Media Comparison for Instruction-based AR Usage in Collaborative Assembly

Lea M. Daling, Anas Abdelrazeq, Frank Hees, and Ingrid Isenhardt
Chair of Information Management in Mechanical Engineering (IMA)
RWTH Aachen University
Aachen, Germany
lea.daling@ima-ifu.rwth-aachen.de, anas.abdelrazeq@ima.rwth-aachen.de, frank.hees@ima.rwth-aachen.de,
ingrid.isenhardt@ima.rwth-aachen.de

***Abstract*— Complex work processes are increasingly performed as a cooperation between humans and robots. In this context, assistive technologies play an important role. The implementation of Augmented Reality as instructional tool offers new possibilities to support workplace-oriented learning processes. This paper studies the suitability and usability of different AR-based media in collaborative assembly between human and robot. For this purpose, three different media have been compared in an experimental within-subject design with regard to the usability of the respective hardware and software. The findings gained from the mixed-method approach show that traditional media are considered easier to use, but that the fundamental potential of Augmented Reality application is clearly recognized.**

Keywords - Augmented Reality; Human-Robot-Interaction; Usability Evaluation; On-the-Job Training; Assistive Technology.

I. INTRODUCTION

New assistance technologies are finding their way into formerly manually and analogously designed areas of the digitized industrial working world. Hereby, employees can be directly supported in the work process [1] and are enabled to cope with new and complex requirements of an increasingly individualized and highly flexible production [2]. The use of Augmented Reality (AR) as an instructional assistance tool is widely expected to be a success factor for digital training programs [3]. It allows employees to be guided through assembly processes step by step, and to train them flexibly for new use cases.

Well-designed assistance systems become particularly important when employees have to be trained for complex or novel work processes. In the course of increased interaction between man and machine, scenarios in which both actors simultaneously work together on a task become more widespread [4]. Therefore, this paper focuses on the design and evaluation of AR-based assistance systems in assembly processes that are neither completely manual nor fully automated and take place in cooperation between human and robot [5]. The presented AR application is designed for the instruction of a collaborative assembly process between human and robot.

Since a high degree of usability can be seen as a prerequisite for further performance measures such as time effectiveness and reducing the error rate, the aim of our current research is to test the usability of an AR-based assistance system tested on different media [6]. This paper represents a detailed elaboration of a usability evaluation for the use of AR assistance systems in human-robot collaboration [1]. In Section II a brief introduction of the basic functions and application possibilities of AR in the manufacturing context are given. The use case of AR as an on-the-job instructional tool in a collaborative assembly cell is presented in Section III. In Section IV, an overview of relevant usability criteria is presented. Furthermore, we present our empirical approach to measure usability of an AR application using different instructional media. Finally, Section V gives an overview of the results and an outlook on further research.

II. THEORETICAL BACKGROUND

The following paragraph gives a brief introduction of the basic functions and application possibilities of AR in manufacturing and the use of AR as an on-the-job instructional tool in a collaborative assembly cell.

A. Instructional AR in the Manufacturing Context

An AR system adds virtual objects to the real world, in a way that both virtual and real components homogeneously appear in the user perception. An AR system “combines real and virtual objects in a real environment; runs interactively and in real time and registers (aligns) real and virtual objects with each other” [7]. In other words, AR systems overlay computer-generated objects onto a real world setting, in real time [8].

Within the last 10-15 years, AR systems greatly improved and have shown an ability to create solutions to various problems [9]. Since then, more and more AR tools are developed and applied in the field of industry. The main use of AR in an industrial context is currently related to maintenance, manufacturing, and assembly related tasks [10]. Using AR, innovative and effective methods can be developed to meet important requirements in simulation, assistance and improvement of manufacturing processes. Volvo, for example, is utilizing the Microsoft HoloLens to enable production line workers to digitally view assembly instructions in real-time while working to put together parts of the vehicle [11]. By adding real-time information to a real (working) environment, AR-based systems can minimize the need for improvement iterations, re-works and modifications by ‘getting it done the right way’ on the first try.

As a result, the possibility of “learning on demand” in on-the-job training sessions arises. To date, there are several approaches that combine learning measures at the workplace with the benefits of new technologies [3]. These on-the-job learning approaches connect theoretical knowledge with practical application [18]. Furthermore, they provide tailor-made learning processes and can be used independent of time and learning pace [3]. The use of AR in training processes promises many positive effects, such as constant access to information, lower error rate, improved motivation and a synchronization between training and performance [13]. For instance, a comparison between paper instructions and AR instructions on a Head Mounted Display (HMD) showed that, although the use of AR in the assembly process gives little “time-advantages”, it significantly reduces the assembly errors [12].

Nevertheless, AR systems still face a couple of challenges, preventing a direct implementation of AR solutions in real world problems. Technical developments are often limited to the capabilities of the specific medium instead of analyzing the requirements of the respective task or work process. The focus is therefore more on the technology to be implemented than on the user or the requirements of the task. Many studies already focus on objective key indicators (e.g., time, error rate, accuracy) as dependent variables, even though basic usability factors such as acceptance, perceived usefulness or enjoyment as well as technical questions (e.g., display and tracking technology, calibration techniques, interfaces to the operating devices [13]) should be of central importance and precede the implementation process [14].

Even with those challenges conquered, other questions still arise. Like whether or not the implementation of such systems would lead to other problems affecting the overall performance. An over-reliance on the AR generated signals and indications can have negative implications on the performance of the user, for example by disrupting the attention or focusing it all in one direction, leading it away from the surrounding context [17]. Further research and evaluation of the technology is therefore necessary to solve existing problems and expand the spectrum of applications.

B. *AR as an interface for collaborative assembly*

The current developments in connection with the increasingly networked and individualized Industry not only impose high demands on interconnected technical systems. More complex, dynamic and individualized production processes are also changing the way work as such is organized. Increased interaction between humans and robots is considered to be a future scenario, in which tasks are accomplished together while working on a product or a component at the same time. Particularly in manufacturing, tasks change from manual work to collaborative work processes between humans and robots or machines. While cooperation means that both interacting partners can have tasks in the (common) workspace, but do not work on the same product or component at the same time, collaboration describes the simultaneous execution of a common task on the same product or component (see **Figure 1**) [4].

On the one hand, this changes the requirements placed on the employees who interact with these machines [2]. On the other hand, it changes the requirements placed on assistance systems, which are intended to provide the best possible support for the fulfilment of the respective task. As already mentioned, AR is potentially suitable as an interface between humans and robots, in order to guide through new tasks and procedures in on-the-job trainings. In addition, when used as an assistance system, it can also create transparency about the current status or work steps of the robot.

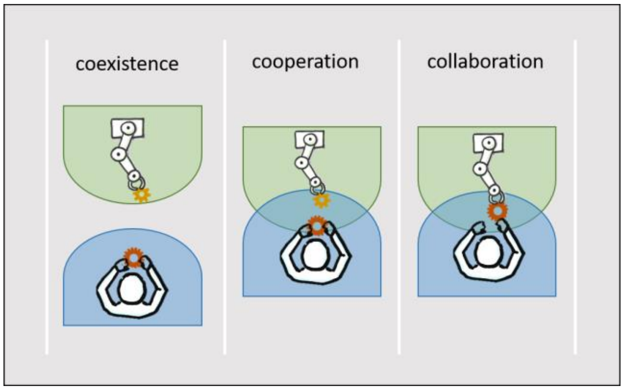


Figure 1. Levels of Human-Robot Interaction [4]

AR as an assistant system for collaborative tasks can be used, to display assembly process information, robot motion and workspace visualization, visual alerts, and production data amongst others [15]. Makris and colleagues, for example, showed that the AR application was able to minimize the time required for operators to access the necessary information and also increased the operator’s acceptance to work with industrial robots without safety fences. Michalos and colleagues used an AR based application that ran on an Android tablet to support human-robot interactive cooperation [16]. Although they achieve positive results, they point out that the AR application should be tested on different media (e.g., head-mounted) and that overall, more focus should be directed towards researching factors such as ergonomics and handling of such media, as well as the design of the AR application itself (e.g., layout of visual aids).

This leads to the assumption that, before the effectiveness of AR and its appropriateness for the use case of human robot collaboration for assembly tasks can be tested, there is still a need for research regarding the usability of AR systems and different media. In addition to that, it is important to gain a deeper understanding on how different aspects of the AR system are perceived and evaluated by the user. In this respect, it is important to first analyze the task or requirements of the work process. The technical development should be iterative and adapted to the needs of the users. When evaluating the AR assistance system, it is advisable to distinguish between hardware (e.g., data glasses vs. tablet) and software (the AR application used) [1]. Therefore, the present qualitative pre-study aims at deriving basic implications on the usability of the developed AR application by not only providing feedback on the AR-capable hardware, but also on the AR application software itself. Throughout the next section, the usability aspects to be considered are explained before the use case and the analysis of the task are presented.

C. *Usability Aspects*

Since 1997, DIN EN ISO 9241 has been an international series of standards that defines usability as the extent to which a technical system can be used by certain users in a certain usage context in order to achieve certain goals effectively, efficiently and satisfactorily [6]. Sarodnick and Brau emphasize that usability particularly considers the fit of system, task and user, while taking into account the quality of goal achievement perceived by the user [6]. For this reason, it is essential to involve potential users in the evaluation process at an early stage.

A survey of Gabbard and colleagues [20] showed that in a total of 1104 articles on augmented reality, only 38 (~3%) addressed some aspect of human computer interaction, and only 21 (~2%) described a formal user-based study. Since, as mentioned, the involvement of users in the evaluation process is crucial for the successful development of a product, a user-centered mixed-method approach will be presented in the following.

A widely used *inductive* approach, characterized by the analysis of early versions and prototypes, is the so-called thinking aloud method [6]. Here, test subjects are encouraged to express their cognitions verbally during the test. The advantage of this approach is the explorative acquisition of qualitative data to receive feedback on design and improvement. However, it should be critically noted that the combined load of task processing and thinking aloud reduces the processing speed. Therefore, this method should not be used in conjunction with a performance measurement. Furthermore, these approaches are barely standardized.

Deductive methods, on the other hand, capture the user's perspective on an already developed system. At this point, however, changes and corrections of a system are often time- and cost consuming. Established evaluation concepts (e.g., IsoMetrics; Isonorm [22]), often make use of the classical questionnaire methodology, which ensures the fulfilment of the quality criteria (validity, reliability, objectivity) to a large extend. The aim of this paper is to combine the advantages of both methods in order to generate feedback on the usability of the AR application based on empirical user surveys - extended by open questions. Furthermore, the thinking aloud method was used to verbalize and record the impressions, reactions and cognitions of the participants during the work process. Before the composition of these approaches is presented in Section IV, the respective use case is presented in the following section.

III. USE CASE AND REQUIREMENTS

The use case in which AR is utilized to enable on-the-job training consists of a collaborative assembly cell equipped with a robot. It represents a common scenario in Industry 4.0, where the digitalization of production is continuously increasing. It also poses special challenges to the interface design of the instructional tool, due to the interaction between humans and robots.

During the assembly process, workers are collaborating with a robot arm (UR-5) to assemble a small gear drive. In total, three plates with gear wheels are assembled. The worker performs five steps, while the robot performs a total of four steps. Once the worker has familiarized himself with the cell, he/she is instructed to position a base plate and rear plate into a holder. The robot then inserts four hexagon socket screws and positions the back plate onto the base plate, while the worker assembles two sets of gear wheels. In the final step, the gear wheels are mounted on the pre-assembled base plate presented by the robot.

Previous studies have shown that an efficient and user-friendly introduction training can contribute significantly to increasing the acceptance of the human-robot interaction [21]. During these studies, the assembly task was guided by a fixed touchscreen with 2D images and 3D animations. Thereby, participants repeatedly had to look up the work steps on the fixed screen.

AR offers the benefit of displaying information and work steps, e.g., 3D overlays, directly within the workspace or the tool required for the respective assembly step. The instructions for the AR application were developed based on the existing work steps and supplemented by virtual objects with real-time animations. During the design of the work instructions, special attention was paid to the fact that the instructions must be comprehensible for laypersons and inexperienced employees. Based on fundamental usability heuristics (e.g., visibility of the system status, consistency and aesthetics) [19] and iterative testing within the interdisciplinary development team, we defined the following requirements for the development the AR-Application:

1. The application is designed to provide non-experts with an **easy and intuitive on-the-job training process** without moving back to the screen.
2. The application should function on **head-mounted** and **handheld AR devices.**
3. The application should be able to **recognize the working space.**
4. The application should illustrate different **step cues** (text, 2D images and 3D animations) for the user.
5. The application should enable a predominantly **hands-free instruction process**. Thus, the worker should be able to either **navigate with touch or voice** control.
6. To ensure a good workflow between human and robot, the AR application should be able to **communicate with the robot** and be aware of its status.

In order to meet these requirements, an AR application was developed that is deployable on both Microsoft HoloLens and Android-based tablets fixed with a tablet arm (see **Figure 2**).

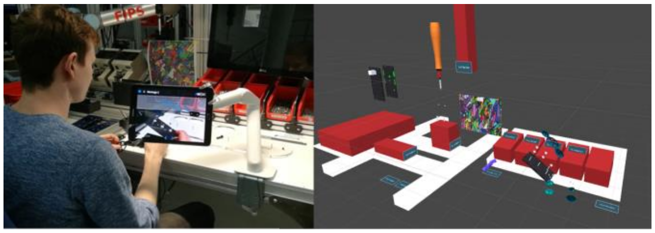


Figure 2. Using HoloLens and Android Tablet as Instructional Media

Unity 3D game engine was adapted for the software development process. The proposed AR application consists of several modules (e.g., game objects) that are organized within a Unity scene (see **Figure 3**). Different components, including the interface layout, are realized within these modules.

In order to provide orientation at the workplace and to locate the devices optimally within the assembly cell, special attention was paid during development to implement *Localization and Tracking*. For this purpose, one marker was placed in the cell for the HoloLens, whereas the tablet uses two different markers for the different work areas in the cell due to its lower tracking capability. If the tracking has been completely lost, the application asks the user to reposition the device to the point where the marker is seen. In addition, guidance arrows are used to indicate the location of the next step. In order to enable dynamic and on demand changes, as part of the *scenario manager*, a configuration ".json" file was developed to configure the steps specifying their text instructions, 2D images and 3D animations, as well as information about the robot status. The *interactive user interface* allows the user to navigate through the steps by either by touching (i.e., clicking) the tablet or by voice control of the HoloLens. *Robot Communication* is ensured based on Robot Operating System (ROS) to initiate a step or to wait for the robot until it finished its step.

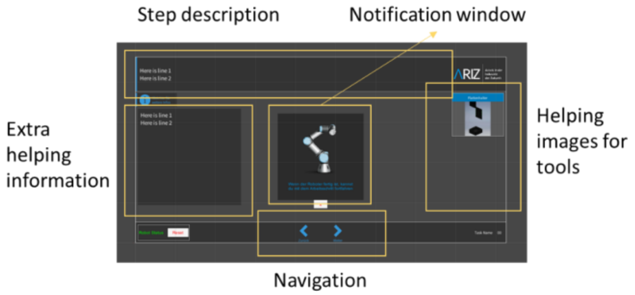


Figure 3. Layout of the AR Application

IV. METHOD

The aim of the present usability evaluation is to collect feedback on an AR application prototype that is tested on different media. The chosen mixed-method approach combines inductive qualitative methods with deductive, quantitatively oriented approaches of data acquisition. In 1993, Nielsen stated that a number of 5-6 test subjects were sufficient to detect significant problems [23]. Since not only the AR application but also the usability of the three media used is to be evaluated, we aimed at a minimum N of 15 persons. According to Faulkner, at least 90% to 97% of all known usability problems can be detected with a number of 15 people [24]. Therefore, we decided on a within-subject design in which every test person performs tests on every medium. The study design, the description of the sample and the used questionnaires will be presented in the following section.

A. *Study Design and Procedure*

In addition to the evaluation of the AR application, the usability of the respective instructional media should also be evaluated. Thus, we have set up a within-subject test design, where the participants have to perform three rounds on the assembly cell (1. Tablet (AR); 2. HoloLens (AR); 3. Touchscreen (non-AR)). Each round was instructed by different instructional media: The AR application is used by two media (the tablet and the HoloLens), so that the evaluation of the AR application can be carried out independently of the medium used. In order to compare these media with previously used media, the touchscreen is also included in the testing. It uses text- and animation-based instructions but is not AR-capable and therefore limited to the dimensionality of the screen. In order to control for repetition and learning effects [25] as far as possible, the order of the instruction media was randomized.

Each participant completed a pre-test questionnaire at the beginning in a paper-pencil format. They were then asked to familiarize themselves with the workstation of the assembly cell. Depending on the randomized condition, the first assembly was instructed by either the tablet, the HoloLens or the touchscreen. The participants had the opportunity to ask the test supervisor for help at any time, but were encouraged to carry out the assembly themselves. After each assembly process, which was completed as soon as the fully assembled gear drive was placed in a box by the robot, there was a post-test questionnaire referring to the medium used. During all three sessions, the subjects were encouraged to express their thoughts aloud. The statements were recorded with a voice recorder. After the third assembly had been completed, participants were asked to fill out the third part of the questionnaire referring to the AR-application itself. The study took about 60 minutes to complete.

B. *Participants*

A total of eight men and seven women took part in the study (N = 15). The mean age of the study participants was 25 years (MW = 25.07, range = 20 - 32). The sample consisted of eleven students and four working persons. Seven participants indicated to have high school graduation and/or the general university entrance qualification as the highest education degree, the remaining eight already have an academic degree (nBachelor = 4, nMaster = 3, nPhD = 1). Twelve participants have never worked with a robot; the other three have rarely worked with a robot. Only one person had already participated in a study on the collaborative assembly cell.

TABLE I. SCALES – INSTRUCTIONAL MEDIA

Scales	***Sources***	***Number of Items***	***Final number and Cronbachs α***	***Example Items***
Task-Load	NASA Task Load Index [29]	6 Items	6 Items Cronbachs α = .68 - .83*	"How much mental and perceptual activity was required? Was the task easy or demanding, simple or complex?"
Perceived Usefulness	Technology Acceptance Model (TAM 3) – Perceived Usefulness [30]	4 Items	4 Items Cronbachs α = .91 - .93	"Using the instruction medium would improve my work performance."
Media Self-Efficacy	TAM 3 – Computer Self-Efficacy [30]	4 Items	2 Items Cronbachs α = .85 - .95	"I would be able to use the instructional medium to do my work if no one were present to tell me what to do."
Perceived Enjoyment	TAM 3 – Perceived Enjoyment [30] Key Components of User Experience (meCue2.0; [28]	3 Items (TAM 3) 3 Items (meCue2.0)	3 Items (TAM 3) 3 Items (meCue2.0) Cronbachs α = .66 - .92	"I would enjoy using the instructional medium.""The instructional medium frustrates me."
Perceived Ease of Use	TAM 3 – Perceived Ease of Use [30] IsoMetrics [22]	4 Items (TAM 3) 2 Items (IsoMetrics)	4 Items (TAM 3) 1 Item (IsoMetrics) Cronbachs α = .56 - .89	"I think the handling of the instructional medium would be clear and understandable for me.""The operating options of the instructional medium support an optimal use of the application."
Open Questions	• What did you particularly like about the instruction medium you used? • What would have to be changed on the instruction medium to make the assembly process even easier? • Please create a ranking of the instruction media, 1 being your strongest preference, 2 being the second choice, etc. Justify your decision.			

* Cronbachs α has been evaluated for three different media and therefore is presented as a range.

C. *Questionnaires*

In the following paragraph, the pre- and post-test questionnaires for both "Instructional Media" and the "AR application" are presented. An overview of all scales with example items can be seen in **TABLE I**.

1) *Pre-Test.*

In addition to the demographic data already reported, the participants were asked about their affinity for technology with five items (e.g., "My enthusiasm for technology is...") on a six-level scale ranging from "very low" to "very high". To complete the data on the participants, we also asked which media (e.g., laptop, smartphone or tablets) are available to them, how often they use them and how easy it is to use the respective medium. In addition, we used the "locus of control for technology" questionnaire (KUT) to assess general control beliefs while dealing with technology [26]. With its eight items (e.g., "Most of the technological problems that I have to face can be solved by myself") on a six-level scale ranging from "not true at all" to "absolutely true" the German questionnaire has a reliability of α = 0.89 [26].

In order to measure the participant's mood before and after the collaborative work process, we decided to use the Affect Grid [27]. The Affect Grid has been designed as a rapid means of evaluating affects in the dimensions of pleasant-unpleasant and arousal-sleepiness. The scale shows reasonable reliability, convergent validity and discriminant validity in studies in which subjects used the Affect Grid to describe their current mood. Since the Affect Grid is particularly suitable for repeated use, the mood was measured after each run with the various media.

2) *Post-Test – Instructional Media.*

The assessment of the usability of the instructional media used is carried out separately from the evaluation of the AR application. Thus, it is possible to separate the findings on software and hardware more clearly. Based on existing usability literature [6]][19][22][28], we decided to select relevant and quantifiable criteria for the task with regard to their face validity in order to determine the suitability of the chosen instructional media.: *a) task load, b) perceived usefulness, c) media self-efficacy, d) perceived enjoyment,* and *e) perceived ease of use.*

a) Task load. The task load was measured by the "NASA Task Load Index (NASA TLX)". It measures subjectively experienced demand using a multidimensional scale that differentiates, for example, between physical and mental strain [29]. The German short version contains six dimensions, namely; mental, physical and temporal demands, as well as performance, effort and frustration. The original scale has 20 gradations from "very low" to "very high". Adapted to the German version, we used a 10-step scale with the poles "little" and "much". Criteria on reliability have been satisfactorily reviewed (Cronbachs α = .68 - .83).

b) Perceived usefulness. The factor perceived usefulness arises from the widespread and empirically well-founded "Technology Acceptance Model (TAM)" [30], which has been incorporated into the development of the usability catalogue. The TAM, currently in its third version, aims at predicting the usage behavior and acceptance of information technologies. To represent the construct, we used four items on a scale from one (strongly disagree) to seven (strongly agree) and adapted them to our application (e.g., “using the instruction medium would improve my work performance”). Cronbach's alpha showed a satisfactory value of α = .91 -.93.

c) Media self-efficacy. Four items from TAM 3's original "Computer Self-Efficacy" scale were used and subsequently adapted (e.g., “I would be able to use the instructional medium to do my work if no one were present to tell me what to do”). Since two of these items - presumably due to a misleading formulation – showed a high standard deviation, they were excluded from further analysis. The remaining two items reached a Cronbach's alpha of α = .85 - .95.

d) Perceived enjoyment. This construct is composed of three adapted items from TAM 3 (perceived enjoyment; e.g., “I would enjoy using the instructional medium.”) and three other items from the “Modular Evaluation of Key Components of User Experience“(meCue2.0; e.g., “The instructional medium frustrates me.”). This questionnaire is based on the analytical “Components of User Experience” model by Thüring and Mahlke [28]. This model distinguishes between the perception of task-related and non-task-related product qualities and includes user emotions as an essential and mediative factor of certain usage consequences. Internal consistency criteria are satisfied for the scale composed in this way (Cronbachs α = .66 - .92).

e) Perceived ease of use. The construct consists of four adapted items from TAM 3 (e.g., "I think the handling of the instructional medium would be clear and understandable for me.") and two further items from the IsoMetrics questionnaire (e.g., "The operating options of the instructional medium support an optimal use of the application."). IsoMetrics was designed for use during the software development process [22]. The focus is set on seven scales, which constitute an operationalization of the seven criteria of the European Committee for Standardization. Here the scale controllability was used to supplement the items from the TAM. Due to its high standard deviation, one item of the IsoMetrics had to be excluded from the analysis. The remaining four items reached a satisfactory internal consistency of Cronbachs α = .56 - .89.

The Post-test on instructional media also contains open questions: “What did you particularly like about the instructional medium you used?”, “What would need to be changed in the instruction medium to make the assembly process even easier?”, and “Please create a ranking of the instructional media, where 1 is your strongest preference, 2 is your second choice, etc. Please give reasons for your decision.”

3) Post-Test – AR application.

The assessment of the usability of the AR application itself was measured by five parameters selected with regard to their fit in terms of early stage evaluation: *(a) perceived usefulness, (b) aesthetic and layout, c) appropriateness of functions,* as well as *d) terminology and terms.* An overview of all scales with example items can be seen in **TABLE II.**

a) Perceived usefulness. To measure perceived usefulness, the same four items were used as in the instructional media post-test. Only the terms were adapted (e.g., "Using the AR application would improve my performance."). Cronbach's alpha showed a satisfactory value of α = .96.

b) Aesthetic and layout. In order to comprehensively depict this construct, four items from the “Visual Aesthetics of Websites Inventory – Short (VisAWI-S)“ were used in the field of aesthetics [31]. The VisAWI-S records how users subjectively perceive the aesthetics of a graphical interface. The used, short version represents the general aesthetic factor [31]. We adjusted the items in terms of terminology (e.g., “Everything matches within the application”) and further added one item from IsoMetrics (“The layout complicates my task processing due to an inconsistent design.”) and another from the “Questionnaire for User Interface Satisfaction (QUIS)”, which was first published in 1987 to ensure feedback on the font as well [32]. This

TABLE II. SCALES – AR APPLICATION

Scales	***Sources***	***Number of Items***	***Final number and Cronbachs α***	***Example Items***
Perceived Usefulness	Technology Acceptance Model (TAM 3) – Perceived Usefulness [30]	4 Items	4 Items Cronbachs α = .96	“Using the AR application would improve my performance.”
Aesthetics and Layout	Visual Aesthetics of Websites Inventory (VisAWI-S; [31] Questionnaire for User Interface Satisfaction (QUIS; [32]	4 Items (VisAWI-S) 1 Item IsoMetrics) 1 Items (QUIS)	4 Items (VisAWI-S) 1 Item (IsoMetrics) 1 Items (QUIS) Cronbachs α = .60	“Everything matches within the application.” “The layout complicates my task processing due to an inconsistent design.”
Appropriate-ness of Functions	IsoMetrics [22]	4 Items	4 Items Cronbachs α = .72	“The information necessary for task processing is always in the right place on the screen.”
Terminology and Terms	QUIS [32] ISONORM [22]	4 Items (QUIS) 2 Items (ISONORM)	4 Items (QUIS) 2 Items (ISONORM) Cronbachs α = .65	“On-screen prompts were confusing.” “Within the AR application, easily understandable terms, descriptions or symbols (e.g., in masks or menus) are used.”
Open Questions	• What did you particularly like about the AR application? • What would have to be changed in the AR application so that the assembly process could be carried out even more easily? • Would you prefer learning via AR classic manuals / manuals? Why?			

composed scale reached an internal consistency of Cronbachs $\alpha = .60$, which is critical for the analysis of this overall scale.

c) Appropriateness of functions. This scale is based on the Task Adequacy Scale of IsoMetrics and with four items (e.g., "The information necessary for task processing is always in the right place on the screen") reached a Cronbach's alpha of $\alpha = .72$.

d) Terminology and terms. To illustrate how understandable the terms and instructions used were, we used four items from QUIS (e.g., "On-screen prompts were confusing.") [32]. Furthermore, the transparency of the robot's activities was queried ("The application always informed me about what the robot does."). Two further items (e.g., “Within the AR application, easily understandable terms, descriptions or symbols (e.g., in masks or menus) are used.”) for this parameter are taken from the Isonorm questionnaire published in 1993 [22]. Like IsoMetrics, Isonorm is based on the criteria of the European Committee for Standardization and therefore uses the same seven factors. This scale reached in total a Cronbachs alpha of $\alpha = .65$.

Similar to the instructional media post-test, the post-test for the AR application also contains open questions: “What did you particularly like about the AR application?”, “What would need to be changed in AR application to make the assembly process even easier?” Finally, the test persons should decide whether and why they would prefer the AR application to traditional manuals.

D. Analysis

The analysis of the collected data was conducted using SPSS. Open questions and recorded comments were analyzed using MAXQDA software. Since this is still a work in progress, the following is a first insight into the results, with a short outlook on qualitative findings. An inferential statistical comparison of the groups is carried out exploratory by subsuming the individual test conditions to the media used. Thus, the comparison groups "tablet", "HoloLens" and "touchscreen" are used for the calculations. Due to the small sample, Friedman's ANOVA [25], used as a non-parametric test procedure, provides an insight into existing group differences, which are further investigated with the help of a post-hoc analysis according to Dunn-Bonferroni [25].

V. RESULTS

Section V gives an overview of the first results for pre- and post-test questionnaires as well as results of the open questions on “Instructional Media” and the “AR Application”.

A. Pre-Test

The participants have a mean technical affinity of 4.61 (min = 3.4; max = 5.60; *SD* = 0.71). General control beliefs while dealing with technology is ranging between min = 3.00 to max = 5.75 (*mean* = 4.73; *SD* = 0.72) within the sample. Media as PC ($n = 7$), Laptops ($n = 11$) and Smartphones ($n = 15$) are used daily by the majority of the test persons, while HoloLens ($n = 12$) and the Oculus Rift (n = 13) are used almost never. Only three participants already used the HoloLens before this study.

The evaluation of the Affect Grid (see **Figure 4**) using a one-way ANOVA, shows that there are significant differences between the measurement points (before and after) in the arousal level ($F(2, 126) = 6.94$, $p = .001$, partial $\eta^2 = .099$) and in the general sensations (F (2, 126) = 30.272, $p < .001$, partial $\eta^2 = .325$) (see **Figure 4**). The diagram shows the comparison of the scales pleasant- unpleasant feelings and sleepiness - arousal before the experiment and after the use of the three different media. The HoloLens shows a tendency towards unpleasant feelings and high arousal, while the use of the touchscreen triggers pleasant feelings and a higher level of sleepiness. However, there are no significant differences between the various conditions (Tablet, HoloLens, and Touchscreen).

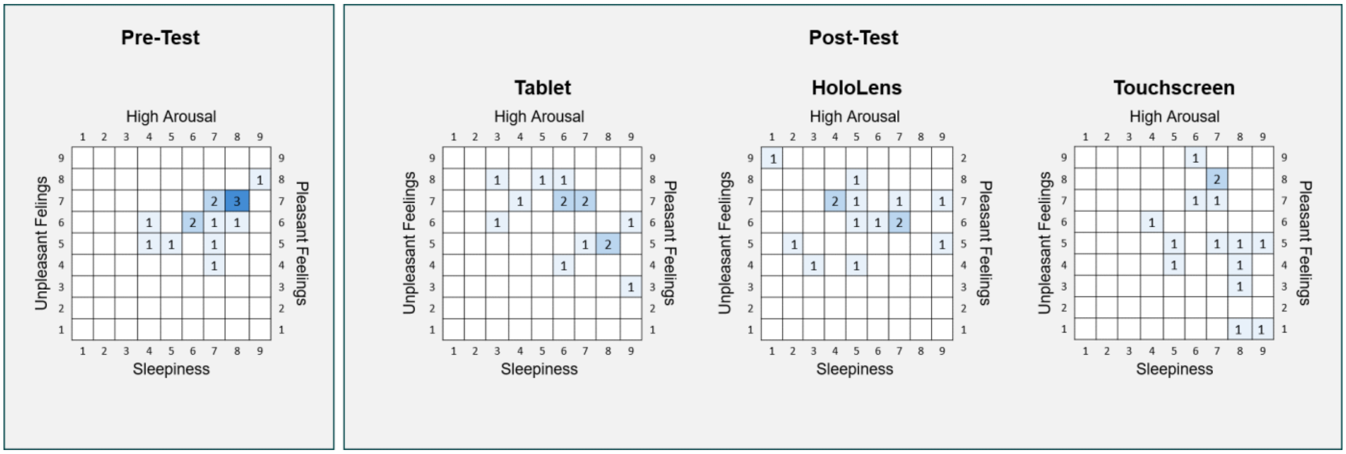


Figure 4. Affect Grid before and after using Instructional Media

B. *Post-Test – Instructional Media*

a) Task-load. Descriptive results of the task load (N = 15; Scale: (0) = “low” to (10) = “high”) can be seen in **Figure 5**, where, for each media, the mean of each scale is shown as a percentages. The mean level of frustration over all tasks and media is 41%, and the highest mean level is reached by the HoloLens with 47%. The lowest mean frustration level of 31% is while using the touchscreen.

The participants stated to achieve their goal on a mean of 67% and the highest performance was achieved using the touchscreen (73%). On average, 40% effort was needed to fulfil the assembly task. The mean temporal demand ranges from 36% (touchscreen) to 42% (HoloLens). The highest mean of physical demand was reported using the tablet (53%), the highest mean of mental demand was reported using the HoloLens (61%).

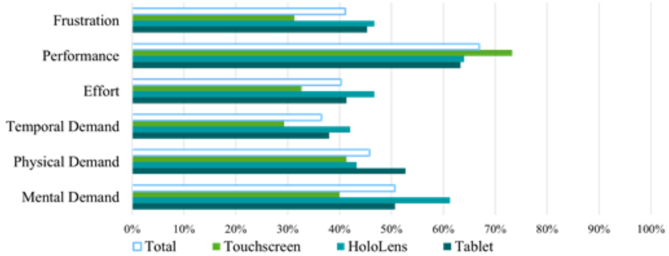


Figure 5. Task Load of Instructional Media

b) Perceived usefulness. All descriptive statistics of the following scales can be seen in **TABLE III**. Within Friedman’s ANOVA, it is always assumed as null hypothesis that there is no difference between the groups. However, the analysis for perceived usefulness shows a statistically significant difference between the groups (x2r (2) = 10.67, p = .005, n = 15). The subsequently performed Dunn-Bonferroni tests with a corrected alpha = .017 show that both the perceived usefulness between tablet and touchscreen differ statistically significantly (z = -2.641, p = .008), as well as the perceived usefulness between HoloLens and touchscreen (z = -2.548, p = .011), indicating that HoloLens and tablet are perceived as less useful than the touchscreen. HoloLens and tablet are not significantly different.

c) Media self-efficacy. Neither mean values nor Friedman's ANOVA show any statistically significant difference between the groups (x^2_r(2) = 4.545, p = .103, n = 15).

d) Perceived Enjoyment. As in the previous scale, neither mean values nor Friedman's ANOVA show a significant difference between the groups with regard to the perceived enjoyment (x^2_r (2) = 2.980, p = .225, n = 14).

e) Perceived ease of use. Both mean values and Friedman’s ANOVA indicate a statistically significant difference between the groups with regard to the perceived ease of use of the media (x^2_r(2) = 21.088, p < .000, n = 15). The subsequently performed Dunn-Bonferroni tests with a corrected alpha = .017 show that both the perceived ease of use between tablet and HoloLens differ statistically significant (z = -2.841, p = .005), as well as the perceived ease of use between HoloLens and touchscreen (z = -3.425, p =.001). Tablet and touchscreen also differ statistically (z = -2.522, p = .012). The results raise an indication that the HoloLens is considered the least easy to use, while the touchscreen reaches the highest value.

TABLE III. DESCRIPTIVE STATISTICS - INSTRUCTIONAL MEDIA

	Descriptive Statistics					
		N	***Mean***	***SD***	***Min.***	***Max.***
Perceived Usefulness	Tablet	15	3.87	.96	2.50	5.75
	Hololens	15	3.85	1.11	2.00	5.50
	Touchscreen	15	4.87	.93	2.25	6.00
Media Self-Efficiacy	Tablet	15	4.80	.95	3.00	6.00
	Hololens	15	4.33	1.51	2.75	6.00
	Touchscreen	15	5.10	.96	3.00	6.00
Perceived Enjoyment	Tablet	14	4.29	.88	2.67	6.00
	Hololens	14	4.38	.66	3.33	5.83
	Touchscreen	14	4.98	.66	3.83	6.00
Perceived Ease-of-Use	Tablet	15	4.60	.60	3.60	5.60
	Hololens	15	3.78	.96	1.80	5.20
	Touchscreen	15	5.17	.59	4.20	6.00

Within the ranking of the instructional tools, the touchscreen was selected as a first choice a total of ten times, after that comes the HoloLens with three times, and lastly the tablet with only two times.

Open questions and comments. The evaluation of the verbal expressions and written comments is done by categorizing them into positive and negative comments for each medium. Individual entries are coded several times. In the following, a brief overview of the most frequently mentioned is given (see **TABLE IV**).

Overall, there are 69 positive comments on the media. 33 of these refer to the touchscreen, which is perceived as easy to use, clearly arranged and overall less restrictive compared to other media. With the tablet (n = 18) it is positively evaluated that the animations can be viewed on demand and from different directions. In combination with markers used, some participants find it easier to orientate themselves at the workplace. The HoloLens is 18 times positively evaluated - the intuitive operation and the hands-free working process are mentioned most frequently. In addition, there are 68 negative remarks, 54 of which are verbal and 14 written comments. 38 of these, are related to the HoloLens due to its lack of wearing comfort and limited vision, e.g., because animations overlay the view of actions to be performed.

There are 28 negative comments about the tablet, mainly related to the perceived difficulty of repositioning the tablet arm, resulting in a limited view of the work surface. The touchscreen has only two negative annotations, namely ‘the fixation does not provide orientation at the workstation’ and ‘animations are not displayed on the work surface’.

TABLE IV. WRITTEN COMMENTS AND VERBAL EXPRESSIONS ON INSTRUCTIONAL MEDIA (AR HARDWARE)

Frequent Categories			Example Statements
Touchscreen	Positive	Ease of Use	“With the touchpad, all necessary information can be grasped most clearly and quickly.”
		No Restrictions (e.g., field of view)	"The field of vision is not restricted, I found that somewhat problematic with the glasses.“ "Especially in comparison to the HoloLens I have the feeling that I am much freer now, because I can operate more easily.”
	Negative	Workplace Orientation	“The fixation does not provide orientation at the workstation” “Animations are not displayed on the work surface”
Tablet	Positive	Support on Demand	“It is more relaxed. I look at the monitor when I need help and can then focus on my task."
		Workplace Orientation	“The markers distinguish between different work areas.” “The component can be viewed from different angles. Green arrows show where the tablet has to be positioned.”
	Negative	Physical Demand	"Tablet must be moved often" "I also need to reach around the tablet's grab arm a little cumbersome here."
		Limited Field of View	“I want to keep looking at the animation, but I also want to work on it at the same time. So I have to use the area on the left side all the time and there is relatively little space to do it.”
HoloLens	Positive	Intuitiveness	“The operation of the HoloLens is intuitive and allows an easy interaction with the environment and reacts automatically to the markers, since no manual positioning is required as it follows the eye.”
		Hands-free Work Process	“There are no movement restriction, the glasses are intuitive” “Free hands and work area”
	Negative	Wearing comfort	"The HoloLens is rather uncomfortable and interferes with vision." "Glasses slip a lot, but if they're tighter, they hurt. It presses on my nose."
		Overlays	"If you have these glasses on and don't look at the animation, but at what you're doing, you don't see it so well. I'd rather look under my glasses".

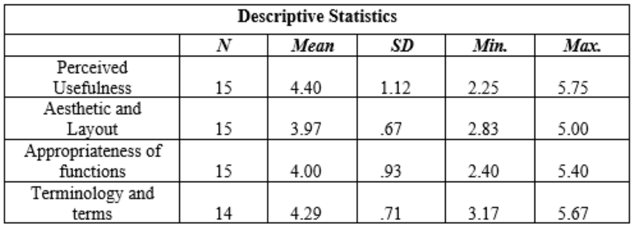
C. *Post-Test – AR Application*

All descriptive statistics of the following scales can be seen in **TABLE V**. The perceived usefulness of the AR application is on a mean of 4.40, which corresponds to an assessment between ‘rather agree’ and ‘agree’. Aesthetics and Layout and Appropriateness of functions (mean = 3.97) corresponds to an assessment of ‘rather agree’. Terminology and terms corresponds to an assessment between ‘rather agree’ and ‘agree’ with a mean of 4.29.

On the question of whether participants would prefer learning via AR to traditional manuals, 13 out of 15 people say they would prefer AR. Main reasons given are, for example, the active learning process, the small steps, the high degree of interaction, the simplicity of use and the perceived fun. In contrast, comments against included the perceived external control of the technology and the possibility of browsing through manuals at one's own pace.

Open questions and comments. There are a total of 85 positive comments on the AR application. The detailed and vividly visualized animations are mentioned particularly frequently here (29 entries) and are accompanied by the clearly perceived instructions (13 entries). The fun and excitement in the process (20 entries) and the active, goal-oriented learning process (7 entries) are also mentioned. In the 182 negatively coded expressions, there are often remarks about the lack of correspondence between reality and displayed animations (e.g., in color, degree of detail, or positioning; 34 entries), such as: "it’s hard to stay focused while the animation continuously moves in the background". In addition, some animations appeared in places not expected by participants or outside the direct field of vision. Since statements were coded multiple times and often referred directly or indirectly to the specific implementation on the respective medium, a further differentiated quantification was not purposeful. An overview of frequent comments can be seen in **TABLE VI**.

TABLE V. DESCRIPTIVE STATISTICS – AR APPLICATION



Descriptive Statistics					
	N	*Mean*	*SD*	*Min.*	*Max.*
Perceived Usefulness	15	4.40	1.12	2.25	5.75
Aesthetic and Layout	15	3.97	.67	2.83	5.00
Appropriateness of functions	15	4.00	.93	2.40	5.40
Terminology and terms	14	4.29	.71	3.17	5.67

VI. DISCUSSION AND OUTLOOK

The study provides insights into the usability and suitability of different media as assistance systems in collaborative assembly. In a within-subject design, two AR-based media (tablet vs. HoloLens) with the same user interface were tested and compared with a classic medium (touchscreen) to instruct the work steps. The results indicate that although the potential of new assistance technologies is recognized, the classic medium is still judged the easiest to use. Furthermore, the study provides results for the separate evaluation of hard- and software usability, using a tailor-made usability catalogue. According to these results, many problems and weaknesses of the technologies are due to the ergonomics and handling of the hardware, whereas the

TABLE VI. WRITTEN COMMENTS AND VERBAL EXPRESSIONS ON AR APPICATION (AR SOFTWARE)

Frequent Categories		**Example Statements**
Positive	Visualization and Localization	“The animations are close to the point where you should perform the next assembly step.”
		"The fact that this is now fully animated makes it even easier to see. So you can really see exactly where the mother has to go now.”
	Clear Instructions and Interactivity	“It is more interactive and you learn to do the work step by yourself. You don't have to look at a whole manual beforehand, but are told step by step what to do.”
		“You learn more because it's easier to memorize it through taking part. It's also fun.”
Negative	Discrepancy between Animation and Reality	"The technology isn't as accurate yet, it doesn't display it perfectly." "I can't see exactly what I'm mounting it on; I'm doing it more intuitively."
	Identification	"If the gears are on top of each other, you can't tell which ones are meant." "You can tell it's the bottom one, but not the one that's on top."
	Localization of Animations	"You don't always know where to look to see the animation." “Because of all the movement, I didn't see that [the animation] was displayed on the far right".

software of the AR application is perceived as very helpful. In general, the methodology and composition of the usability catalogue from inductive and deductive methods was proven to be a successful approach. A high degree of objectivity could be achieved through the questionnaires. The individual results of the scales could be explained in detail by open comments and verbal statements and were thus made comprehensible afterwards. However, the validation of scales in a larger sample should precede further studies.

The results of the evaluation of the instructional media shows that a high level of frustration occurs when processing the task with the HoloLens, which goes hand in hand with a high cognitive demand. Here, possible connections between the ease of use of the media and the perceived cognitive demand could be an interesting starting point for further research. The touchscreen on the contrary causes a low frustration and is evaluated with the highest performance. The tablet's evaluation usually lies between the other media, but shows the highest physical demand. These findings are supported by the assessment of the usability scales, where HoloLens and tablet are rated with a lower usefulness than the touchscreen. In terms of media self-efficacy and perceived enjoyment, there is no difference between the media tested. In future studies, the possible influence of the high technical affinity of the sample on these variables should be clarified. Both the open questions and the ranking support the impression that the touchscreen convinces users with its simple operation. In addition, it becomes clear that the innovative character of the HoloLens is perceived as enjoyable. Above all, the restricted field of vision and the overlapping of animations with reality still poses a problem, which can be prevented, for example, by using the tablet and moving the holder.

In the evaluation of the AR Application, it becomes clear that AR is generally assessed with a relatively high usefulness. This finding is also supported by the open comments. Comparing only the number of positive and negative statements, the impression could arise that the AR application has been increasingly perceived as negative. The high number of negative comments could be mainly due to the nature of the task. During the thinking aloud method, the test persons are explicitly asked to express all thoughts and ideas. Thereby, it is obvious that the participant first deals with deficits and usability problems - but the reflection and evaluation after the task shows a positive assessment of the application. Many of the comments also contain ideas and suggestions for improvement in order to improve the use of the media.

Especially the clear and small step instructions are perceived as useful. The aesthetics and layout of the application, as well as the appropriateness of functions should be worked on in the further course. It should be considered that images and animations are perceived as helpful, whereas text descriptions are sometimes described as obstructive or misleading.

With regard to the suitability of the assistance technologies presented for the collaborative working scenario, it can be stated that by focusing on the task and the respective medium, the test persons hardly paid attention to the robot and also scarcely made any comments on it. It is problematic that this was not explicitly measured in the questionnaire. The perception of the robot should definitely be investigated more closely in future studies and brought into focus.

The analysis of the Affect Grid shows that, although there is a difference in the level of excitement and pleasure before and after the test, there is no significant difference between the media. However, the descriptive evaluation shows that participants feel increasingly stressed by the HoloLens (axis unpleasant - aroused), whereas the touchscreen shows more frequent entries towards relaxation (axis pleasant - sleepy).

A main restriction of the study refers to the high academic degree and the young average age of the sample, which is accompanied by a comparatively high affinity for technology. In addition, almost all participants are novices in the field of assembly, which severely limits the transferability of study results. Another limitation refers to the laboratory setting of the study, where no real working conditions (e.g., lighting, noises) occur. The results achieved by such a small number of participants should not be interpreted without caution. Conclusions on the quality of the questionnaire used should not yet be derived, as this requires a larger sample. The items and constructs used here were selected based on the specific use case, which limits the transferability of this selection. Further, we only used already established questionnaires and the thinking aloud method and did not include any other usability instruments (e.g., usability cards).

This paper provides deeper insights into the earlier usability evaluation of the same use case [1]. Special focus was paid to the detailed analysis of the task process in order to align the development of the AR application directly with the use case. This certainly limits the transferability of the results to other use cases. In the meantime, another study has been conducted to test Head Mounted Displays in the field, which has revealed similar results [33]. After the revision of the detected usability problems, following studies should be carried out referring to the effectiveness of the use of AR. Hereby, the influence of AR on the general performance, error rate and satisfaction with the work process should be examined. Furthermore, the influence of AR on the acceptance of robots should be further investigated within the context of collaborative workspaces.

ACKNOWLEDGMENT

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L14Z000) and implemented by the Project Management Agency Karlsruhe (PTKA). The author is responsible for the contents of this publication.

REFERENCES

[1] L. Daling, A. Abdelrazeq, M. Haberstroh, and F. Hees, "Usability Evaluation of Augmented Reality as Instructional Tool in Collaborative Assembly Cells," The Twelfth International Conference on Advances in Computer-Human Interactions (ACHI 2019) IARIA, Feb. 2013, pp. 199-205, ISBN: 978-1-61208-686-6.

[2] Acatech - Deutsche Akademie der Technikwissenschaften in Kooperation mit Fraunhofer IML, and equeo, Ed., "*Kompetenzentwicklungsstudie Industrie 4.0: Erste Ergebnisse und Schlussfolgerungen (Industry 4.0 Competence Development Study: First Results and Conclusions),*" München, 2016.

[3] Q. Guo, "Learning in a Mixed Reality System in the Context of Industry 4.0, "*Journal of Technical Education*, vol. 3, no. 2, pp. 92–115, 2015, ISSN 2198-0306.

[4] W. Bauer, M. Bender, M. Braun, P. Rally, and O. Scholtz, "Leichtbauroboter in der manuellen Montage - einfach einfach anfangen: Erste Erfahrungen von Anwenderunternehmen," Frauenhofer IOA, 2016.

[5] S. L. Müller, S. Stiehm, S. Jeschke and A. Richert, "Subjective Stress in Hybrid Collaboration", in vol. 10652, *Social Robotics*, A. Kheddar et al., Eds., Cham: Springer International Publishing, pp. 597–606, 2017, doi: 10.1007/978-3-319-70022-9.

[6] F. Sarodnick and H. Brau, "*Methoden der Usability Evaluation (Methods of Usability Evaluation),*" (2nd. ed). H. Huber, Hogrefe AG, Bern, 2006.

[7] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier and B. MacIntyre, "Recent Advances in Augmented Reality", Naval Research Lab, Washington DC, pp. 34-47, 2001, doi: 10.1109/38.963459.

[8] B. Furht, Ed., "Handbook of Augmented Reality," Springer Science & Business Media, 2011, doi: 10.1007/978-1-4614-0064-6.

[9] A. Y. Nee, S. K. Ong, G. Chryssolouris and D. Mourtzis, "Augmented Reality Applications in Design and Manufacturing", *CIRP Annals-manufacturing technology*, vol. 61(2), pp. 657-679, 2012, doi: 10.1016/j.cirp.2012.05.010.

[10] A. Dey, M. Billinghurst, R.W. Lindeman, and J.E. Swan, "A systematic review of 10 years of augmented reality usability studies: 2005 to 2014," Frontiers in Robotics and AI vol. 5:37, 2018, doi: 10.3389/frobt.2018.00037.

[11] A. Little, "*How Automotive Manufacturers are Utilizing Augmented Reality,*" August 2018, retrieved January, 08, 2019 from:https://www.manufacturingtomorrow.com/article/2018/03/how-automotive-manufacturers-are-utilizing-augmented-reality-/11117.

[12] S. Werrlich, D. Austino, A. Ginger, P.-A. Nguyen, and G. Notni, "Comparing HMD-based and paper-based training," Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE, Munich, Oct. 2018, pp. 134–142, doi: 10.1109/ismar.2018.00046.

[13] M. Akcayır and G. Akcayır, "Advantages and Challenges Associated with Augmented Reality for Education: A systematic review of the literature," *Educational Research Review*, vol. 20, pp. 1-11, 2017, doi:10.1016/ j.edurev.2016.11.002.

[14] X. Wang, S.K. Ong, A.Y.C. Nee, "A comprehensive survey of augmented reality assembly research," Adv. Manuf., vol. 4 (1), pp. 1–22, 2016, doi: 10.1007/s40436-015-0131-4.

[15] S. Makris, P. Karagiannis, S. Koukas, and A. S. Matthaiakis, "Augmented reality system for operator support in human–robot collaborative assembly," CIRP Annals, vol. 65(1), pp. 61-64, 2016, doi: 10.1016/j.cirp.2016.04.038.

[16] G. Michalos, P. Karagiannis, S. Makris, Ö. Tokcalar, and G. Chryssolouris, "Augmented reality (AR) applications for supporting human-robot interactive cooperation," Procedia CIRP, vol. 41, pp. 370-375, 2016, doi: 10.1016/j.procir.2015.12.005.

[17] A. Tang, C. Owen, F. Biocca and W. Mou, "Comparative effectiveness of augmented reality in object assembly," *In Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 73-80, ACM, April 2003, doi:10.1145/642611.642626.

[18] A. Ullrich and G. Vladova, "Qualifizierungsmanagement in der vernetzten Produktion - Ein Ansatz zur Strukturierung relevanter Parameter (Qualification Management in Networked Production - An Approach to Structuring Relevant Parameters)," *Lehren und Lernen für die moderne Arbeitswelt*, pp. 58–80, GITO, 2015.

[19] J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 152-158, ACM. 1994, doi: 10.1145/191666.191729.

[20] L. Gabbard, J. E. Swan II, D. Hix, S,-J. Kim and G. Fitch, "Active Text Drawing Styles for Outdoor Augmented Reality: a User-based Study and Design Implications," in *Virtual Reality Conference VR'07, IEEE*, pp. 35-42, 2007, doi: 10.1109/VR.2007.352461.

[21] S. L. Müller-Abdelrazeq, K. Schönefeld, M. Haberstroh, and F. Hees, "Interacting with Collaborative Robots - A Study on Attitudes and Acceptance in Industrial Contexts," in Social Robots: Technological, Societal and Ethical Aspects of Human-Robot Interaction, O. Korn, Ed. Springer International Publishing, pp. 101-117, 2019.

[22] K. Figl, "Deutschsprachige Fragebögen zur Usability-Evaluation im Vergleich (German-language questionnaires for usability evaluation in comparison)," *Zeitschrift für Arbeitswissenschaft*, *4*, pp. 321-337, 2010.

[23] J. Nielsen, "Usability Engineering," London: AP Professional Ltd., 1993.

[24] L. Faulkner, "Beyond the five-user assumption: Benefits of increased sample sizes in usability testing," *Behavior Research Methods, Instruments, and Computers,* 35 (3), pp. 379-383, 2003.

[25] A. Field and G. Hole, "How to design and report experiments," Sage, 2002.

[26] G. Beier, "Kontrollüberzeugungen im Umgang mit Technik: ein Persönlichkeitsmerkmal mit Relevanz für die Gestaltung technischer Systeme (Control Control beliefs in dealing with technology: a personality trait with relevance for the design of technical systems)," *Ph.D. Dissertation*, Humboldt-Universität zu Berlin. Available from GESIS database, Record No. 20040112708.

[27] J. A. Russel, A. Weiss and G. A. Mendelsohn, "Affect grid: a single-item scale of pleasure and arousal", *Journal of personality and social psychology,* 57 (3), pp 493-502, 1989.

[28] M. Thüring and S. Mahlke, "Usability, Aesthetics and Emotions in Human–Technology Interaction", *International Journal of Psychology*, *42*(4), pp. 253-264, 2007, doi: 10.1080/00207590701396674.

[29] S. G. Hart, "NASA-Task Load Index (NASA-TLX); 20 Years Later," *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*, Santa Monica: HFES, pp. 904-908, 2006, doi: 10.1177/154193120605000909.

[30] F. D. Davis, "A Technology Acceptance Model for Empirically Testing New End-user Information Systems: Theory and Results," *Ph.D. Dissertation,* Massachusetts Institute of Technology, 1985.

[31] M. Moshagen and M. T. Thielsch, "A Short Version of the Visual Aesthetics of Websites Inventory", *Behaviour & Information Technology*, 32 (12), pp. 1305-1311, 2013, doi: 10.1080/0144929X.2012.694910.

[32] J. P. Chin, V. A. Diehl, and K. L. Norman, "Development of a Tool Measuring User Satisfaction of the Human-Computer Interface", in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 213-218. ACM, 2018.

[33] L. Daling, A. Abdelrazeq, C. Sauerborn, & F. Hees."A Comparative Study of Augmented Reality Assistant Tools in Assembly"; in *International Conference on Applied Human Factors and Ergonomics* (AHFE 2019) Springer, Cham., July 2019, pp. 755-767.