, *learning outcomes* and *learner variables*. In the following section, we will now examine these categories and the related factors in detail and present available measuring methods.

Table 2: Evaluation factors and components for AR and VR applications.

Source	Interactive Experience				Usage (Learning) Experience		(Learning) Outcomes	Learner Variables		
	Simulator Sickness*	Presence*	Usability	User Experi- ence	Cognitive Load	Satis- faction	Human performance measures (knowledge, errors, task performance)	Demo- graphical Data	Back- ground / Experience	Open Questions (bugs...)
Jerald 2015 – The VR Book [1]	x		x			x			x	
Hale and Stanney (2015) – Handbook of virtual environments [17]	x	x	x				x			
Dörner et al. (2013) – Virtual und Augmented Reality (VR / AR) [16]			x	x	x	x		x	x	x

* Only possible in VR applications

5 EVALUATION COMPONENTS

In education, AR and VR can enable user experiences, not possible for students elsewhere. One of the most important design objectives for VR and AR systems is a human-centered interaction design, focusing on the user perspective. The users need to be able to work in an intuitive manner, the applications need to provide a pleasurable experience, and user frustration must be avoided [1] Thus, when aiming at a high user acceptance and also at high learning outcomes of AR and VR trainings, the *interactive experience* and the *learning experience* of users require special attention.

The *interactive experience* relates to the AR/VR application itself and results from its usability and the created user experience. While usability assesses how easy and pleasant the AR / VR application can be used, the user experience also takes subjective feelings of the users into account, and also considers effects that arise before or after the use of an application [18]. The *learning experience* of users is influenced by the learners' satisfaction and their flow experience. The *interactive experience* thereby influences the *learning experience*. Both, *learning outcomes* and *technology acceptance* depend on the two explained experience components. *Learning outcomes* can be measured in a subjective or an objective way, or they can reveal themselves indirectly as a transfer of knowledge (in real-world problems). Finally, *learner variables,* like e.g. age, previous knowledge, experience with AR or VR technology, need to be considered as covariates. Figure 3 summarizes the derived holistic evaluation concept.

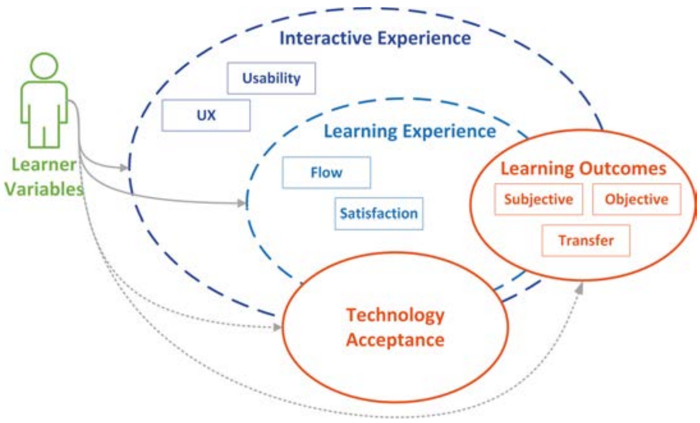

Figure 3. Components of a holistic evaluation concept.

The above evaluation concept allows for various hypothesis that could be assessed in the context of the ARSuL project, like e.g.:

1 A better interactive experience leads to a better learning experience.

2 A better learning experience leads to higher learning outcomes.

3 A better interactive experience leads to a higher technology acceptance.

4 A better learning experience leads to a higher technology acceptance.

5 More experience with AR/VR technology leads to higher technology acceptance.

6 Higher prior knowledge and professional practice leads to higher learning outcomes.

5.1 Interactive Experience

Carruth [19] states that „a virtual training tool should be assessed for usability and for user experience“ (UX). To measure usability, he suggests using a questionnaire such as the System Usability Scale (SUS) [20]. Brooke [20] defines usability as the “general quality of the appropriateness to a purpose”, which means that the systems need to fit a specific context. Brooke developed the SUS, which allows to gain global and subjective information on the users’ perception of the ‘ease of use’. The SUS is a 10-item 5-point Likert questionnaire, ranging from “strongly disagree” to “strongly agree” [20], which results in an easy-to-understand score from 0 (negative) to 100 (positive) [21].

In addition, Bangor et al. [21] worked on the meaning of SUS scores in an absolute sense. They added an adjective rating scale to the SUS, which can help practitioners to interpret individual SUS scores. [21]

For user experience, Carruth [19] suggests to consider two components, presence on the positive side, and simulator sickness on the negative side. Witmer and Singer [22] state that simulator sickness and presence have a negative relationship, since symptoms associated to simulator sickness draw attention away from the virtual environment and thus decrease involvement. In their study, they prove this correlation as significant.

In general, user sickness is a major problem of VR environments [23]. Terms like motion sickness, cyber sickness, simulator sickness, etc. are often used interchangeable, but in fact refer to different causes of user sickness. While motion sickness names symptoms caused by the exposure to real (physical or visual) and/or apparent motion, cyber sickness is visually induced motion sickness, which results from immersion in computer-generated virtual worlds. The term cyber sickness does not include non-motion sickness, like accomondation-vergence conflict, display flicker or fatigue. [24] Simulator sickness results from shortcomings of simulations, e.g. when the simulated motion that does not match physical motion. It includes sickness caused by accomodation-vergence conflict and flicker, but sickness caused by the actual situation simulated, like e.g. acrophobia, is not included. In summary, currently there is no generally accepted term covering all sicknesses resulting from VR usage. [1]

The Simulator Sickness Questionnaire (SSQ) introduced by Kennedy et al. [25] measures symptom presence and severity both before (pre-) and after (post-) the exposure to VR and was originally developed to assess simulator motion sickness. Nevertheless, the SSQ is also often used to measure sickness in VR applications [1, 26]. Subjects are asked to rank each of the 16 symptoms included in the SSQ on a 4-point scale as “none”, “slight”, “moderate” or “severe”. The symptoms are usually clustered into the three categories oculomotor, disorientation and nausea, each of the categories representing a dependent subscore, and resulting in a total (overall) sickness score. The SSQ can thus be used to identify improvement needs for VR applications [1].

According to Slater and Wilbur [27], immersion is the objective and quantifiable degree to which a VR system is able to deliver an inclusive, extensive, surrounding, and vivid illusion of reality to the senses of the users. Immersion is an important part of the VR experience and has the potential to engage users, as it aids them to perceive and interpret the stimuli from within the virtual environment. [1] Presence, on the contrary, results from a user’s feeling of being immersed in a virtual environment. The virtual environment is temporarily becoming reality and the user forgets the real world and fades out stimuli from there for the time using the VR application. [23] Thus, immersion creates presence in VR. The better the immersion in a VR application, the higher the chances of a user feeling presence. [1] Presence refers to a state in which a user consciously acts in the VR environment as if he was in the real world. It can be facilitated by different parameters of the virtual environment, like e.g. a large field of vision, active head tracking and real movement. [23] Presence can be measured with the help of self-reported questionnaires, through physiological data, like e.g. skin conductance level, pulse, etc., or through observation of user behavior. [23]

Witmer and Singer [22] state that “the effectiveness of virtual environments […] has often been linked to the sense of presence reported by users […]”. They define presence “[…] as the subjective experience of being in one place or environment, even when one is physically situated in another.” [22] They developed a Presence Questionnaire (PQ) to measure and evaluate the degree to which individuals experience presence in VR environments [22]. The authors deem that an increased presence will lead to higher learning and performance because of the more meaningful experiences. They group the factors affecting the sense of presence into control, sensory, distraction and realism factors, each influencing presence by affecting involvement, immersion, or both. Version 4.0 [28] (found in [29]) of the PQ is the latest version of the questionnaire. The PQ is a four-factor model with 29 items answered on a 7-point Likert scale. The four subscales that represent aspects that lead to the experience of presence are involvement, sensory fidelity, adaptation/immersion, and interface quality. The PQ also allows for the calculation of an overall presence score.

To summarize, for the evaluation of the *interactive experience*, we choose to use the SUS questionnaire [20] to assess the usability of AR and VR trainings. Further, we will interpret our findings with the help of the approach from Bangor et al. [21]. In addition, the VR prototype will be evaluated regarding its user experience by assessing VR sickness via the SSQ [25] and presence with the PQ [22].

5.2 Learning Experience & Learning Outcomes

Flow is a state, in which people lose themselves in a smoothly running activity. They still have control over the activity despite high demands. [30] Flow occurs at a proper level of difficulty: difficult enough to challenge people and not too difficult to create frustration. [1] The flow state is experienced as pleasant, and can be considered positive, if not connected to highly risky activities. This is especially true for learning activities. [31]

Rheinberg et al. [32] developed the flow-short-scale (FSS, FKS German version), which fits well a broad range of activities and can be used in an easy and fast way. It consists of ten items that represent different components of a *flow experience*. Three additional items measure a user's current *worry* to fail at important things. Rheinberg states that it became apparent that some people react with worry to challenges, whereas others get into a state of flow. Besides the overall flow value (items 1-10), two factorial justifiable sub-scores can be formed: *fluency* (items 2,4,5,7,8,9) and *absorption* (items 1,3,6,10). The items 11,12,13 together build the *worry* score. [31]

The flow score of the FSS proved to be a predictor for performance, as flow correlates significantly with learning performance [33, 34]. Furthermore, Engeser et al. [33] showed, that even when taking competence factors into account, flow still has a significant impact on the performance of students.

The Training Evaluation Inventory (TEI) [35] is an instrument to measure the training outcome dimension, addressing *subjective enjoyment*, *perceived usefulness*, *perceived difficulty*, *subjective knowledge gain* and *attitude towards training* as well as aspects of problem-based training design on ten scales. The training outcome dimension of the instrument is based on Kirkpatrick's widely applied hierarchical evaluation model [36, 37] with the four levels *reactions*, *learning*, *behavior,* and *results*. The level *reactions* measures the enjoyment and the emotional reactions to a training. The level of *learning* instead focuses on the acquisition of knowledge as well as attitude change through a training. The *behavioral* level is focusing on job-related behavior and performance change, indicating a transfer of learning. The level *results* finally refers to the organizational impact of a training, e.g. regarding time and cost.

In the TEI, the levels *reactions*, *learni*ng and attitudes of Kirkpatrick's model are included, since they can be measured in paper-and-pencil format [35]. The basic *reactions* level of this model measures the enjoyment of a training in general. Since other studies further differentiate reactions into *enjoyment of training* (affective reaction), *perceived usefulness of training* (utility reaction) and *perceived difficulty* of training [38, 39], these subcategories were also incorporated into TEI. On the *learning* level, the components *subjective knowledge gain* and *attitude towards training* are measured in a content-independent and universal way [35]. These five concepts are referred to as *training outcome dimensions*. All items of the TEI are answered on a five-point Likert scale ranging from "strongly disagree" to "strongly agree". [35]

Sitzmann et al. [40] were able to show that self-assessment of knowledge correlates moderately with objective learning measures. Thus, self-reported knowledge gain can serve as an approximation for learning success, "although the relationship is less than optimal and results have to be interpreted with caution" [35]. Therefore, it is reasonable to include an objective content-based learning measure in a systematic evaluation setting like the one being developed here for the ARSuL-Project.

For the ARSuL-Project, an objective paper-and-pencil 25-point knowledge test was created to assess the knowledge gain of each participant. The questions were created with the help of a training engineer of the heating manufacturer and project partner Vaillant, who also provided the sample solutions. As question types, gap fill, matching and short answers questions were used.

In summary, to evaluate the *learning experience* we decided to use the FSS [32] to measure flow as a predicator for performance in all training types (AR, VR and f2f). From the TEI [35] we included the subcategories *enjoyment of training*, *perceived usefulness of training* and *perceived difficulty* to measure the satisfaction of the learners with each training, since these are related to the reactions level of Kirkpatrick's model.

The *learning outcomes* are measured trough the mentioned objective knowledge test and the *subjective knowledge gain* component included in the TEI [35]. We do not measure the transfer of learning explicitly at this point.

5.3 Technology Acceptance

The Technology Acceptance Model (TAM) [41] is one of the oldest and one of the most widely used theoretical approaches measuring the acceptance of new technologies. The model aims to explain the determinants of acceptance and to reflect on them in generalized structures.

Technology acceptance models have their origins in psychology, information systems research, and sociology. TAM is based on the theory of reasoned action (TRA) [42] and the theory of planned behavior (TPB) [43]. According to TAM, the major factors influencing and determining users' intentions to use a new technology are *perceived usefulness* and *perceived ease of use*. Vekatesh and Davis [44] define *perceived usefulness* as "the extent to which a person believes that using the system will enhance his or her job performance", and they define *perceived ease of use* as "the extent to which a person believes that using the system will be free of effort".

TAM has been applied to a multitude of technologies and has been tested in various contexts. Furthermore, new factors have been added to increase the predictive validity of the model and to better understand the determinants of technology acceptance. [45]

For the ARSuL-Project, we decided to measure *technology acceptance* through questions included in the SUS questionnaire [20] together with self-developed questions representing the factors *perceived usefulness*, *perceived use of use*, *attitude towards using*, *motivation*, *future* and *behavioral intention to use*.

5.4 Learner Variables

Learner variables, such as previous knowledge, experience, and practice as well as demographical data can serve as predictors or covariates for learning outcomes and technology acceptance. We presume that also previous experience with computer games can impact e.g. the *interactive experience* of a user. Therefore, in our evaluation setup we retrieve previous knowledge with an objective knowledge test, and we ask the participants about prior professional practice as regards the learning content of the training, but also about prior experience with VR/AR technology and/or computer games. Demographical data like gender, age and education is collected as well.

6 EVALUATION CONCEPT

Figure 4 shows the resulting evaluation concept with its components, their factors as well as how and when they were measured in the test setting. We conducted four test runs with 58 craftsmen and apprentices of the HVAC industry. For the test setting, we prepared three different trainings (AR/VR/classical face-to-face) with the same learning content, to be able to compare them. The AR and VR training were developed using the results from the requirements analysis conducted before [5]. Each participant went through all three trainings.

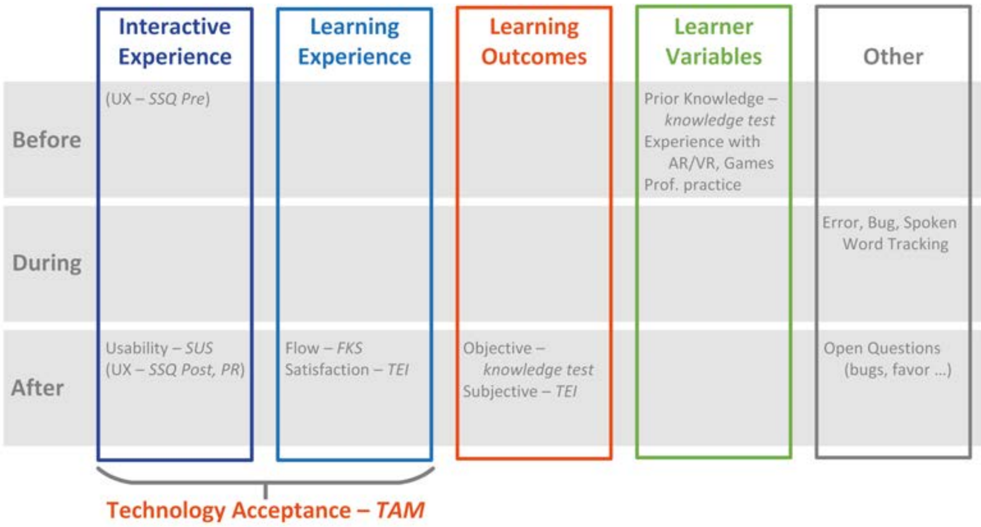

Figure 4. Evaluation Setup.

Subjects filled out the questionnaires regarding their individual prior knowledge, experience and practice before they started with the first training, including a knowledge pre-test. Since the individual knowledge gain can only be measured after the first training, we divided the participants into three groups, and each of the groups started with a different training (VR, AR, classical f2f). During each training, the test coordinators tracked errors made by the participants, documented bugs that became apparent in the applications, and tracked any spoken conversation. After each training, the participants filled out the questionnaires on the state of flow and the training (satisfaction, learning), and they were asked to report on any bugs, to point out possible improvements and to summarize what they liked and what they disliked about the training. Further, the participants evaluated the AR and VR trainings with regard to their usability, and the VR training also on the UX components simulator sickness and presence. Finally, only after their first training, the subjects were asked to complete the knowledge post-test. Technology acceptance was measured through questions, included in the questionnaires on the *interactive* and *learning experience*.

7 CONCLUSION & OUTLOOK

The present paper reveals the need for a holistic evaluation of AR and VR trainings. As the examined literature did not provide any adoptable integrated evaluation concept for the ARSuL-Project, we derived a holistic evaluation framework for AR and VR from according suggestions in textbooks. The resulting evaluation concept comprises the components *interactive experience* and *learning experience, learning outcomes* and *technology acceptance*, with the two latter being affected by the two former. Major goals of the iterative design and development process in the ARSuL-Project must be to foster the acceptance of the AR and the VR learning applications, and to improve the learning outcomes of the users by adjusting the key experience components.

For all factors considered in the evaluation concept, we found suitable models and questionnaires that can be considered proven and reliable. Not yet included and thus not measured during the test runs is the user experience in the AR application.

As a follow-up to this paper, in [46], we provide and discuss first evaluation results from the test runs to derive necessary improvements for the ARSuL-Prototypes [46]. We thereby combine the evaluation feedback that we received with the recommendations of the Universal Design for Learning (UDL), to make sure that our AR and VR solutions contain as few barriers as possible [47, 48].

REFERENCES

[1] J. Jerald, *The VR Book*: Association for Computing Machinery, 2015.

[2] A. Sag, *CES 2018: Virtual Reality And Augmented Reality Get Another Shot.* [Online] Available: https://www.forbes.com/sites/moorinsights/2018/01/25/ces-2018-virtual-reality-and-augmented-reality-get-another-shot/#45983c5aab04. Accessed on: Nov. 23 2018.

[3] T. Hutzschenreuter and C. Burger-Ringer, “Impact of Virtual, Mixed, and Augmented Reality on Industries,” Technische Universität München, 2018. [Online] Available: https://mediatum.ub.tum.de/1454069. Accessed on: Jan. 05 2019.

[4] L. Johnson, S. Adams-Becker, and M. Cummins, *The NMC Horizon Report: 2016 Higher Education Edition*. Toronto, ON, CA: New Media Consortium, 2016.

[5] J. Beckmann, K. Menke, and P. Weber, “AR and VR meet Boiler Suits: Deriving Design Elements for AR / VR Trainings in the HVAC Industry,” in *Proceedings of EdMedia + Innovate Learning 2018*, 2018, pp. 1533–1542.

[6] P. Milgram and F. Kishino, “A taxonomy of mixed reality visual displays,” *IEICE TRANSACTIONS on Information and Systems*, vol. 77, no. 12, pp. 1321–1329, 1994.

[7] P. Milgram, H. Takemura, A. Utsumi, and F. Kishino, “Augmented reality: A class of displays on the reality-virtuality continuum,” in *Telemanipulator and telepresence technologies*, 1994, pp. 282–293.

[8] C. Cruz-Neira, “Virtual Reality Overview,” in *SIGGRAPH ‘93: Course Notes*, 1993.

[9] W. Broll, “Augmentierte Realität,” in *EXamen.press, Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*, R. Dörner, W. Broll, P. Grimm, and B. Jung, Eds., Berlin: Springer, 2013, pp. 241–294.

[10] R. T. Azuma, “A survey of augmented reality,” *Presence: Teleoperators and Virtual Environments*, vol. 6, no. 4, pp. 355–385, 1997.

[11] E. Dale, *Audiovisual methods in teaching,* 3rd ed. New York: Dryden Press, 1969.

[12] A. Mehler-Bicher and L. Steiger, *Augmented Reality: Theorie und Praxis,* 2nd ed. Berlin: De Gruyter Oldenbourg, 2014.

[13] H. M. Niegemann *et al., Kompendium multimediales Lernen*. Berlin, Heidelberg: Springer, 2008.

[14] J. Webster and R. T. Watson, “Analyzing the Past to prepare for the Future: Writing a Literature Review,” *MIS quarterly*, 2002.

[15] N. Statt, *AR has inherited all the promise and hype of VR.* [Online] Available: https://www.theverge.com/2018/1/18/16906640/ar-vr-promise-hype-microsoft-hololens-google-glass-ces-2018. Accessed on: Jan. 11 2019.

[16] R. Dörner, W. Broll, P. Grimm, and B. Jung, Eds., *Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*. Berlin: Springer, 2013.

[17] K. S. Hale and K. M. Stanney, Eds., *Handbook of virtual environments: Design, implementation, and applications,* 2nd ed. Boca Raton, Fla.: CRC Press, 2015.

[18] T. Tullis and W. Albert, *Measuring the user experience: Collecting, analyzing, and presenting usability metrics*. Amsterdam: Elsevier, 2013.

[19] D. W. Carruth, “Virtual reality for education and workforce training,” in *ICETA 2017: 15th IEEE International Conference on Emerging eLearning Technologies and Applications: proceedings : October 26-27, 2017, Starý Smokovec, The High Tatras, Slovakia*, Stary Smokovec, 2017, pp. 1–6.

[20] J. Brooke, “SUS: A quick and dirty’ usability scale,” in *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and I. L. McClelland, Eds., London: Taylor and Francis, 1996, pp. 189–194.

[21] A. Bangor, P. Kortum, and J. Miller, “Determining what individual SUS scores mean: Adding an adjective rating scale,” *Journal of usability studies*, vol. 4, no. 3, pp. 114–123, 2009.

[22] B. G. Witmer and M. J. Singer, “Measuring presence in virtual environments: A presence questionnaire,” *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[23] R. Dörner and F. Steinicke, “Wahrnehmungsaspekte von VR,” in *EXamen.press, Virtual und Augmented Reality (VR / AR): Grundlagen und Methoden der Virtuellen und Augmentierten Realität*, R. Dörner, W. Broll, P. Grimm, and B. Jung, Eds., Berlin: Springer, 2013, pp. 33–63.

[24] B. D. Lawon, “Motion Sickness Symptomatology and Origins,” in *Human factors and ergonomics series, Handbook of virtual environments: Design, implementation, and applications*, K. S. Hale and K. M. Stanney, Eds., 2nd ed., Boca Raton, Fla.: CRC Press, 2015, pp. 531–600.

[25] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal, “Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness,” *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203–220, 1993.

[26] T. Weißker, A. Kunert, B. Fröhlich, and A. Kulik, “Spatial Updating and Simulator Sickness During Steering and Jumping in Immersive Virtual Environments,” in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR 2018): Tuebingen/Reutlingen, Germany, 18-22 March 2018*, Reutlingen, 2018, pp. 97–104.

[27] M. Slater and S. Wilbur, “A Framework for Immersive Virtual Environments (FIVE): Speculations on the Role of Presence in Virtual Environments,” *Presence: Teleoperators & Virtual Environments*, vol. 6, pp. 603–616, 1997.

[28] J. Moss, “Characteristics of Head-Mounted Displays and Their Effects on Simulator Sickness,” Dissertation, Human Factors Psychology, Clemson University, Clemson, South Carolina, 2008.

[29] B. G. Witmer, C. J. Jerome, and M. J. Singer, “The Factor Structure of the Presence Questionnaire,” *Presence: Teleoperators and Virtual Environments*, vol. 14, no. 3, pp. 298–312, 2005.

[30] M. Csikszentmihalyi, *Beyond boredom and anxiety*. San Francisco: Jossey-Bass, 1975.

[31] F. Rheinberg, *Motivationsdiagnostik*. Göttingen: Hogrefe, 2004.

[32] F. Rheinberg, R. Vollmeyer, and S. Engeser, “Die Erfassung des Flow-Erlebens,” in *Tests und Trends*, vol. 2, *Diagnostik von Motivation und Selbstkonzept*, J. Stiensmeier-Pelster and F. Rheinberg, Eds., Göttingen: Hogrefe, 2003.

[33] S. Engeser, “Motivation, Lernaufwand und Lernleistung in der Statistikausbildung Psychologie,” Dissertation, Institut für Psychologie, Universität Potsdam, Potsdam, 2005.

[34] S. Engeser, F. Rheinberg, R. Vollmeyer, and J. Bischoff, “Motivation, Flow-Erleben und Lernleistung in universitären Lernsettings.,” *Zeitschrift für Pädagogische Psychologie*, vol. 19, no. 3, pp. 159–172, 2005.

[35] S. Ritzmann, V. Hagemann, and A. Kluge, “The Training Evaluation Inventory (TEI) - Evaluation of Training Design and Measurement of Training Outcomes for Predicting Training Success,” *Vocations and Learning*, vol. 7, no. 1, pp. 41–73, 2014.

[36] J. D. Kirkpatrick and W. K. Kirkpatrick, *Kirkpatrick’s four levels of training evaluation*. Alexandria, VA: ATD Press, 2016.

[37] D. L. Kirkpatrick, “Evaluation of Training,” in *Training and Development Handbook: A Guide to Human Ressource Development*, R. L. Craig and L. R. Bittel, Eds., New York: McGraw-Hill, 1967, pp. 87–112.

[38] G. M. Alliger, S. I. Tannenbaum, W. Bennett, H. Traver, and A. Shotland, “A Meta-Analysis of the Relations Among Training Criteria,” *Personnel Psychology*, vol. 50, no. 2, pp. 341–358, 1997.

[39] P. Warr and D. Bunce, “Trainee Characteristics and the Outcomes of Open Learning,” *Personnel Psychology*, vol. 48, no. 2, pp. 347–375, 1995.

[40] T. Sitzmann, K. Ely, K. G. Brown, and K. N. Bauer, “Self-Assessment of Knowledge: A Cognitive Learning or Affective Measure?,” *AMLE*, vol. 9, no. 2, pp. 169–191, 2010.

[41] F. D. Davis, “A technology acceptance model for empirically testing new end-user information systems: Theory and results,” Doctoral Dissertation, Massachusetts Institute of Technology.

[42] I. Ajzen and M. Fishbein, *Understanding attitudes and predicting social behavior*. Englewood Cliffs, N.J: Prentice-Hall, 1980.

[43] I. Ajzen, “From Intentions to Actions: A Theory of Planned Behavior,” in *Springer series in social psychology, Action control: From cognition to behavior*, J. Kuhl and J. Beckmann, Eds., 1st ed., Berlin: Springer, 1985, pp. 11–39.

[44] V. Venkatesh and F. D. Davis, “A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies,” *Management Science*, vol. 46, no. 2, pp. 186–204, 2000.

[45] N. Marangunić and A. Granić, “Technology acceptance model: A literature review from 1986 to 2013,” *Universal Access in the Information Society*, vol. 14, no. 1, pp. 81–95, 2015.

[46] J. Beckmann, K. Menke, and P. Weber, “AR/VR-Trainings in the ARSuL-Project: Results from a Prototype Evaluation,” in *INTED2019 Proceedings*, 2019.

[47] J. Beckmann, K. Menke, and P. Weber, “Universal Design for Learning in Augmented and Virtual Reality Trainings,” in *Universal Access Through Inclusive Instructional Design: International Perspectives on UDL*, S. L. Gronseth and E. M. Dalton, Eds.: Routledge, accepted and forthcoming in 2019.

[48] Cast, *Universal Design for Learning Guidelines version 2.2.* [Online] Available: http://udlguidelines.cast.org. Accessed on: Sep. 09 2018.