Usability Evaluation of Augmented Reality as Instructional Tool in Collaborative Assembly Cells

Lea M. Daling, Anas Abdelrazeq, Max Haberstroh, and Frank Hees
Chair of Information Management in Mechanical Engineering (IMA)
RWTH Aachen University
Aachen, Germany
lea.daling@ima-ifu.rwth-aachen.de

***Abstract*—The increasing digitalization and customization of the production sector, which is commonly referred to as Industry 4.0, poses new challenges for the qualification of employees. The integration of Augmented Reality (AR) as an instructional tool provides new opportunities to support learning processes close to the workplace. Even though the use of this technology seems promising, there is still little empirically founded knowledge about the performance and fit of the system for collaborative assembly processes. This paper presents an empirical approach for the usability evaluation of a developed AR application, which can be used to assess software and hardware factors separately. In a pre-study, this catalogue was tested in combination with an experimental study design. First results for the evaluation of the developed AR solution and the suitability of different media are presented.**

Keywords-Usability Criteria; Augmented Reality; On the Job Training; Industry 4.0; Human-Robot-Collaboration (key words).

I. INTRODUCTION

The distribution of digitalization and networking within the field of 'Industry 4.0' is associated with increasingly individualized and highly flexible production [1]. Thus, fast and efficient training processes will be necessary to prepare both experienced and temporary workers for the respective assembly processes. This becomes particularly important when assembly processes are neither completely manual nor fully automated and take place in cooperation between human and robot [2].

In order to increase the efficiency of training processes, the integration of new technology and digital learning formats is needed to guide employees step by step through assembling processes and to train them flexibly for new use cases. The use of Augmented Reality (AR) as instructional assistant tool is widely expected to be a success factor of digital training programs [3]. However, its adequacy is not yet sufficiently empirically proven [4] .

Since a high degree of usability can be seen as a prerequisite for further performance measures such as increasing effectiveness and reducing the error rate, the aim of our current research is to test the usability of AR in an Industry 4.0 use case [5]. Thus, the presented AR application is designed for the instruction of a collaborative assembly process between human and robot. In Section 2, a brief introduction of the basic functions and application possibilities of AR in the manufacturing context are given. Furthermore, the use case of AR as an on the job instructional tool in a collaborative assembly cell is presented. In Section 3, the methodological approach is presented. We present an overview of relevant usability criteria and our empirical approach to measure usability of an AR application using different instructional media. Finally, section 4 gives an overview of the first 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. 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"[6]. It can therefore be said that AR systems overlay computer generated objects onto a real world setting, in real time [7].

Within the last 10-15 years, AR systems have shown great improvement and an ability to create solutions to various problems [8]. Using AR, for example, innovative and effective methods can be developed to answer important needs 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 [9]. Through such innovative uses, one can minimize the need for improvement iterations, re-works and modifications by 'getting it done the right way' from the start. The use of AR in the current state promises many positive effects, such as constant access to information, lower error occurrences, improved motivation and a synchronized training and performance [10]. 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 reduces the assembly errors significantly [10].

Nevertheless, AR systems still face a couple of challenges that prevent the direct implementation of AR solutions in real world problems. The current status, e.g., in display and tracking technology, as well as calibration techniques, still faces many difficulties [10]. Even with those challenges conquered, other questions still arise, like whether or not the implementation of such systems would lead to

 ISBN: 978-1-61208-686-6

other problems affecting the overall performance. An over-reliance on the AR generated signals and indications can for example have negative implications on the performance of the user, by disrupting the attention or focusing it all in one direction, leading it away from the surrounding context [11]. Further research and evaluation of the technology is therefore necessary to solve existing problems and expand the spectrum of applications. In the following, an exemplary use case is presented in which AR is to be used as an on the job instructional tool for collaborative assembly processes.

B. AR as On the Job Instructional Tool

Since AR is used to add real-time information to a real (working) environment, we expect it to enable "learning on demand" in an on the job training session. 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 [12]. Further, they provide tailor-made learning processes and can be used time- and learning pace-independent [3]. The use case in which AR is used to enable on the job training consists of a collaborative assembly cell equipped with a robot.

During the assembly process, man and robot are working together to assemble a small gear drive. In collaboration with the UR-5, the participants put together three plates with gear wheels. The human operator performs five steps, while the robot performs a total of four steps. Once the participant has familiarized himself with the cell, he is instructed to position a base plate and rear plate in a holder. The robot inserts four hexagon socket screws and positions the back plate on the base plate, while the human operator assembles two sets of gear wheels. In the final step, these are mounted on the pre-assembled base plate presented by the robot. Previously, the assembly task was guided by a fixed touchscreen with 3-D animations. The use of AR offers the added value of displaying information and work steps directly at the workspace or at the tool required for the respective assembly step. The instructions for the AR application were developed on the basis of the existing work steps and supplemented by virtual objects with real-time animations. Based on fundamental usability heuristics (e.g., visibility of the system status, consistency, aesthetics) [13], we decided to use a minimalistic design, small work steps and an avatar to guide the test persons through the assembly process, which can be seen in Figure 1.

Figure 1. Layout of the AR Application.

However, before the effectiveness of AR and its appropriateness for the use case of human robot collaboration can be tested in an experimental setting collecting quantitative data, the present qualitative pre-study aims at deriving basic implications on the usability of the developed AR application. The evaluation is intended to not only provide feedback on the AR-capable hardware, but also on the AR application software itself. Thus, two AR-capable see-through devices (tablet and Microsoft HoloLens) are used as hardware to test the AR application (software). Using the Microsoft HoloLens as HMD, we intended to enable a hands-free assembly process. Figure 2 shows that the use of the tablet is supported by a desk mount tablet arm. The study design explained below aims to derive implications of the usability of both the respective media and the AR application. For this purpose, relevant usability criteria are evaluated, which are explained in more detail below.

Figure 2. AR as Instrcutional Tool for HMD and Tablet.

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 [5]. Sarodnick and Brau emphasize that usability particularly considers the fit of system, task and user, taking into account the quality of goal achievement perceived by the user [5]. For this reason, it is essential to involve potential users in the evaluation process at an early stage.

Within a survey of Gabbard and colleagues [14] , it was found 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.

The most widely used *inductive* approach, which is characterized by the analysis of early versions and prototypes, may be the so called thinking aloud method [5]. 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 double load of task processing and loud thinking reduces the processing speed. For this reason, this method should not be used in conjunction with a performance measurement. Furthermore, the test conditions in the execution are often few standardized.

Deductive methods, on the other hand, capture the user's perspective on an already developed product. At this point,

 ISBN: 978-1-61208-686-6

however, changes and corrections of a system are often time- and cost-consuming. Established evaluation concepts (e.g., IsoMetrics; Isonorm [15]), often make use of the classical questionnaire methodology, which ensures the fulfillment 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. The thinking aloud method was further used to verbalize and record the impressions, reactions and cognitions of the participants during the work process. The composition of these approaches will be presented in the following.

III. 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, quantiatively oriented approaches of data acquisition. In 1993, Nielsen stated that a number of 5-6 test subjects were sufficient to detect significant problems [16]. 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 [17]. 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. 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 [18] 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 has 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 8 men and 7 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 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.

C. Questionnaires

In the following paragraph, the pre- and post-test questionnaires for both “Instructional Media” and the “AR application” are presented.

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, 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 [19]. 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 $\alpha = 0.89$ [19].

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 [5]][13][15][20], 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 [21]. 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)"[22], 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 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 [20]. 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 [15]. 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.*

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 [23]. The VisAWI-S records how users subjectively perceive the aesthetics of graphical interface. The used short version represents the general aesthetic factor [23]. 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 [24]. This composed scale reached an internal consistency of Cronbachs α = .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") reaches a Cronbach's alpha of α = .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.")[24]. 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 [15]. 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 α = .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.

 ISBN: 978-1-61208-686-6

D. Analysis

The analysis of the collected data was conducted using SPSS. Open questions and the analysis of recorded comments was done using MAXQDA software. Since this is still 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 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 [18] 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 [18].

IV. First Results

Section 4 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.

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 3, where means of each scale are shown as percentages for each media. 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 is 31% while using the touchscreen. The participants reported 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 fulfill 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%).

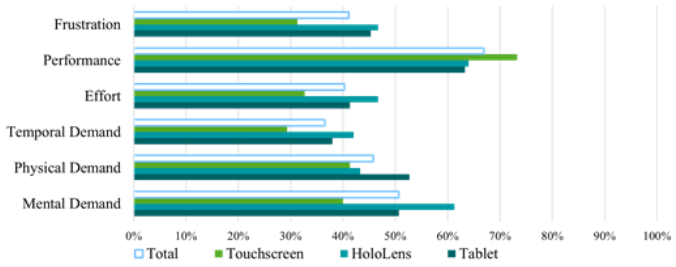

Figure 3. Task Load of Insstructional Media.

b) Perceived usefulness. All descriptive statistics of the following scales can be seen in Table. I. 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 (x^2_r (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 as 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 significantly (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 significantly (z = -2.522, p = .012). The results raise an indication that the HoloLens is considered to be the least easy to use, while the touchscreen reaches its highest value.

TABLE I. 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-Efficacy	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 chosen ten times as a first choice, the

 ISBN: 978-1-61208-686-6

HoloLens three times, and the tablet two times as a first choice.

The evaluation of the verbal expressions and written comments was done by categorizing them into positive and negative comments for each medium. Individual entries were coded several times. In the following, a brief overview of the most frequently mentioned is given.

Overall, there were 69 positive comments on the media. 33 of these referred to the touchscreen, which was perceived as easy to use and clearly arranged. With the tablet (n = 18) it was positively evaluated that the animations can be viewed on demand. The HoloLens was 18 times positively evaluated - most frequently the intuitive operation and the innovative character were mentioned. In addition, there were 68 negative remarks, 54 of which were verbal and 14 written comments. 38 of these were related to the HoloLens and the lack of wearing comfort or limited vision, e.g., "The HoloLens impairs vision". There were 28 negative comments about the tablet, mainly relating to the difficult positioning of the tablet arm and the resulting limited view of the work surface. The touchscreen had only two negative annotations, namely 'the fixation does not provide orientation at the workstation' and 'animations are not displayed on the work surface'.

C. Post-Test – AR Application

All descriptive statistics of the following scales can be seen in Table. II. 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.

TABLE II. 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

On the question of whether participants would prefer learning via AR to traditional manuals, 13 out of 15 people said they would prefer AR. Main reasons given were, 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 were a total of 85 positive comments on the AR application. The detailed and vividly visualized animations were mentioned particularly frequently here (29 entries) and are accompanied by the clearly perceived instructions (13 entries). The fun and excitement (20 entries) in the process and the active, goal-oriented learning process (7 entries) were 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 moves continuously in the back". In addition, the text instructions were sometimes perceived as cumbersome: "I find it annoying that I always have to read so much text".

V. DISCUSSION AND OUTLOOK

The study reveals first results for the separate evaluation of the usability of different instruction media as well as the developed AR application using a tailor-made usability catalogue. In general, the 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. Here it becomes apparent that 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 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, which is also supported by the open comments. This shows that 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.

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

 ISBN: 978-1-61208-686-6

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 on the basis of the specific use case which limits the transferability of this selection. Further, we only used already established questionnaires and the thinking aloud method but did not include any other usability instruments (e.g., usability cards).

The first impressions of the study should be examined in a real use case within further research. In order to generate a better transferability of the results, a survey of assembly workers is planned for the further course of our investigations. After the revision of the detected usability problems, following studies will 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 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] acatech - Deutsche Akademie der Technikwissenschaften in Kooperation mit Fraunhofer IML & equeo, Ed., *Kompetenzentwicklungsstudie Industrie 4.0: Erste Ergebnisse und Schlussfolgerungen (Industry 4.0 Competence Development Study: First Results and Conclusions).* München, 2016.

[2] 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_59.

[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.

[4] M. Funk, T. Kosch and A. Schmidt, „Interactive worker assistance", in Proceedings of the 2016 ACM, 2016 UbiComp, pp. 934–939. doi: 10.1145/2971648.2971706.

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

[6] 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.

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

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

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

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

[11] 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.

[12] 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.

[13] 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

[14] 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.

[15] 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.

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

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

 ISBN: 978-1-61208-686-6