Computers in Human Behavior Reports 18 (2025) 100638

Contents lists available at ScienceDirect

Computers in Human Behavior Reports

journal homepage: www.sciencedirect.com/journal/computers-in-human-behavior-reports

Full length article

Productivity and learning effects of augmented reality head-mounted displays

Özgür Gürerk [a,*], Thomas Bohné [b], Guillermo Alvarez Alonso [c]

[a] Faculty of Management, Economics and Social Sciences, University of Cologne, Germany
[b] Department of Engineering, University of Cambridge, United Kingdom
[c] School of Business and Economics, RWTH Aachen University, Germany

ARTICLE INFO

Keywords:
Augmented reality
Cyber-human systems
Experiment
Productivity
Quality
Wearable technology

ABSTRACT

Augmented reality (AR) applications promise to substantially amplify human capabilities in a variety of industrial work contexts. However, there is a lack of repeated-measures studies on productivity effects of AR HMDs on humans doing real tasks in realistic industrial settings. Moreover, productivity is mostly measured as time to completion, but other important measures, such as quality and usability, are often not considered over time. To address this research gap, we carried out a repeated experiment with participants using AR HMDs for a real industrial repair task. We find that even though there are no immediate task efficiency gains, the use of AR HMDs results in significantly faster repairs over time. The use of AR HMDs also has an immediate and positive effect on work quality. However, this effect diminishes as participants gain more experience. Finally, we find that usability and comfort were no significant problems for the subjects in our experiment.

1. Introduction

Among XR devices, Augmented Reality (AR) HMDs are especially prominent in the digital transformation of manufacturing and service contexts (Bohné, 2018; Moencks, Roth, Bohné, Basso, & Betti, 2022; Moencks, Roth, Bohné, & Kristensson, 2022). Research shows that XR devices may reduce cognitive workload and improve task performance in industry (Jeffri & Rambli, 2021). Research on the effects of XR-based training on objective performance measures has gained momentum (Daling & Schlittmeier, 2022). AR based training in manual assembly tasks seems to induce promising results (Daling & Schlittmeier, 2022; Eswaran & Bahubalendruni, 2023). In some contexts, AR or VR based trainings are just as good or even better than training in a non-simulated control environment (Bohné et al., 2021; Kaplan et al., 2021).

However, as is evident from multiple scoping and systematic reviews of XR (Chiang, Shang, & Qiao, 2022; Dey, Billinghurst, Lindeman, & Swan, 2018; Kadir, Broberg, & da Conceição, 2019; Pietschmann, Bohné, & Tsapali, 2022), there is a distinct lack of field studies in industry, on industrial applications, and on the effects of XR devices on humans and their work over time.

To address this research gap and make progress with a better understanding of how XR might affect human operations in industrial settings, we report here the results of a series of field experiments we conducted with workers using an Augmented Reality (AR) head-mounted display (HMD) for an industrial repair task. AR is a particularly promising technology for skill development and human augmentation in industry as it potentially allows more intuitive, interactive, and efficient experiences (Segura et al., 2018; Westerfield, Mitrovic, & Billinghurst, 2015), and it has the potential to transform human–computer interaction (Mahr, Heller, & de Ruyter, 2023).

AR can be broadly defined as "all cases in which the display of an otherwise real environment is augmented by means of virtual (computer graphic) objects" (Milgram & Kishino, 1994). According to this definition, AR encompasses a variety of concepts and devices including HMDs and smart glasses as used in our experiments.

In our experiment we investigated the potential of AR to assist with a real repair task over time. The repair task was performed by apprentices with industry experience but no task-specific prior expertise. Participants in our study had to restore a damaged car windshield, a damage that is frequently caused by stone chips. The repair is a highly standardized process in the car industry. In our experiment, we used a between-subjects design, i.e., in our *experimental treatment*, a group of participants received step-by-step instructions via an AR HMD. In the other treatment, the same instructions were conveyed to another group of subjects via smartphones. The latter group constituted our *control treatment*, as receiving instructions via smartphones, tablets, or some handheld analog device like a handbook is currently the dominant

* Corresponding author.
E-mail address: guererk@wiso.uni-koeln.de (Ö. Gürerk).

https://doi.org/10.1016/j.chbr.2025.100638
Received 9 February 2024; Received in revised form 30 January 2025; Accepted 26 February 2025
Available online 15 March 2025

industrial practice of instructing workers. To investigate the possible effect of AR HMDs on productivity, we measured the time subjects required to repair the damaged windshield. Additionally, a third-party expert evaluated the quality of their work. To explore potential learning effects, we repeated the experiment with the same participants after four weeks in a second round.

We found, in the first round of our experiment, that the subjects who used the AR HMD delivered higher quality work than the subjects who used smartphones. There were, however, no differences between the two groups concerning the time to complete the task (TCT). In the second round of the experiment, the quality differences between the experimental groups leveled up. TCT decreased in both groups, but the decrease was significantly larger in the group with AR HMDs (28 percent) compared to the group with smartphones (14 percent).

We contribute to the literature by conducting one of the first repeated-measures experimental field studies investigating the effectiveness of AR HMDs in a repair process. We provide initial evidence that AR may lead to faster repair times over time compared to more conventional (handheld) devices. However, despite initial quality gains of AR HMDs over handheld devices, differences in quality ratings diminished over time.

2. Background

Remarkably, to date, only few experiments have examined the effect of *head-mounted* AR technology on real industrial assembly or maintenance tasks (Pietschmann et al., 2022). Baird and Barfield (1999) carried out one of the first experiments with 15 participants to study the effectiveness of AR displays on the manual assembly of a computer motherboard. They found that subjects who used HMDs performed the assembly task almost 50 percent faster than the participants who used paper instructions. Henderson and Feiner (2011) conducted a pilot experiment with an HMD to augment 18 tasks carried out by military mechanics. They found that the HMD allowed mechanics to locate tasks more quickly and resulted in less head movement. However, the experiment only involved six participants across three treatments, which limits the experiment's statistical power. Hanson, Falkenström, and Miettinen (2017) and Fager, Hanson, Medbo, and Johansson (2019) carried out experiments with five participants on the effectiveness of AR for conveying picking information. They found that if batch preparation was supported by AR this resulted in significantly shorter picking times. Seeliger, Netland, and Feuerriegel (2022) compared a HoloLens2 and tablet for a machine set-up performed by eight operators. Both devices performed similarly in task completion time and errors, but tablets were rated higher by users in perceived workload and usefulness.

2.1. Learning effects of AR HMDs in industrial settings

While AR might be able to accelerate learning processes (Ibrahim et al., 2018), its effects on the learning process of operators in industrial settings remains poorly understood, both empirically and conceptually (Dunston & Wang, 2011; Kraut, Fussell, & Siegel, 2003). According to a review conducted by Bacca et al. (2014), between 2003 and 2013 only one study focused on the use of AR technology for vocational educational training. A recent review of research on AR in vocational training (Chiang et al., 2022) found only 17 empirical studies between 2000 and 2021. This suggests both a substantial lack of and considerable potential for research on to better understand the effect of AR on learning in industrial contexts.

Most experimental studies published to date have either applied AR technology to artificial tasks, such as assembly tasks with LEGO (Alves, Marques, Ferreira, Dias, & Santos, 2022; Hou & Wang, 2013; Loch, Quint, & Brishtel, 2016) or did not use HMDs (Gavish et al., 2015; Longo, Nicoletti, & Padovano, 2017; Sirakaya & Kilic Cakmak, 2018). While these studies offer important initial insights, research has shown that transferring skills from artificial tasks or virtual environments into real world applications is not straightforward (Bossard, Kermarrec, Buche, & Tisseau, 2008; Catrambone, 1990). Moreover, empirical evidence on the effect of a wide range of HMDs on learning-related human factors such as cognitive load remains largely inconclusive (Yang et al., 2019): studies have found positive (Kalawsky, Hill, Stedmon, Cook, & Young, 2000; Strzys, Thees, Kapp, & Kuhn, 2018; Thees et al., 2020), negative (Frederiksen et al., 2020) and no influence (Ikiz, Atici-Ulusu, Taskapilioglu, & Gunduz, 2019; Kearney, Starkey, & Miller, 2020) of the technology on cognitive load. In short, what is missing are experiments that go beyond a cross-sectional design and instead use a longitudinal design to better understand the effect of AR HMDs on operators over time (Atici-Ulusu, Ikiz, Taskapilioglu, & Gunduz, 2021).

3. Experimental design and procedures

In our experiment, we focus on the potential effects of AR HMDs on productivity, measured both as time to complete the task (TCT) and work quality, as well as usability. We employed a between-subjects design, i.e., in our experimental treatment *AR* a group of subjects used AR HMDs to receive instructions, while in our control treatment *SP*, a different group of subjects used smartphones to receive instructions. AR HMDs allow the use of an AR application that provides information via a wearable head-mounted display, whereas smartphones represent the traditional approach to access information via a handheld screen. The current learning and work practice in the automotive industry, as in many other industries, is still heavily based on either a paper manual or instructions on handheld devices. The manual in our experiment describes all stages to finish the repair task (see Fig. A.1 in the Appendix).

To investigate possible learning effects, we ran two experimental rounds for each treatment. The second round took place four weeks after the first run. In both rounds, subjects either used only AR HMDs, or they only used smartphones, i.e., with respect to learning, we employed a within-subjects experimental design.

3.1. Repair task

Subjects had to complete a real task, which was the complete repair process of a damaged windshield. Such damage usually occurs when small stones hit a moving car. The repair task is a standard repair process in the car industry. There are different types of damage that can occur when a stone hits a windshield. Damages are generally classified as a fissure, star break, bulls-eye, crack, or combined damage.

In the experiment, subjects had to repair a combined damage, which is not only one of the most common damages, but also easier to replicate in an experimental research design, compared to the other damage types. Fig. 1 shows an example of the repair task in our experiment, in this case done with AR HMDs.

To make the repair process as similar as possible for each subject, an expert applied the same impact procedure for each windshield. This was done with a metal ball attached to a rubber band that was released several times against the windshield. The resulting damage classifies as a combined damage. Small variations between damaged windshields were inevitable. However, given that the repair process for combined damage is comprehensive and highly standardized, the exact shape of the damage was negligible in the actual repair process.

3.2. Repair app and devices

For the experiment, we developed an application ("app") that conveyed the repair instructions to the subjects. The app was programmed in *Java* and developed using the *Android* SDK platform, and it is compatible for any device with an API 16 or higher. The app had a step-by-step structure leading the subjects through the repair process, with

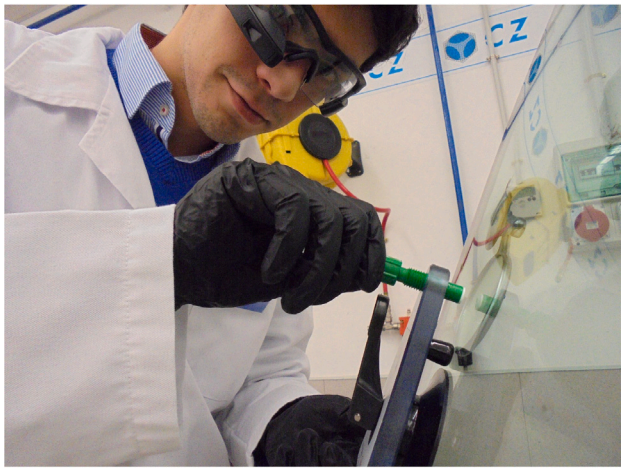


Fig. 1. Repair task.

each step shown on a new screen. Fig. A.2 in the Appendix provides further particulars of the step-by-step instructions.

After completing each step, and in order to progress with the next screen, the subjects had to tap a button on the HMD (or tap on the smartphone). During some steps, the subject could choose between a button to acquire more information on the actual step or a button to continue with the next step. To move back to the main path, the subject had to tap the Android back button that was located at the bottom of the display. The last screen corresponded with the last step when the app informed the subject that he/she had finished the task. Fig. A.3 in the Appendix shows how the app looked like in a smart phone and in the AR HMD.

Subjects in the SP treatment were provided with AQIFON 4776300 devices with a 4.0-inch touchscreen. Subjects in the AR treatment were provided with Recon Jet devices with a WQVGA 16:9 screen and a lateral touch sensor, shown in Fig. A.4 in the Appendix. The AR HMDs weighted 81.5 g. The apps running on both devices were almost identical. The display of the smartphone was larger than the display of the AR HMDs, which only slightly changed the format of the shown information. The AR app had a black background with white font, and the smartphone app had a white background with a black font, which resulted from the different design recommendations for each device. The AR HMDs required a dark background because of the weak available brightness of the display, and research suggests that a dark background is most efficient for text readability in see-through devices (Debernardis, Fiorentino, Gattullo, Monno, & Uva, 2013). This was not a limitation for the smartphones, which allowed a standard format of the background and font colors for the smartphone app.

3.3. Workplace and subjects

We conducted both rounds of the experiment at an industrial training center in Zaragoza, Spain; we ran Round 2 four weeks after Round 1. In Round 2 we replaced the windshields to replicate the experiment. 24 apprentices aged 15–23 years participated in our experiment, of which one subject was female. We had to exclude one subject from the data analysis because this participant had myopia but only reported this condition at the end of Round 2.

All subjects were enrolled in a vocational degree course in the automotive industry. 6 subjects were enrolled in their last year and therefore the second year of their professional degree; 7 subjects were enrolled in the first year of their professional degree; 8 subjects were in their second year of the obligatory vocational degree, and the other 3 were in the first year of their obligatory vocational degree.

Table 1 outlines the descriptive statistics of the treatment groups. The participants were on average 17.9 years old (SD 1.4 years). Neither the age of the participants ($z = -1.072$, $p = 0.284$), nor their vocational experience ($z = -1.176$, $p = 0.294$) significantly differed between both treatments.

Table 1
Participants' age and education.

Treatment	N	Age	Education[a]
SP	12	18.1	2.8
		(2.2)	(1.1)
AR	11	17.3	2.3
		(1.2)	(0.9)

Treatment averages. Standard deviations are in parentheses.
[a] Education is coded as 1 being the first year of the two-year obligatory vocational degree to 4 being the second year of the two-year professional degree.

Participation in the experiment was voluntary, and subjects received no monetary incentives. All procedures complied with the ethical requirements of the RWTH Aachen University for experimental studies involving human subjects. Each day, from Tuesday to Friday, we ran two sessions at 10:00 am and 12:30 pm. In each session, three subjects performed the repair in the presence of an expert, employed by the training center, simultaneously on three different windshields. The workplace, the tool case used for the repair task, and the arrangement of the subjects during an experimental session are shown in Fig. 2.

Subjects were instructed not to communicate with each other. Before the experiment started, they were given an instruction sheet that clarified the whole procedure. The translated version of the instructions are provided in the Appendix. In all sessions in Round 1 and 2, the same expert was present, who also ensured that subjects were only focused on their task.

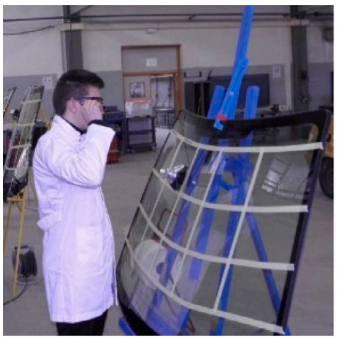
We focused on apprentices firstly, because they are one of the target groups for many training programs of industrial and manufacturing organizations. The insights of the experiment should help delineate the possible advantages and limitations that the users of AR HMD might encounter when adopting this technology. Secondly, due to the high costs of the set-up per subject, we aimed as well to have a small but representative sample of randomly selected participants, where the dexterity of the subjects was as homogeneous as possible in order to facilitate the experiment to more clearly discern the effects that each treatment could have in the users' capabilities to execute a given but unknown task.

3.4. Controlled environment and experimental set-up

To investigate the usability of both treatments in an industrial complex, we selected an experimental setting that was familiar for the subjects. The participants could reach the testing facilities as well, without the need to arrange any transportation from the vocational center. The familiarity and the reachability of the facilities had the aim to decrease differing confounding factors that each subject could encounter before and while performing their task during the experiment. The industrial training center in Zaragoza enabled the expert to control the sound exposure and limit any external distractions during the sessions. For this the center was instructed to restrict access to the facilities to anyone that was not participating in the experiment. Besides of the locational considerations, the testing sessions were organized in the morning to temper the differing effects of accumulated fatigue during the day on the subjects abilities to execute the repair process. Nevertheless, we did not perform any physical or mental examinations on the subjects to evaluate their fitness before of the experimental sessions.

3.5. The course of action of an experimental session

Our experiment consisted of two rounds. In Round 1, subjects received their respective instructions for the session upon arrival. The full instructions can be found in the Appendix. In the AR treatment,



(a) Workplace

(b) Experimental setting

(c) Tool Case

Fig. 2. The workplace and the experimental settings.

after the expert handed out the HMD and briefly explained the task, the subject could adjust the device to feel comfortable with it. Once the subject was ready, they started the task. The expert measured the time it took them to complete the task (TCT) and, separately, the time the subject needed to clean up the workplace. At the end of the repair process, the expert evaluated the quality of the repair by applying a standardized quality checklist, which is displayed in Fig. A.5 in the Appendix.

After each session of the experiment, subjects completed a questionnaire including questions about their background, and evaluated the device used with respect to the application itself, the device, and the perceived difficulty of the task. Subjects then had to answer a question about their willingness to use the device in their daily tasks. Finally, the subjects could leave a comment about things they especially liked or disliked about the task or the device.

We conducted the second experimental round four weeks after Round 1. After the first round, participants did not know that there will be a Round 2, and between the Rounds 1 and 2, they had no access to the application or the devices used in experiment. The course of action during Round 2 was exactly as during Round 1. Subjects performed the task again, without receiving any feedback on their individual performance from the previous round. Subjects also completed a questionnaire to capture subjects' progress in the repair task.

4. Results and discussion

We first test the effect of AR HMDs on repair times (TCT). Then we compare the quality ratings scored by subjects who used AR HMDs or smartphones. Finally, we investigate the usability and the comfort of both technologies. In our data analyses, we rely on the non-parametric Mann–Whitney's rank sum test to detect significant differences between treatments (also called U test, abbreviated as MWU). For within treatment comparisons, we use the non-parametric Wilcoxon matched-pairs test (abbreviated as WMP), for example to test improvements from Round 1 to Round 2. We report the p-values of two-sided tests. To investigate the impact of different variables on individual behavior, we also conduct multi-variate (parametric) regression analysis.

4.1. Effects of AR HMDs on TCT

We take TCT (in minutes) as our measure. Fig. 3 shown below presents the different findings on TCT and displays therefore the median, and the upper and the lower quartiles of TCT as observed in both treatments and rounds. There is no significant difference between subjects in the SP and AR treatment in terms of repair time in Round 1 (average TCT in SP: 47.9 min and in AR: 49.4 min, MWU test, $z = -0.431$, $p = 0.667$). The standard deviation of TCT in Round 1 amounts to 7.7 in the AR treatment, and to 5.9 in the SP treatment.

In Round 2, the subjects with AR HMDs achieved lower TCT, on average, than the subjects with smartphones (SP: 40.1 min and AR: 34.9 min, MWU test, $z = -2.462$, $p = 0.014$). Compared to Round 1, subjects in the AR treatment improved considerably in the Round 2. Their progress, measured as repair time reduction, is significant (from 49.4 min to 34.9 min, WMP test, $z = -2.934$, $p = 0.003$). We also observe a significant reduction in TCT for subjects using smartphones (from 47.9 min to 40.1 min, WMP test, $z = -2.589$, $p = 0.010$). In absolute terms as well as in percentages, however, subjects in the AR treatment display a significantly greater reduction in their TCT compared to subjects in the SP treatment. The relative TCT reduction with AR HMDs is significantly larger than with smartphones: it amounts to 28.2 percent in AR, but only to 14.4 percent in the SP treatment, MWU test, $z = -2.216$, $p = 0.027$). The standard deviation of TCT in Round 2 amounts to 4.5 in the AR treatment, and to 2.3 in the SP treatment.

Result 1. *There is no significant repair time difference between subjects using AR HMDs and smartphones when performing the task for the first time.*

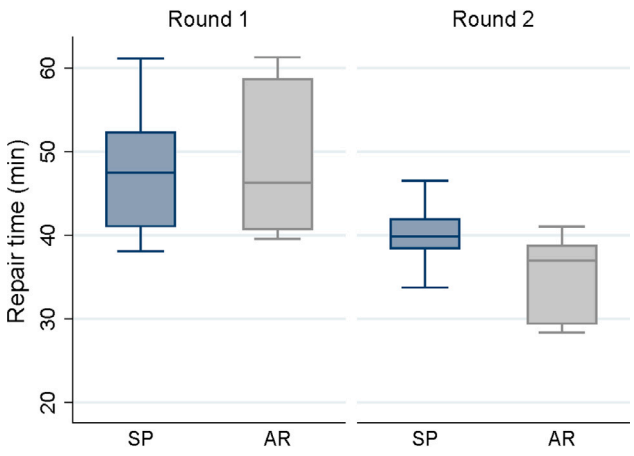


Fig. 3. Time to completion (TCT).

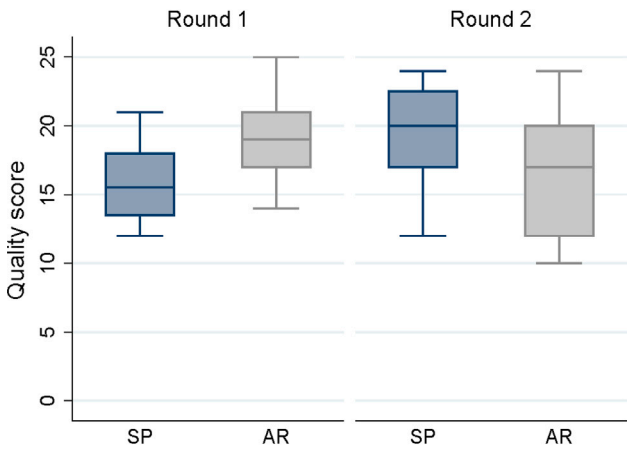


Fig. 4. Quality score.

When the task is repeated, both experimental groups become significantly faster. The relative improvement of the subjects using AR HMDs is almost double that of the subjects using smartphones.

A potential explanation for the significant reduction in TCT between both rounds may be that during the first round, subjects had to perform the task for the very first time, which required similar levels of time to understand the task. At this stage, repair times of the AR HMDs and smartphone groups are similar. In the second round, however, subjects already had a basic understanding of the repair task, and only needed cues to complete the procedure, e.g., on how long they have to polish the resin, which tool is necessary for each stage, etc. During the second round, AR HMDs hence might have allowed subjects to decrease their interaction time with the available information, as the glasses are more easily accessible, e.g., subjects did not have to reach in their pockets or to the table to access the information. This line of argument is also supported by the questionnaires after Round 1 (see Table 2) and after Round 2 (see Table 3). Subjects in both treatments considered their first interaction with the task more difficult than the second one, in terms of understanding the information provided at each stage of the repair task. On a Likert scale from 1 (very easy) to very difficult (5), in Round 1, subjects in the AR treatment rated their understanding with 2.55, and with 2.45 in the SP treatment, while in Round 2, with 2.00 in the AR treatment and 1.92 in the SP treatment. Our result is consistent with previous research that found AR HMDs leading to improved TCT (Atici-Ulusu et al., 2021; Hou & Wang, 2013), but departs from research that did not find TCT improvements (Seeliger et al., 2022).

Effects of AR HMDs on quality

Fig. 4 displays the quality scores obtained by the subjects in both experimental rounds for each of the treatments. The quality rating is reported on a scale from 1 to 25, where 25 is the highest possible quality score.

As can be seen from Fig. 4, in Round 1, subjects who used AR HMDs obtained on average higher quality scores than subjects using smartphones (SP: 15.9 and AR: 19.2, MWU test, $z = -2.381$, $p = 0.017$). In Round 2, however, there is no significant difference between both treatments (SP:19.3 and AR: 16.6, MWU test, $z = 1.518$, $p = 0.129$). In terms of the differences between the first and second trial, we observe that subjects who used smartphones improved their quality ratings (WMP test, $z = -1.932$, $p = 0.053$), while there is a negative but statistically non-significant tendency for the subjects who used AR HMDs (WMP, $z = 1.386$, $p = 0.166$).

Result 2. *AR HMDs have an immediate and positive effect on the work quality of untrained workers. The difference between AR HMDs and smartphones diminished as workers gained experience. In the second round, we observed no significant quality differences.*

An interesting question is why subjects immediately achieved a higher quality score with AR HMDs compared to subjects using smartphones? A starting point to answer this question may be the components of the quality rating. In other words, are subjects in the AR treatment better than subjects in the SP treatment in some particular sub-measures? Indeed, in Round 1, subjects who used AR HMDs were rated better in polishing (SP: 2.91, AR: 3.55, MWU test, $p = 0.132$), and significantly better in resin treatment (SP: 2.73, AR: 3.64, MWU test, $z = -2.243$, $p = 0.025$). Why were they better in these sub-tasks? The questionnaire data after Round 1 show that subjects rated the app used for the AR HMDs as more intuitive compared to the respective ratings of the smartphone group, though the difference is not statistically significant (SP: 4.18, AR: 4.45, MWU test, $z = -1.330$, $p = 0.184$). Subjects who used AR HMDs also felt that the task was easier than did the group with smartphones, though the difference again is not statistically significant (SP: 2.50, AR: 1.91, MWU test, $z = 1.500$, $p = 0.134$).

Another interesting question is why the quality ratings decreased in the AR HMDs group between the first and second round? One explanation might be that there is a productivity-quality trade-off. On the one hand, while subjects who used AR HMDs got quicker with their task (28 percent quicker in our experiment), they might have lost some of their focus on the task and quality performance might have dropped as a result. On the other hand, subjects who used smartphones were slower (they improved only 14 percent in our experiment). This might have kept participants in a more focused state, explaining the higher quality performance in the second round.

We ran a regression analysis to investigate a potentially negative relationship between TCT and quality. In our regressions, the dependent variable is the quality rating, independent variables are the TCT, and treatment, the latter one being a dummy variable, taking the value 1 for the AR, and 0 for the SP treatment. As the respective column in Table 4 shows, in Round 1, TCT had no significant effect on the obtained quality rating. In Round 2 there was a weakly significant treatment effect, as the dummy variable treatment shows, indicating that subjects with AR HMDs obtained a lower quality rating. While time still had no effect on quality for the SP treatment, there is indeed a weakly significant positive effect of time on quality in the AR treatment. The significant interaction term between TCT and the treatment dummy shows: the more time (in minutes) is taken, the higher the quality rating in the AR treatment. This positive effect alleviates the weakly significant negative treatment effect we observed in the AR treatment. Our result adds important new repeated-measures evidence to the literature, as previous cross-sectional studies found that AR HMDs led to fewer errors (Alves et al., 2022; Büttner, Funk, Sand, & Röcker, 2016; Hao & Helo, 2017)

Table 2
Results from the questionnaire after Round 1.

Questionnaire after Round 1	SP	AR
1. How intuitive was the application?*	4.18	4.45
	(0.40)	(0.69)
2. How would you rate the quality of the explanations?*	4.27	4.27
	(0.47)	(0.79)
3. How do you rate the design of the layout?*	4.36	4.27
	(0.92)	(0.47)
4. How comfortable was the device?*	4.64	4.09
	(0.62)	(1.04)
5. How do you rate the resolution of the display?*	4.36	3.91
	(0.74)	(0.70)
6. Interaction with the device during the task*	4.73	4.00
	(0.65)	(0.45)
7. Design of the glasses*	n/a	4.00
		(0.63)
8. Understanding the information at each stage**	2.45	2.55
	(0.87)	(0.93)
9. Performing the tasks described in the display**	2.55	2.18
	(0.78)	(0.75)
10. Performing the tasks while using/ wearing the device**	2.18	1.91
	(0.85)	(0.90)
11. In general terms, how do you rate the task?**	2.73	1.91
	(1.00)	(0.70)
12. How would you feel if you had to work for 8 h with this equipment?***	2.64	2.64
	(0.79)	(0.67)

Treatment averages. Standard deviations are in parentheses.
* Likert scale from very good (5) to very bad (1).
** Likert scale from very easy (1) to very difficult (5).
*** Likert scale from "I would not enjoy it" (1) to "I would enjoy it" (3).

Table 3
Results from the questionnaire after Round 2.

Questionnaire after Round 2	SP	AR
1. In general, how do you rate your last experience?*	8.25	8.64
	(1.76)	(1.29)
2. In general, how do you rate the results from your last task?*	7.25	6.73
	(1.82)	(1.49)
3. In general, how do you rate this experience?*	8.08	8.91
	(2.11)	(0.83)
4. In general, how do you rate your results from this task?*	6.83	6.82
	(2.12)	(2.09)
5. How do you rate the usability of the application?*	8.67	8.36
	(1.37)	(1.86)
6. How do you rate the information displayed?*	8.58	7.45
	(1.38)	(1.63)
7. How do you rate the interaction with the device?*	8.17	8.36
	(1.99)	(1.50)
8. Did you felt more comfortable with the device than last time?**	0.83	0.73
	(0.39)	(0.47)
9. Did you consider that this task was more difficult than the last one?**	0.00	0.00
	(0.00)	(0.00?)
10. Understanding the information at each stage***	1.92	2.00
	(0.79)	(0.63)
11. Performing the tasks described in the display***	1.92	1.73
	(0.67)	(0.47)
12. Performing the tasks while using/wearing the device***	1.58	2.27
	(0.51)	(0.65)
13. In general terms, how do you rate the task?***	2.33	2.27
	(0.89)	(0.79)

Treatment averages. Standard deviations are in parentheses.
* Likert scale from very good (10) to very bad (1).
** Binary option with yes (1) and no (0).
*** Likert scale from very easy (1) to very difficult (5).

Table 4
Multivariate analysis for quality.

Dependent variable: Quality score	Round 1	Round 2
Task completion time (TCT)	−0.016	−0.222
	(0.117)	(0.357)
Treatment ($0 = SP, 1 = AR$)	7.615	−34.041
	(7.792)	(16.718)*
TCT x Treatment	−0.087	0.862
	(0.158)	(0.432)*
Constant	16.693	28.238
	(5.674)***	(14.358)*
Adjusted R^2	0.19	0.25
N	23	23

Regression coefficients. Standard errors in parentheses.
Significance levels: (*) = 0.1, (**) = 0.05, (***) = 0.01.

Table 5
Evaluation of the application.

Round 1	SP	AR
Usability of the application	4.18	4.45
	(0.40)	(0.69)
Interaction with the device	4.33	4.00
	(0.65)	(0.45)
Comfort	4.25	4.09
	(0.62)	(1.04)
Round 2*	SP	AR
Usability of the application	8.67	8.36
	(1.37)	(1.86)
Interaction with the device	8.17	8.36
	(1.99)	(1.50)

Treatment averages. Standard deviations are in parentheses.
Likert scale from very good (10) to very bad (1).
* In Round 2, we used a more fine-grained Likert scale.

Usability of the AR HMDs

Previous studies have shown that a lack of comfort and usability is a common concern and problem preventing wider adoption of HMDs in the industrial sector (Booher, 1975; Quint et al., 2016; Stocker, Spitzer, Kaiser, Rosenberger, & Fellmann, 2017). To capture usability, we asked subjects several questions after each experiment. Table 5 summarizes our analysis of these questions. We find that the usability of the application evoked positive and consistent feedback in both treatments. Subjects rated the usability of both devices similarly. The difference between AR and SP is neither significant in the first round nor in the second round. Table 5 also shows that users did not perceive any significant difference in the interaction with their devices. Additionally, we find that subjects rated both technologies as comfortable to a similar extent. Our result departs from research that found low usability for AR HMDs compared to tablet, projection or paper-based instructions (Büttner et al., 2016; Seeliger et al., 2022).

Conclusion, limitations and future research opportunities

We started this paper with the observation that XR has the potential to amplify human capabilities in a variety of industrial contexts and may be able to play a significant role in the digital transformation of manufacturing and service systems. However, there remain significant research gaps in experimentally examining the performance effects of XR in real industrial settings. In our experiment, we focused on one particular XR technology, namely AR HMDs. We found that even though there were no immediate repair time effects, the use of AR HMDs resulted in significant gains over time compared to the control group in the second round of our experiment. The use of AR HMDs also had an immediate and positive effect on work quality for untrained workers, but this effect diminished fairly quickly as workers gained more experience. Finally, we found that the usability and comfort of the HMD presented no problems for the subjects in our experiment.

We believe our findings, which are based on a repeated experiment and real industrial repair task, make new and important contributions to our understanding of how AR HMDs affect TCT, quality and usability in industrial settings. Nevertheless, our experiment is subject to some limitations. Peer effects, such as the ones found by Gürerk, Bönsch, Kittsteiner, and Staffeldt (2019), can play a role because subjects in our experimental setting typically worked in parallel at three workstations. The distance in itself between the subjects performing the task, where meticulous attention to detail was required, was large enough to hinder imitation. However, the first subject in each round to finish their task might have had an impact on the TCT of the remaining participants. We mitigated this effect as much as possible, while at the same ensuring that the experiment reflected a realistic industrial setting. Subjects were explicitly instructed not to communicate with each other, and the expert observing the work also ensured that subjects were only focused on their tasks. Possible peer effects could be further mitigated in future studies by placing non-intrusive but visual barriers that separate each subject within the industrial facilities.

We see several opportunities to extend our experiment in future research projects. Although our data already allowed some longitudinal insights on learning and productivity effects, future research could extend our insights with a multi-year long-term study of performance effects. There is also potential for future research to experimentally study the effects of AR HMDs for different industrial tasks and in different industrial settings or explore its use in collaborative settings where peer effects could be more pronounced. Moreover, we did not measure cognitive load in our study, which could be a useful theoretical framework and measurement to consider in future studies (Ayres, Lee, Paas, and Van Merrienboer (2021), particularly as our results seem counterintuitive in that quality for users using AR diminished over time even though we would have expected cognitive load to reduce over time and improve actual task performance. Similarly, future studies could also include commonly used measures of cognitive fatigue, i.e., a decrease in cognitive resources developing over time on sustained cognitive demands (Trejo et al., 2005), to explore its potential effect on quality decrease when the task is executed repeatedly.

CRediT authorship contribution statement

Özgür Gürerk: Writing – review & editing, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Thomas Bohné:** Writing – review & editing, Writing – original draft, Project administration, Formal analysis, Data curation, Conceptualization. **Guillermo Alvarez Alonso:** Writing – original draft, Visualization, Supervision, Project administration, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We acknowledge the financial support from the Excellence Initiative (ZUK II) of the German Research Foundation DFG.

Appendix

The manual describing the stages of the repair task

See Fig. A.1.


Carrocería y pintura

Reparación de lunas

Los servicios de reparación de lunas contribuyen a fidelizar clientes al ofrecer una opción diferente frente a la sustitución.

Mª Concepción Pérez García

Las pequeñas roturas en las lunas laminadas se producen con una frecuencia muy alta, de forma que pequeños daños en el cristal que no afectan a la seguridad del vehículo, sin embargo, afectan negativamente a la estética del mismo. Frente a estos daños los usuarios pueden optar por dos posibilidades: la sustitución de la luna o su reparación.

La reparación de lunas es una alternativa interesante que deben considerar los usuarios frente a la sustitución.

Actualmente la reparación en las lunas de pequeños daños producidos por impactos de piedras u otros objetos es cada vez más frecuente, debido a las ventajas que presenta frente a la sustitución. Economía, breve tiempo de inmovilización del vehículo y simplicidad del proceso son los puntos clave a destacar en este tipo de reparación.

Cuando el daño en una luna laminada es reparable, debe considerarse la reparación y las ventajas que aporta frente a la sustitución.

En numerosas ocasiones, dependiendo del tipo y magnitud de los daños producidos en la luna se considera que no es necesario realizar su sustitución, pero si estos daños no son reparados pueden propiciar contribuyendo a que se produzca la rotura total de la misma y como consecuencia disponer como única opción de la sustitución de la misma.

La reparación de una Luna laminada frente a la tradicional sustitución por una nueva, presenta varias ventajas: el ahorro económico en materiales utilizados,

evita los desplazamientos innecesarios ya que las reparaciones pueden ser realizadas a domicilio, el tiempo de inmovilización del vehículo es muy reducido, se mantiene la luna original montada en fábrica y sus uniones evitando los posibles riesgos derivados de un proceso más complejo como la sustitución, además de contribuir a un mejor medio ambiente al disminuir el número de desechos.

Gracias a estas ventajas, los usuarios deben considerar la opción de la reparación frente a la sustitución, no obstante se debe tener en cuenta que el aspecto final conseguido en la reparación puede depender de la magnitud y del tipo de daño, así como conocer que todos los daños no es recomendable repararlos.

Las reparaciones en lunas deben realizarse por técnicos experimentados que dispongan de los conocimientos necesarios y de un equipo adecuado para conseguir un buen acabado de la reparación en el que se reduzca al mínimo la visualización del daño una vez reparado.

Debido a que en ocasiones el daño reparado puede ser apreciable visualmente y evitar posibles pérdidas de atención del conductor, se ha establecido una zona de exclusión para la luna parabrisas donde no se recomienda reparar daños, ésta debe mantenerse para no afectar a la visión directa del conductor. Para establecer la amplitud y límites de esta zona puede tomarse como referencia las siguientes líneas:

- El límite superior marcado por el barrido del limpiaparabrisas.
- La línea horizontal tangente al volante en su límite superior.
- Tomando como referencia la línea media del volante, el límite por la izquierda se fija a 120 mm. y por la derecha a 130mm.

Fundamento de la reparación

La técnica de reparación de lunas se basa en la extracción del aire atrapado en el interior de la rotura y el relleno de la misma con una resina de propiedades ópticas adecuadas para permitir una correcta visión a través de ella.

Existen varios factores que pueden afectar al resultado final de la reparación. El tipo o forma que presentan los daños con roturas más o menos abiertas en su interior pueden dificultar en mayor o menor grado la entrada de la resina de relleno. Otro factor importante es la limpieza del interior del daño, por ello, una acción recomendable frente a una rotura de la luna es tapar el punto de impacto con una lámina de plástico para evitar la entrada de suciedad hasta el momento en el que se produzca la reparación.

Zona de exclusión de daños reparables

Tipos de daños

Los daños que aparecen en la rotura de las lunas laminadas se pueden clasificar según la forma que presentan en los siguientes:

- Estrella
- Ojo de buey o diana
- Combinada
- Media Luna. (Se trata de un daño con forma igual a la mitad de un ojo de buey)
- Rotura en grieta o fisura

Equipo de reparación

En el mercado existen distintos equipos de reparación de lunas, no obstante todos basan su proceso de reparación en el fundamento descrito anteriormente, diferenciándose básicamente entre ellos en el formato que presentan los elementos que componen el equipo de reparación y las resinas utilizadas siendo más o menos viscosas para facilitar su inserción en el interior del daño.

Las ventajas a destacar en la reparación de lunas son: breve tiempo de inmovilización del vehículo, mantenimiento de la luna y uniones de origen, simplicidad del proceso de trabajo, y bajo coste económico del proceso.

Los equipos se suelen presentar en un maletín en el que se transportan los diferentes elementos utilizados variando entre ellos la inserción o no de algún elemento complementario, pero no fundamental para la realización de la reparación. Los componentes básicos son los siguientes:

- El inyector de la resina y su soporte para situarlo sobre el cristal. También puede incluir unos elementos adaptadores al soporte para acceder a diferentes zonas de la luna.
- Uno o varios tipos de resinas (de reparación y de acabado).
- Pulimento para el acabado de la reparación.
- Una lámpara de rayos ultravioleta.
- Un minitaladro, con juego de brocas muy finas y disco de pulir.
- Diverso material complementario como láminas de plástico, punzón, espejo, cuchilla, encendedor, limpiocristales, linterna o gafas de protección.
- En algunos casos pueden presentar una batería portátil que proporciona la energía eléctrica necesaria para los diferentes componentes eléctricos incluidos en el equipo.

Curado de la resina mediante la acción de una radiación ultravioleta.

Analizando las diferencias entre los distintos equipos de reparación existentes en el mercado, se destacan las siguientes: la efectividad del inyector para ejercer una mayor presión y vacío sobre el daño y la disponibilidad de resinas más o menos fluidas para su inserción en el daño.

Por otra parte, estos equipos de reparación pueden ser adquiridos a través de distribuidores de equipamiento de reparación de carrocerías. En algunos casos concretos, los equipos se distribuyen mediante el sistema de franquicias en exclusiva, siendo los franquiciados y sus equipos de técnicos los encargados de la prestación del servicio de reparación.

Proceso de reparación

Los etapas que componen el proceso reparación de un daño son las siguientes:

1 Preparación del equipo de reparación.
2 Identificación y limpieza del daño. Se identifica el tipo de daño y se limpia la zona del cristal a trabajar evitando que se introduzca en el interior el producto de limpieza utilizado. Después con un punzón se retiran los restos de cristales que se encuentren todavía en el daño.

Limpieza del daño.

3 Acetado de fisuras y aberturas de vías. En determinadas ocasiones necesario utilizar el taladro y una broca fina para abrir vías de acceso de la resina al interior del daño cuando éste es muy cerrado impidiendo la entrada de ella. En el caso de fisuras o grietas se deben estabilizar los extremos de éstas mediante un taladrado para evitar su propagación.
4 Colocación del inyector. El inyector se suele montar sobre un soporte que se fija al cristal mediante ventosas. El inyector debe situarse perpendicular al punto de impacto del daño y con la boca de forma que cubra por completo el punto de impacto, en caso contrario el inyector no podrá ejercer el efecto presión-vacío de la siguiente etapa.
5 Ciclo vacío (extracción de aire) y presión (inyección de resina). Aplicando sucesivos ciclos de vacío-presión se extrae el aire que se encuentra atrapado en el interior de la misma para, a continuación, rellenar el interior del daño con resina. Tanto los tiempos de vacío como los de presión varían en función del tipo de daño que estemos reparando y del tipo de resina que estemos utilizando, más o menos viscosa, adecuando ésta al tipo de daño y a la temperatura ambiente a la que se va a efectuar la reparación. En ocasiones es necesario calentar el daño, ligeramente y durante un tiempo muy breve, para facilitar la entrada de la resina. Una vez introducida la resina se coloca una lámina de plástico sobre el punto de impacto para evitar que salga la resina.

El inyector es el elemento fundamental del equipo de reparación de lunas.

6 Curado de la resina. Una vez que el daño se ha rellenado totalmente se procede al secado de la resina mediante el uso de una lámpara de luz ultravioleta que acelera el proceso de secado.
7 Acabado de la reparación. Una vez que hemos conseguido una correcta curación de la resina se retira mediante una cuchilla o un rascador el material sobrante y posteriormente se aplica una resina de acabado para reparar el punto de impacto con el fin de mantener la continuidad superficial del vidrio. Con la resina de acabado se actúa de forma similar a la anterior resina, se saca con la lámpara de luz ultravioleta y se elimina el material sobrante con la cuchilla.
8 Acabado final. Finalmente y para mejorar la estética de la reparación debe aplicarse un pulimento que garantiza el acabado final y con el que se consigue brillo en la superficie de la zona reparada.
9 Finalmente se recoge el puesto de trabajo.

Fig. A.1. Guidelines: Ma Concepción Pérez Garcia (Volume No 45 – Juli/September 2010, www.centro-zaragoza.com).

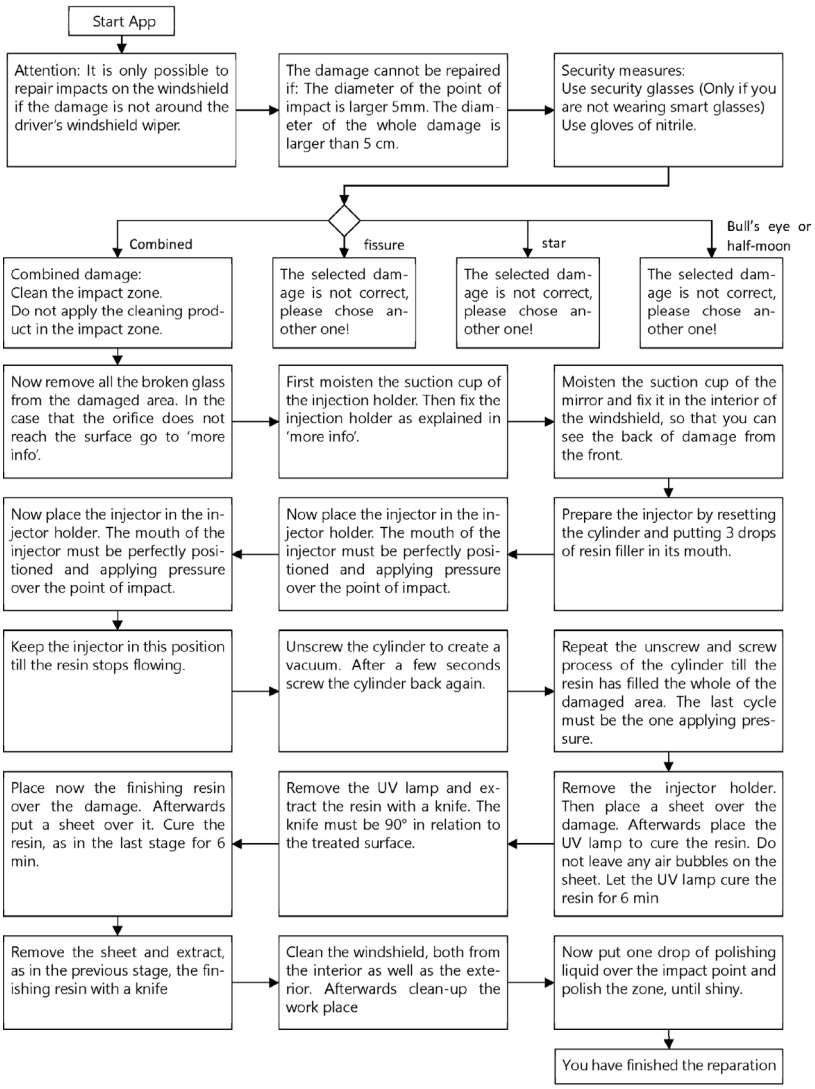


Fig. A.2. Full instructions.

(a) Smart glass display

(b) Subject uses the app on the smartphone

(c) Example of instructions

(d) Example of instructions

Fig. A.3. The user interface of the application for smart phones and AR HMDs.

Full instructions of the repair task

See Fig. A.2.

The development of the application

The application was developed in cooperation with Centro Zaragoza. First, we shot a video showing the whole repair process. The video helped to develop the application. We divided the procedure into several screens. Each screen contained only around three to five lines of explanations to ensure that the user was not overwhelmed with too much information. At the bottom of each screen, the user could select between one or two buttons. With one button a user could scroll forward. In some screens, an additional button could be selected to display more information about the current stage. After we developed all the required screens, an expert used the AR HMDs with the application to repair a damaged windshield to check that the order of the screens and information shown were correct. Once we had a final version for the AR HMDs, we copied the application into a compatible format for the smartphones.

The team considered a few aspects during the development of this application to ensure that users could navigate correctly through the different screens. Firstly, the expert controlled the size of the letters to ensure that the text was large enough so users could read the information properly and know which button to select to avoid unnecessary confusion during the repair process. The guidelines displayed had intentionally a simple user interface, style, and instructions.

Experimental instructions for the participants

Please read these instructions carefully before starting the repair process! In the first place, make sure that you feel comfortable with the AR HMDs. For this you should turn on the display. To do this, press the turn in/turn out button for 10 s and then release it. Afterwards adjust the glasses so that you can read the initial options easily. To adjust the glasses, make use of the figures below. If you have any doubts, please ask for assistance. Then locate the application “reparation of the front moon”. During the initial adaptation phase, please do not use the application with the name “reparation of the front moon”. Wait for the instructor to open it. Once inside the application, you will have to conclude the first stage and then you can select the damage that matches the one you are confronted with. In general, the application allows you to select between different options, go forwards or go backwards. Additional information about the current stage can also be displayed.

Please only conclude each operation in the manual once you have finished the task. Once you feel comfortable with the glasses go to the instructor and inform him that you are ready. He/she will then take you to the place where you will repair the front moon. Please wait for the instructor to tell you when you can start.

Please do not interrupt or pause your work during the whole period of this experiment unless it is absolutely necessary.

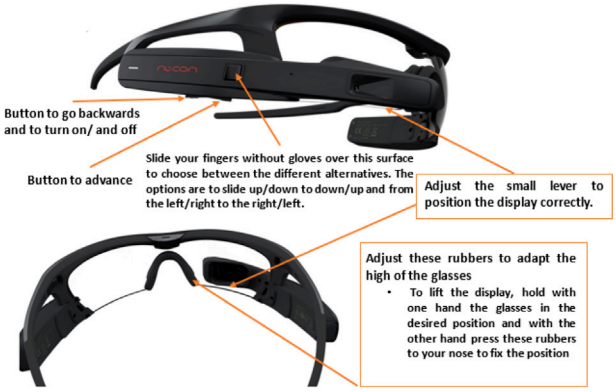


Fig. A.4. Recon Jet devices with a WQVGA 16:9 screen and a lateral touch sensor.

Evaluation

QUALITY CHECKLIST

ESTHETIC EVALUATION	VERY GOOD	GOOD	NEUTRAL	BAD	VERY BAD	POINTS
POLISHING	☐	☐	☐	☐	☐	1
RESIN TREATMENT	☐	☐	☐	☐	☐	3

ADDITIONAL QUESTIONS	YES	NO	POI-NTS
CREATED NEW DAMAGES IN THE WINDSHIELD?	☐	☐	-5
DAMAGED THE EQUIPMENT?	☐	☐	-5
CLEAN-UP THE WINDSHIELDS?	☐	☐	2
PICKED THE EQUIPMENT UP?	☐	☐	2
USED THE SAFETY EQUIPMENT CORRECTLY?	☐	☐	1
REPAIRED A WRONG DAMAGE?	☐	☐	-5

TOTAL POINTS	
END VALUE	

TIME

	TIME (MIN, SEC)
PARTICIPANTS MAIN TIME:	
PARTICIPANTS TIME TO CLEAN-UP THE WORKPLACE (OPTINAL):	

NAME OF THE EXPERT, DATE, SIGNATURE	

Page 1 von 3

Fig. A.5. Quality check.

Data availability

Data will be made available on request.

References

Alves, J. B., Marques, B., Ferreira, C., Dias, P., & Santos, B. S. (2022). Comparing augmented reality visualization methods for assembly procedures. *Virtual Reality*, 1–14.

Atici-Ulusu, H., Ikiz, Y. D., Taskapilioglu, O., & Gunduz, T. (2021). Effects of augmented reality glasses on the cognitive load of assembly operators in the automotive industry. *International Journal of Computer Integrated Manufacturing*, *34*(5), 487–499.

Ayres, P., Lee, J. Y., Paas, F., & Van Merrienboer, J. J. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, *12*, Article 702538.

Bacca, J., Baldiris, S., Fabregat, R., Graf, S., et al. (2014). Augmented reality trends in education: a systematic review of research and applications. *Journal of Educational Technology & Society*, *17*(4), 133–149.

Baird, K. M., & Barfield, W. (1999). Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality*, *4*(4), 250–259.

Bohné, T. (2018). Can cyber-human technology help us learn skills? *The Manufacturer, Special Issue on Future Tech and AI*.

Bohné, T., Heine, I., Gürerk, Ö., Rieger, C., Kemmer, L., & Cao, L. Y. (2021). Perception engineering learning with virtual reality. *IEEE Transactions on Learning Technologies*, *14*(4), 500–514.

Booher, H. R. (1975). Relative comprehensibility of pictorial information and printed words in proceduralized instructions. *Human Factors*, *17*(3), 266–277.

Bossard, C., Kermarrec, G., Buche, C., & Tisseau, J. (2008). Transfer of learning in virtual environments: a new challenge? *Virtual Reality*, *12*(3), 151–161.

Büttner, S., Funk, M., Sand, O., & Röcker, C. (2016). Using head-mounted displays and in-situ projection for assistive systems: A comparison. In *Proceedings of the 9th ACM international conference on pervasive technologies related to assistive environments* (pp. 1–8).

Catrambone, R. (1990). Specific versus general procedures in instructions. *Human-Computer Interaction*, *5*(1), 49–93.

Chiang, F.-K., Shang, X., & Qiao, L. (2022). Augmented reality in vocational training: A systematic review of research and applications. *Computers in Human Behavior*, *129*, Article 107125.

Daling, L. M., & Schlittmeier, S. J. (2022). Effects of augmented reality-, virtual reality-, and mixed reality–based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. *Human Factors*, Article 00187208221105135.

Debernardis, S., Fiorentino, M., Gattullo, M., Monno, G., & Uva, A. E. (2013). Text readability in head-worn displays: Color and style optimization in video versus optical see-through devices. *IEEE Transactions on Visualization and Computer Graphics*, *20*(1), 125–139.

Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, *5*, 37.

Dunston, P. S., & Wang, X. (2011). A hierarchical taxonomy of AEC operations for mixed reality applications. *Journal of Information Technology in Construction (ITcon)*, *16*(25), 433–444.

Eswaran, M., & Bahubalendruni, M. R. (2023). Augmented reality aided object mapping for worker assistance/training in an industrial assembly context: Exploration of affordance with existing guidance techniques. *Computers & Industrial Engineering*, *185*, Article 109663.

Fager, P., Hanson, R., Medbo, L., & Johansson, M. I. (2019). Kit preparation for mixed model assembly–efficiency impact of the picking information system. *Computers & Industrial Engineering*, *129*, 169–178.

Frederiksen, J. G., Sørensen, S. M. D., Konge, L., Svendsen, M. B. S., Nobel-Jørgensen, M., Bjerrum, F., et al. (2020). Cognitive load and performance in immersive virtual reality versus conventional virtual reality simulation training of laparoscopic surgery: a randomized trial. *Surgical Endoscopy*, *34*, 1244–1252.

Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., et al. (2015). Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, *23*(6), 778–798.

Gürerk, Ö., Bönsch, A., Kittsteiner, T., & Staffeldt, A. (2019). Virtual humans as co-workers: A novel methodology to study peer effects. *Journal of Behavioral and Experimental Economics*, *78*, 17–29.

Hanson, R., Falkenström, W., & Miettinen, M. (2017). Augmented reality as a means of conveying picking information in kit preparation for mixed-model assembly. *Computers & Industrial Engineering*, *113*, 570–575.

Hao, Y., & Helo, P. (2017). The role of wearable devices in meeting the needs of cloud manufacturing: A case study. *Robotics and Computer-Integrated Manufacturing*, *45*, 168–179.

Henderson, S. J., & Feiner, S. K. (2011). Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE international symposium on mixed and augmented reality* (pp. 191–200). IEEE.

Hou, L., & Wang, X. (2013). A study on the benefits of augmented reality in retaining working memory in assembly tasks: A focus on differences in gender. *Automation in Construction*, *32*, 38–45.

Ibrahim, A., Huynh, B., Downey, J., Höllerer, T., Chun, D., & O'donovan, J. (2018). Arbis pictus: A study of vocabulary learning with augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, *24*(11), 2867–2874.

Ikiz, Y. D., Atici-Ulusu, H., Taskapilioglu, O., & Gunduz, T. (2019). Usage of augmented reality glasses in automotive industry: Age-related effects on cognitive load. *International Journal of Recent Technology and Engineering*, *8*(3), 1–6.

Jeffri, N. F. S., & Rambli, D. R. A. (2021). A review of augmented reality systems and their effects on mental workload and task performance. *Heliyon*, *7*(3).

Kadir, B. A., Broberg, O., & da Conceição, C. S. (2019). Current research and future perspectives on human factors and ergonomics in industry 4.0. *Computers & Industrial Engineering*, Article 106004.

Kalawsky, R., Hill, K., Stedmon, A. W., Cook, C., & Young, A. (2000). Experimental research into human cognitive processing in an augmented reality environment for embedded training systems. *Virtual Reality*, *5*, 39–46.

Kaplan, A. D., Cruit, J., Endsley, M., Beers, S. M., Sawyer, B. D., & Hancock, P. A. (2021). The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: A meta-analysis. *Human Factors*, *63*(4), 706–726.

Kearney, K. G., Starkey, E. M., & Miller, S. R. (2020). Digitizing dissection: A case study on augmented reality and animation in engineering education. In *International design engineering technical conferences and computers and information in engineering conference, vol. 83921*. American Society of Mechanical Engineers, Article V003T03A015.

Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human–Computer Interaction*, *18*(1–2), 13–49.

Loch, F., Quint, F., & Brishtel, I. (2016). Comparing video and augmented reality assistance in manual assembly. In *2016 12th International conference on intelligent environments* (pp. 147–150). IEEE.

Longo, F., Nicoletti, L., & Padovano, A. (2017). Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & Industrial Engineering*, *113*, 144–159.

Mahr, D., Heller, J., & de Ruyter, K. (2023). Augmented reality (AR): The blurring of reality in human-computer interaction. *Computers in Human Behavior*, *145*, Article 107755. http://dx.doi.org/10.1016/j.chb.2023.107755, URL https://www.sciencedirect.com/science/article/pii/S0747563223001061.

Milgram, P., & Kishino, F. (1994). A taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, *77*(12), 1321–1329.

Moencks, M., Roth, E., Bohné, T., Basso, M., & Betti, F. (2022). Augmented workforce: Empowering people, transforming manufacturing. *World Economic Forum White Paper*, 1–31.

Moencks, M., Roth, E., Bohné, T., & Kristensson, P. O. (2022). Augmented workforce: contextual, cross-hierarchical enquiries on human-technology integration in industry. *Computers & Industrial Engineering*, *165*, Article 107822.

Pietschmann, L., Bohné, T., & Tsapali, M. (2022). Extended reality visual guidance for industrial environments: A scoping review. In *2022 IEEE 3rd international conference on human-machine systems* (pp. 1–6). IEEE.

Quint, F., Loch, F., Weber, H., Venitz, J., Gröber, M., & Liedel, J. (2016). Evaluation of smart glasses for documentation in manufacturing. In *Mensch und computer 2016–workshopband*. Gesellschaft für Informatik eV.

Seeliger, A., Netland, T., & Feuerriegel, S. (2022). Augmented reality for machine setups: Task performance and usability evaluation in a field test. *Procedia CIRP*, *107*, 570–575.

Segura, Á., Diez, H. V., Barandiaran, I., Arbelaiz, A., Álvarez, H., Simões, B., et al. (2018). Visual computing technologies to support the operator 4.0. *Computers & Industrial Engineering*.

Sirakaya, M., & Kilic Cakmak, E. (2018). Effects of augmented reality on student achievement and self-efficacy in vocational education and training. *International Journal for Research in Vocational Education and Training*, *5*(1), 1–18.

Stocker, A., Spitzer, M., Kaiser, C., Rosenberger, M., & Fellmann, M. (2017). Datenbrillengestützte checklisten in der fahrzeugmontage. *Informatik-Spektrum*, *40*(3), 255–263.

Strzys, M. P., Thees, M., Kapp, S., & Kuhn, J. (2018). Smartglasses in STEM laboratory courses—The augmented thermal flux experiment. In *Proceedings of the physics education research conference*.

Thees, M., Kapp, S., Strzys, M. P., Beil, F., Lukowicz, P., & Kuhn, J. (2020). Effects of augmented reality on learning and cognitive load in university physics laboratory courses. *Computers in Human Behavior*, *108*, Article 106316.

Trejo, L. J., Kochavi, R., Kubitz, K., Montgomery, L. D., Rosipal, R., & Matthews, B. (2005). Measures and models for predicting cognitive fatigue. In *Biomonitoring for physiological and cognitive performance during military operations, Vol. 5797* (pp. 105–115). SPIE.

Westerfield, G., Mitrovic, A., & Billinghurst, M. (2015). Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education*, *25*(1), 157–172.

Yang, Z., Shi, J., Jiang, W., Sui, Y., Wu, Y., Ma, S., et al. (2019). Influences of augmented reality assistance on performance and cognitive loads in different stages of assembly task. *Frontiers in Psychology*, *10*, Article 458057.