Comparing Conventional and Augmented Reality Instructions for Manual Assembly Tasks

Jonas Blattgerste
Bielefeld University
Universitätsstr. 25
33615 Bielefeld, Germany
jblattgerste@techfak.uni-bielefeld.de

Benjamin Strenge
CITEC - Cluster of Excellence
Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
benjamin.strenge@uni-bielefeld.de

Patrick Renner
CITEC - Cluster of Excellence
Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
prenner@techfak.uni-bielefeld.de

Thies Pfeiffer
CITEC - Cluster of Excellence
Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
tpfeiffer@techfak.uni-bielefeld.de

Kai Essig
CITEC - Cluster of Excellence
Cognitive Interaction Technology
Bielefeld University
Inspiration 1
33619 Bielefeld, Germany
kessig@techfak.uni-bielefeld.de

ABSTRACT

Augmented Reality (AR) gains increased attention as a means to provide assistance for different human activities. Hereby the suitability of AR does not only depend on the respective task, but also to a high degree on the respective device. In a standardized assembly task, we tested AR-based *in-situ* assistance against conventional pictorial instructions using a smartphone, Microsoft HoloLens and Epson Moverio BT-200 smart glasses as well as paper-based instructions. Participants solved the task fastest using the paper instructions, but made less errors with AR assistance on the Microsoft HoloLens smart glasses than with any other system. Methodically we propose operational definitions of time segments and other optimizations for standardized benchmarking of AR assembly instructions.

CCS Concepts

•**Human-centered computing** → **Human computer interaction (HCI);**

Keywords

Assistance Systems; Head-Mounted Displays; Smartglasses; Benchmarking

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

PETRA '17 June 21–23, 2017, Rhodes, Greece

© 2017 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ISBN 978-1-4503-5227-7/17/06…$15.00

DOI: http://dx.doi.org/10.1145/3056540.3056547

1. INTRODUCTION

Due to increasing scientific and commercial interest and frequent releases of new devices, AR also comes into focus regarding its applicability in assistive systems. The most frequently used AR-capable devices include smartphones and smart glasses, also known as optical see-through Head-Mounted Displays (HMDs), but other approaches have been developed as well. For example, projectors have been used as a means for AR presentations, e.g. mounted above the workplace [5, 3] or worn with a helmet [4].

Each of these approaches has different advantages and drawbacks. Smartphones are readily available to almost everyone, but must be held with (at least) one hand, thus only one hand is available for physical interaction with the environment, e.g. solving a manual task. Today's smart glasses have a limited field of view (usually below 30° while humans have a horizontal field of view of almost 180°) and often questionable aesthetic properties. Projectors are comparatively heavy, lack mobility and may raise privacy concerns because everyone around can see projections possibly meant for a specific user only. Because of these very different characteristics, the suitability of AR techniques for supporting a given task highly depends on the employed devices.

As a benchmark for AR instructions in manual assembly, a standardized task using Lego Duplo bricks has been proposed [2]. Funk et al. [3] utilized this task to compare in-situ projection with paper-, HMD- and tablet-based instructions. Their results led the authors to believe that *"HMD instructions have problems being accepted by workers"* and that *"locating a part is significantly faster using in-situ projection and paper-based instructions, [...] locating assembly positions is significantly slower using HMD instructions compared to tablet and paper instructions, and assembling is significantly faster using in-situ projection compared to*

HMD. [...] Further, participants made significantly fewer errors using the tablet and in-situ instructions compared to the HMD instructions. Moreover, the perceived cognitive load using the NASA-TLX questionnaire is significantly lower for the in-situ instructions compared to the HMD instructions." However, in this study the HMD (an Epson Moverio BT-200) was not actually used as an AR device in the narrower sense, but rather to show a still image instruction in the center of participants' field of view. Therefore we do not think that this serves as proof of general unserviceableness of the smart glasses technology for assembly assistance, but rather indicates that the particular implementation and device caused unfavorable results.

In order to provide evidence for this, we used the same assembly task to evaluate the performance of our own implementation on Microsoft HoloLens smart glasses against three other instruction systems. For the HoloLens system we used AR-based in-situ visualizations conceptually comparable to the in-situ projection introduced by Funk et al. [3] which were able to show the location of the brick container for the current step and the correct postion for assembling this brick. Hereby we adjusted the visualizations to be in line with the capabilities of smart glasses. In contrast to the two-dimensional in-situ projections by Funk et al. [3], our implementation showed three-dimensional virtual objects to indicate the correct assembly position, and instead of highlighting a container with a green light we placed a crosshair on it. The same visualizations were also used on a smartphone to assess the influence of the chosen device. In addition to these two newly-created implements we reassessed two approaches previously used by Funk et al. [3]: The in-view display of pictorial instructions using the Epson Moverio HMD, and paper-based instructions. Concerning methodological advancement we applied a more thorough analysis for the AR instruction benchmark data.

2. RELATED WORK

Research regarding instructions on AR systems is not a particularly new endeavor, and especially AR at workspaces has been researched for quite a while by now. In 1992 Caudell et al. [1] used AR to give workers instructions for a subtask of building an airplane. E.g., they used AR to project needed boreholes onto an unattached wing of an airplane. This was done by showing the instructions on an HMD. Henderson et al. [8] used HMDs to support maintenance staff for military vehicles by projecting step-by-step instructions with info boxes and three-dimensional arrows for specific maintenance tasks that had to be done regularly. Gauglitz et al. [6] used AR instructions on a tablet computer to give instructions in airplane cockpits.

In the specific context of assembly tasks, displaying instructions using AR smart glasses should reduce users' head movements as information are directly shown in their field-of-view. Tang et al. [15] have shown that AR instructions reduced errors and cognitive load of participants. However, they also found that occlusions of target objects by AR content or presenting information over a cluttered background can decrease task performance. Petersen et al. [10] projected video overlays into the environment at the correct position and time using a piecewise homographic transform. By displaying a color overlay of the user's hands, feedback can be given without occluding task-relevant objects.

The majority of assembly tasks include a picking component, as the necessary parts have to be located firstly. Thus, one critical component is guiding the user to these parts. Schwerdtfeger and Klinker [14] compared different visualizations to give positional and directional guidance to the target. They found that a frame (serving as positional information) in combination with an opaque tunnel or an arrow (as directional information) performed best. In our own previous work, we compared a number of different guidance techniques in simulated AR, making use of the user's eye gaze to give more specific feedback [11].

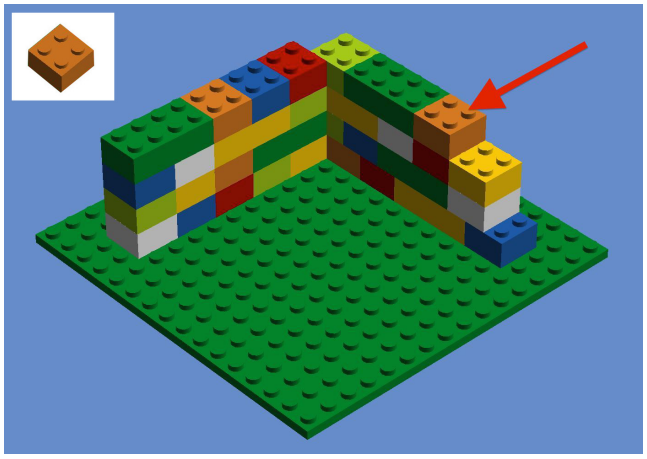

Figure 1: The 32nd step of the paper instructions introduced by Funk et al. [2]

In order to avoid the drawbacks of AR glasses, projection-based approaches are a broadly used alternative. E.g., Sand et al. [13] projected instructions into the physical workspace of a user, which enabled them to assemble products without prior knowledge. Rodriguez et al. [12] proposed a solution in which instructions are directly overlaid with the real world using projection mapping.

In the context of this paper, the AR benchmark introduced by Funk et al. [2] is of particular importance. This benchmark is supposed to allow for a standardized evaluation of different AR instruction systems. To achieve this, they proposed the General Assembly Task Model (GATM), which separates each step of an assembly task into four different phases and categorized these into two task-dependent and two task-independent phases. They also introduced two easily reproducible assembly tasks, one in the professional context (assembly of nuts and washers) and a more abstract task using Lego Duplo bricks. They provided paper instructions as a reference for assembling constructions with 4, 8, 16 and 32 steps (cf. Figure 1). In addition to assembly times and errors, the authors suggested using the NASA Task-Load Index questionnaire [7] to measure the perceived cognitive load of participants. The benchmark that resulted from those components was then used to compare multiple instruction systems: The paper instructions, an in-view implementation on an HMD, a tablet, and an in-situ projection [3]. While the in-situ method projected the instructions onto the environment in real time, the other methods only displayed the still images that were also printed as paper instructions. They came to the conclusion that the in-situ projection they developed outperformed the other implementations.

Figure 2: In-situ assembly assistance and picking location marker as used in our study

Figure 3: HoloLens smart glasses providing in-situ AR assistance

Similarly, Khuong et al. [9] also used a Lego Duplo assembly task to evaluate instructions on HMDs. In contrast to [3], they did not use optical see-through but video see-through HMDs. Moreover, they also made their implementation context-aware, meaning it was able to automatically detect if the user attached the current step and activate the next step of the task as well as to automatically detect errors. They found out that users tended to prefer instructions to not be directly projected onto the actual assembly situation but peripherally beside the target region, thus not occluding relevant objects.

3. INSTRUCTION TECHNIQUES

We compared four different techniques for presenting instructions. Each technique was implemented on the most promising device at our disposal: A Microsoft HoloLens, an Epson Moverio BT-200, a smartphone, and the paper instructions. In general, the implementations can be split into two categories: The in-situ implementations where the instructions were displayed in 3D directly at the respective target positions, and in-view implementations where two-dimensional pictorial instructions were displayed in the field of view.

Figure 4: Smartphone providing in-situ AR assistance

3.1 In-situ instructions

The in-situ AR assistance was implemented for Android devices (like smartphones in Figure 4 and the Epson Moverio) and the Microsoft HoloLens (see Figure 3). In each step a simple cuboid with size and color corresponding to the Lego Duplo brick that had to be assembled was displayed at the correct assembly position (see Figure 2). We suspected that using a more detailed 3D model of an actual Lego Duplo brick would not only have degraded system performance, but could possibly have demanded more cognitive ressources of users by adding unnecessary clutter. We displayed a white crosshair in each step to mark the bin containing the required type of bricks (see Figure 2). We decided to use this instead of just highlighting the container by color to increase the visibility as it can be challenging to distinguish certain colours in certain enviroments on optical see-through AR devices like the Microsoft HoloLens and Epson Moverio.

A non-representative preliminary study comparing all instruction techniques suggested that our in-situ implementation worked better (i.e. more stable) on the HoloLens than on the Moverio BT-200. Therefore in this study we used the HoloLens smart glasses and a smartphone to provide in-situ assistance.

3.2 In-view instructions (Epson Moverio BT-200)

Funk et al. [3] already introduced a two-dimensional in-view instruction technique for the Epson Moverio BT-200 (see Figure 5) that displayed the images of the paper instructions (cf. Figure 1) into the field of view of the user. We re-implemented this based on their description. Funk et al. connected the HMD via Wi-Fi and let the examiner activate each new step for the participant ("Wizard of Oz"). As this might introduce a (negative) Rosenthal effect and bias results, we decided to let participants control the task on their own by pressing a button to advance to the next instruction. In doing so, we also established more comparable conditions, because this way task progress was both controlled by participants when using paper instructions (by turning pages) and in-situ assistance (by pressing a button to advance).

4. METHODOLOGY

Analogous to [3], our study followed a within-subjects design. The independent variable was the instruction system

Figure 5: Moverio BT-200 smart glasses showing central in-view pictorial instructions

(four levels). The dependent variables were the number of errors a participant made, the NASA Task-Load Index RTLX scores [7] for measuring the cognitive load, and the Task Completion Times (TCT).

We used complete counterbalancing to prohibit any possible systematic bias due to order effects (e.g. learning effects). As the conditions consisted of the paper instructions that were introduced with the GATM, an in-situ implementation on the smartphone, an in-situ implementation on the Microsoft HoloLens, and an in-view implementation on the Epson Moverio BT-200, a total of $4! = 24$ permutations existed. Therefore $N = 24$ participants were tested in this study.

Besides demographic data we also asked the participants to rate their experience with different AR methods and their experience with video games to get the opportunity to correlate this with their performance and/or cognitive load. We also gave participants the opportunity to write down comments, observations or suggestions regarding the hardware, methodology or implementations.

4.1 Apparatus

The assembly environment closely followed the standardized Lego Duplo assembly task [2]. It consisted of two areas, the first being the spare part area with eight blue container bins where the bricks were stored, and the second being the assembly area with a green 24x24 Lego Duplo plate (see e.g. Figure 2). We added an AR marker between the spare part area and the assembly area to aid tracking.

A Samsung Gear 360 camera was attached to a stand to record the experiment. It was placed to the left above the participants head to ensure that the participant and the inside of the containers are clearly visible.

Furthermore the four used instruction systems were stored to the right of the participant and an explanation sheet for the in-situ implementations was placed to the left. The green assembly plate was fixed to the table which in turn was to a wall. The chair was adjustable in its height and the room was adequately illuminated at all times.

4.2 Procedure

Participants were asked to sit down at the workplace and given a demographic and the four NASA Task-Load Index questionnaires. After having completed the demographic questionnaire, participants were given a short explanation of the experiment and were handed an explanation sheet containing the handling of all the instruction systems and a general explanation for the in-situ method. Furthermore they were instructed that the first priority of the experiment is to finish the task without making errors and the second priority is to do this as fast as possible. After explaining the experiment to the participants they were told that they could finish the task with either their right or left hand and that they were allowed to adjust the height of their chair.

Afterwards, participants started with the first of the four instruction systems. Before solving each of the main tasks, participants were given a test task with eight steps and comparable difficulty in order to understand and get used to the respective instruction system. The test task could be repeated as often as the participant wanted, until they felt ready to proceed. Participants were then given the main task with 32 steps and the camera was activated to record the Task Completion Time (TCT) according to the GATM [2]. After completing the task, participants were asked to complete the associated NASA TLX [7] questionnaire. This procedure was repeated for all of the four instruction systems.

Optionally, participants were given the opportunity to write down comments, observations or suggestions regarding the hardware, the methodology or the implementation after finishing all of the tasks.

4.3 Participants

The 24 participants were aged between 20 and 33 ($\oslash = 23.63$, $SD = 2.9$); 16 participants were male, 8 were female. All participants were students of Bielefeld University and had no prior experience with the Lego Duplo assembly task that was used in the experiment.

4.4 Results

4.4.1 Task Completion Times

According to the GATM [2] each assembly step is split into four phases with associated times: t_{locate}, t_{pick}, t_{locate_pos}, $t_{assemble}$. As these phases are not clearly defined in [2, 3], we propose a definition which we based our data analysis on: t_{locate} is measured from the onset of the instruction regarding an assembly step until the participant's fingers enter the bin. t_{pick} is the timespan between the hand entering the bin and exiting the container with the brick. t_{locate_pos} is the time needed from exiting the container until hovering the brick over the correct assembly position, i.e. before the brick has physical contact with the assembly position. $t_{assemble}$ is the time the participant needs to actually assemble the brick (i.e. pressing it down and releasing it).

The times for each phase according to the GATM, the number of errors and the RTLX scores have been statistically compared between the four instruction systems using a one-way repeated measures ANOVA. When the ANOVA showed a significant difference between systems, pairwise comparisons have been conducted using Bonferroni correction.

When analyzing the results based on our proposed definition one can observe (see Figure 6) that the paper instructions required the least time to locate the position of the current container and entering it with the hand(t_{locate}) with an average of 0.92s (SD = 0.26s), followed by the Microsoft HoloLens with an average of 1.11s (0.43s), the Epson Move-

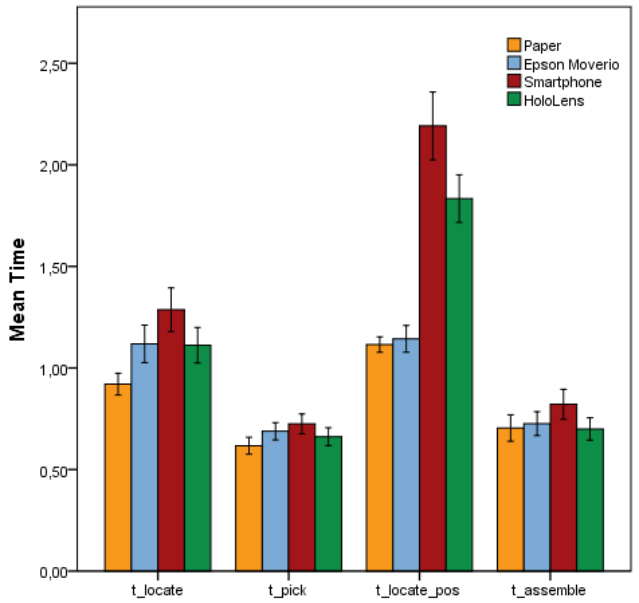

Figure 6: Task Completion Times according to the GATM. Error bars represent the standard error of the mean.

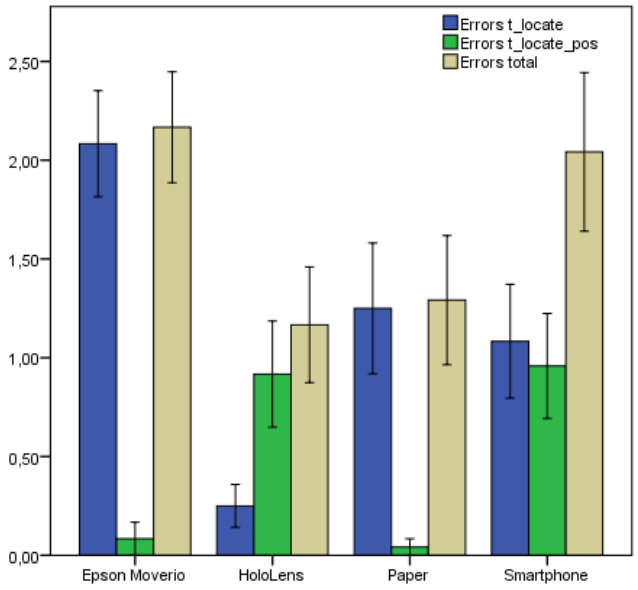

Figure 7: Errors that were made either during the whole task, in finding the correct part and in finding the correct assembly position. Error bars represent the standard error of the mean.

rio with an average of 1.12s (SD= 0.45s) and the smartphone with an average of 1.29s (SD = 0.82s). Mauchly's Test of Sphericity indicated that the assumption of sphericity was violated, $\chi^2(5) = 11.262$, p = .047, therefore a Greenhouse-Geisser correction was used ($\epsilon = 0.777$). There was a significant difference between the systems, $F(2.332, 53.63) = 9.572$, $p < .001$. The effect size estimate was $\eta^2 = 0.294$. The post-hoc tests revealed that t_{locate} was significantly shorter ($p < .05$) for paper instructions than for the Epson Moverio and the smartphone.

Considering the average time the participants needed to perform the picking of the brick of the current step (t_{pick}) all the techniques were close to each other, but the paper instructions required the least average time with 0.62s (SD = 0.2s), followed by the Microsoft HoloLens with an average of 0.66s (SD = 0.21s), the Epson Moverio 0.69s (SD = 0.21s) and the smartphone with an average of 0.73s (SD = 0.24s). There was a significant difference between the systems, $F(3, 69) = 5.684$, $p = .002$. The effect size estimate was $\eta^2 = 0.198$. The post-hoc tests revealed that t_{pick} was significantly shorter ($p < .05$) for paper instructions than for the Epson Moverio and the smartphone.

Participants needed the least time for finding the correct position to assemble the current brick (t_{locate_pos}) while using the paper instructions with an average of 1.12s (SD = 0.18s), closely followed by the Epson Moverio with an average of 1.14s (SD = 0.32). Participants needed more time when using the Microsoft HoloLens with an average of 1.83s (SD = 0.58s) and the smartphone with an average of 2.19s (SD = 0.82s). Mauchly's Test of Sphericity indicated that the assumption of sphericity was violated, $\chi^2(5) = 34.044$, $p < .001$, therefore a Greenhouse-Geisser correction was used ($\epsilon = 0.668$). There was a significant difference between the systems, $F(2.003, 46.076) = 9.572$, $p < .001$. The effect size estimate showed a large effect ($\eta^2 = 0.628$). The pairwise post-hoc tests revealed that t_{locate_pos} was significantly shorter ($p < .001$) for both pictorial instruction systems (paper and Moverio) compared with both in-situ systems (smartphone and HoloLens).

Regarding the average time participants needed to assemble the current step ($t_{assemble}$), once again the paper instructions were the fastest with an average of 0.7s (SD = 0.32s), followed by the Microsoft HoloLens with an average time of 0.7s (SD = 0.27s), the Epson Moverio with an average of 0.73s (SD = 0.29s) and the smartphone with an average of 0.82s (SD = 0.36s). Mauchly's Test of Sphericity indicated that the assumption of sphericity was violated, $\chi^2(5)$= 17.457, $p = .004$, therefore a Greenhouse-Geisser correction was used ($\epsilon = 0.666$). The ANOVA then indicated a significant difference between the systems, $F(1.998, 45.947) = 3.78$, $p = .03$, the effect size estimate was $\eta^2 = 0.141$. However, the post-hoc tests using the Bonferroni correction revealed no significant pairwise differences.

4.4.2 *Errors*

We also analyzed the average errors participants made while completing the task with each instruction techniques. While Funk et al. [3] only counted the average errors for the total assembly task ($Error_{total}$) we also decided to differentiate at which phases of the GATM the error occurred. As no errors occurred while picking or assembling a brick, we only split the errors into the phases $Error_{locate\_pos}$ and $Error_{locate}$. An $Error_{locate}$ was counted when participants clearly stuck their hand into the wrong brick container, and an error $Error_{locate\_pos}$ was counted if participants assembled the current brick onto the wrong position and released it afterwards.

Figure 7 shows the errors which were made in total and in the different phases. Overall ($Error_{total}$), participants made least mistakes using the Microsoft HoloLens (1.17, SD

= 1.43). Using the paper instructions participants made an average of 1.29 (SD = 1.6) mistakes. More mistakes were made using the implementation on the smartphone with an average of 2.04 (SD = 1.97) mistakes, closely followed by the in-view implementation on the Epson Moverio BT-200 with 2.17 (SD = 1.37) average mistakes. The ANOVA indicated a significant difference between the systems regarding the overall errors made by participants, $F(3, 69) = 3$, $p = .036$, the effect size estimate was $\eta^2 = 0.115$. The post-hoc tests using the Bonferroni correction revealed no significant pairwise differences though.

Considering only errors made while locating the correct brick container ($Error_{locate}$) we discovered that participants made with 0.25 (SD = 0.53) by far the least average errors using the Microsoft HoloLens. With the implementation on the smartphone participants made 1.08 (SD = 1.41) average errors. Using the paper instructions participants on average made 1.25 (SD = 1.62) errors. Distinctly most errors were made while using the in-view implementation on the Epson Moverio (2.08; SD = 1.32). Mauchly's Test of Sphericity indicated that the assumption of sphericity was violated, $\chi^2(5) = 18.855$, $p = .002$, therefore a Greenhouse-Geisser correction was used ($\epsilon = 0.725$). There was a significant difference between the systems, $F(2.174, 49.995) = 9.657$, $p < .001$. The effect size estimate was $\eta^2 = 0.296$. The post-hoc tests also confirmed that during locating a brick significantly less errors were made with the HoloLens in-situ system than with any other instruction system (all $p < .05$). The smartphone in-situ system caused significantly less errors during t_{locate} than the Epson Moverio in-view system as well.

While placing the brick of the current step onto the right position of the plate ($Error_{locate\_pos}$) participants on average only made 0.04 (SD = 0.2) errors with the paper instructions and 0.08 (SD = 0.41) errors with the Epson Moverio BT-200. Both in-situ implementations caused the participants to make more errors. Participants on average made 0.92 (SD = 1.32) errors while using the Microsoft HoloLens and 0.96 (SD = 1.30) errors while using the implementation on the smartphone. Mauchly's Test of Sphericity indicated that the assumption of sphericity was violated, $\chi^2(5) = 34.710$, $p < .001$, therefore a Greenhouse-Geisser correction was used ($\epsilon = 0.620$). There was a significant difference between the systems, $F(1.861, 42.807) = 9.657$, $p = .003$. The effect size estimate was $\eta^2 = 0.238$. The pairwise post-hoc tests revealed that during locating the assembly position significantly less errors were made with both pictorial instruction systems (paper and Moverio) compared with both in-situ systems (smartphone and HoloLens).

4.4.3 Cognitive load

Regarding the cognitive load, our results (see Figure 8) show that participants had the least perceived cognitive load while using the paper instructions with an RTLX score of 33.13 (SD = 17.53). This was followed by the in-view implementation on the Epson Moverio BT-200 with a score of 40.5 (SD = 20.92) and the Microsoft HoloLens with a score of 48.71 (SD = 20.3). The smartphone caused the highest perceived cognitive load with an RTLX of 49.13 (SD = 17.84). There was a significant difference between the systems regarding the RTLX score, $F(3, 69) = 10.653$, $p < .001$. The effect size estimate was $\eta^2 = 0.317$. The post-hoc tests revealed that the paper instructions caused significantly less cognitive load than the other systems (all $p < .05$), and the Moverio also less than the smartphone.

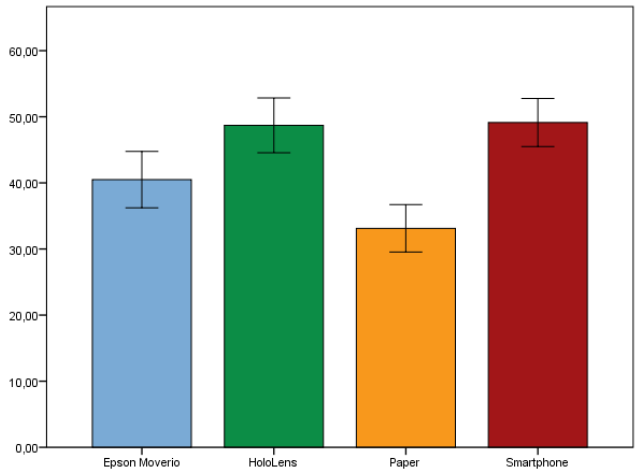

Figure 8: The cognitive load of the participants in the different conditions (NASA-TLX). Error bars represent the standard error of the mean.

4.4.4 Qualitative results

Besides the quantitative results we also collected qualitative results in form of comments, observations and suggestions from the participants and analyzed them.

A large number of participants criticized that handling the smartphone felt unnatural for the task, that the in-situ implementation on the smartphone was hard to interpret and/or that the handling of the smartphone interfered with completing the assembly task.

Regarding the in-view implementation on the Epson Moverio participants stated that displaying the instructions directly in the middle of the view interfered with the task and deteriorated the general visibility of the task. Two participants even commented that this form of presenting instructions caused a headache for them. Additionally participants complained that it was hard for them to interpret the color of the current step on the Epson Moverio and that no information was available regarding in which container the current brick is located, unlike in our in-situ AR implementations.

One half of our participants remarked that the field of view on the Microsoft HoloLens felt too small. Participants underlined positively that the container pointers (see Figure 3) in the in-situ implementations in general, but especially on the Microsoft HoloLens, were a helpful additional feature and that the three-dimensional display of the current step directly on the target made sense and felt natural.

Multiple participants chose to make a ranking of all the compared instruction techniques. In this ranking most participants chose the paper instructions as their favorite technique with the argument that it is intuitive and easy to interpret. Several participants also chose the HoloLens smart glasses either as their favorite or second-favorite system. Furthermore the advantage of having their hands free while using the Microsoft HoloLens or Epson Moverio smart glasses was explicitly mentioned by two participants.

Additionally we observed that the paper instructions that

were delivered with the GATM were interpreted inconsistently between participants. About 1/3 of the participants interpreted the paper instructions in a wrong way and tried to build the Lego Duplo task rotated by 90° to the right. As this would have caused problems for comparing the different techniques, we corrected the participants correspondingly in these cases. Another interesting observation we made was a distinct lack of understanding by some participants for the three-dimensional spatial information that was given with the in-situ techniques: While many participants could easily interpret all of the given in-situ AR instructions, others had issues understanding that these implementations did not take occlusions of bricks by other objects into account. For example, when the current Lego Duplo brick was supposed to be placed behind one that was already placed before, some participants were confused and unsure how to interpret the AR superimposition. They often assembled such bricks in a wrong way then. This might, at least partially, explain the higher number of average errors $Error_{locate\_pos}$ for the in-situ implementation.

5. DISCUSSION

While the paper instructions and the in-view instructions on the Epson Moverio were realized similarly to the ones proposed in [3], our in-situ instructions using smartphone or the Microsoft HoloLens differed.

Comparing the task completion time for the reproduced instructions types to the results found in [3], the times t_{pick} and $t_{assemble}$ are similar (approximately in the order of 1s). In our implementation, t_{locate_pos} is slightly longer than in [3], which might be due to our definition of the phases. Interestingly, t_{locate} is significantly shorter for both paper instructions as well as using the in-view display, and our in-view implementation on the Moverio smart glasses performed way better than in [3]. As these values were expected to be very similar, we suspect that external factors, e.g. lighting conditions, might be causative for this. In general, consistent to the results found in [3], paper instructions performed best with regard to task completion time.

For our newly realized in-situ instructions, the Microsoft HoloLens outperformed the smartphone in all phases of task completion time. With the HoloLens, getting instructions simply requires moving the head to focus the target, while the smartphone requires participants to hold the device in a way that information can be displayed correctly. This is why we expected the HoloLens implementation to be superior. For the HoloLens, t_{locate} was shorter than for all instruction types in [3] including (but close to) the projection-based implementation they propose.

The perceived cognitive load of the participants was lowest when using paper instructions, followed by in-view instructions. Presumably the still very limited field of view of the HoloLens was a pivotal reason why understanding AR instructions felt like a more demanding task.

Regarding the errors, in general the in-situ instructions performed best. However, dividing the errors into $Error_{locate}$ and $Error_{locate\_pos}$, we could make an interesting observation: In the phase locating the correct assembly position, there were close to zero errors using the paper instructions. However, in the phase locating the correct part for picking, the in-situ instructions using the Microsoft HoloLens performed significantly better than all other instruction types. Obviously, while it seemed to be demanding for participants to disambiguate the AR instructions at assembly positions (which in reality are sometimes occluded by previously assembled parts), providing instructions in 3D at the correct picking position can help preventing errors.

5.1 Critical reflection of the AR benchmark

By using the AR benchmark and GATM introduced by Funk et al. [2] we were able to collect some experiences and identify potential optimizations. The most important observations we made are:

- The paper instructions that were delivered with the GATM as a reference are not fully unambiguous. We observed that about 1/3 of the participants interpreted the paper instructions differently and tried to build the Lego Duplo construction rotated by 90 degrees to the right. This problem could e.g. be mitigated by putting a little marker in the lower right corner of the image on the paper instructions, or rotating the 3D model of Lego Duplo items on the images.
- While the average errors per assembly task are undoubtedly a good indicator for the stability and performance of the tested technique, we observed that it makes sense to not only count the average errors per assembly task in general, but also take into account at which phases of the GATM the error occurred (see Figure 7).
- The separation of the different phases must be unequivocally defined. The definitions we proposed and used here (see section 4.4.1) may be slightly different than the corresponding operations used in [3] as some of their results could not be reproduced.

6. CONCLUSION

We evaluated different instruction techniques based on the AR benchmark introduced by Funk et al. [2]. Beside the paper instructions that were introduced with the benchmark and the in-view implementation on the Epson Moverio that was introduced in [3], we evaluated a newly developed in-situ instruction technique for the the Microsoft HoloLens and a smartphone.

We provided evidence that using in-situ instructions with the HoloLens significantly reduced errors, while the time to find the correct part was comparable to using projected instructions as decribed in [3]. However, the performance fell short with regard to errors and time when assisting the assembly of a part. With respect to these results, a promising approach to use on smart glasses like the HoloLens could be the combination of in-situ feedback for picking and pictorial feedback for assembly. Overall, the results suggest that current AR glasses are indeed capable of providing mobile assistive instructions in a helpful manner.

In order to optimize reproducibility of future experiments, we proposed several improvements of the GATM-based AR benchmark. For our future evaluations, we plan to include even more elaborated ways of guiding participants to the correct target positions. Moreover, we will evaluate whether combined in-situ and in-view feedback can further reduce error rates. Given that to date people are commonly much more accustomed to using paper instructions than HMDs or other AR devices, it might also be worthwhile to investigate

whether over time smart glasses perform relatively better as users' experiences with this technology accumulate.

7. ACKNOWLEDGMENTS

This research was partly supported by the Cluster of Excellence Cognitive Interaction Technology 'CITEC' (EXC 277) at Bielefeld University, which is funded by the German Research Foundation (DFG), and partly by BMBF project "ADAMAAS".

8. REFERENCES

[1] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, 1992.

[2] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt. A benchmark for interactive augmented reality instructions for assembly tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, MUM '15, pages 253–257, New York, NY, USA, 2015. ACM.

[3] M. Funk, T. Kosch, and A. Schmidt. Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 934–939. ACM, 2016.

[4] M. Funk, S. Mayer, M. Nistor, and A. Schmidt. Mobile in-situ pick-by-vision: Order picking support using a projector helmet. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 45. ACM, 2016.

[5] M. Funk, S. Mayer, and A. Schmidt. Using in-situ projection to support cognitively impaired workers at the workplace. In *Proceedings of the 17th international ACM SIGACCESS conference on Computers & accessibility*, pages 185–192. ACM, 2015.

[6] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '12, pages 241–250, New York, NY, USA, 2012. ACM.

[7] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the human factors and ergonomics society annual meeting*, 2006.

[8] S. J. Henderson and S. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, 2009.

[9] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. L. Viola, T. Mashita, and H. Takemura. The effectiveness of an ar-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*, pages 57–62, March 2014.

[10] N. Petersen, A. Pagani, and D. Stricker. Real-time modeling and tracking manual workflows from first-person vision. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 117–124, Oct. 2013.

[11] P. Renner and T. Pfeiffer. Attention Guiding Techniques using Peripheral Vision and Eye Tracking for Feedback in Augmented-Reality-based Assistance Systems. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2017.

[12] L. Rodriguez, F. Quint, D. Gorecky, D. Romero, and H. R. Siller. Developing a Mixed Reality Assistance System Based on Projection Mapping Technology for Manual Operations at Assembly Workstations. *Procedia Computer Science*, 75:327–333, Jan. 2015.

[13] O. Sand, S. Büttner, V. Paelke, and C. Röcker. smARt.Assembly - Projection-Based Augmented Reality for Supporting Assembly Workers. In S. Lackey and R. Shumaker, editors, *Virtual, Augmented and Mixed Reality*, Lecture Notes in Computer Science, pages 643–652. Springer International Publishing, July 2016.

[14] B. Schwerdtfeger and G. Klinker. Supporting Order Picking with Augmented Reality. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '08, pages 91–94, Washington, DC, USA, 2008. IEEE Computer Society.

[15] 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*, CHI '03, pages 73–80, New York, NY, USA, 2003. ACM.