

# Comparing effectiveness of paper based and Augmented Reality instructions for manual assembly and training tasks

Sri Sudha Vijay Keshav Kolla<sup>a,\*</sup>, Andre Sanchez<sup>a</sup>, Peter Plapper<sup>a</sup>

<sup>a</sup> University of Luxembourg, 6, rue Richard Coudenhove-Kalergi, L-1359, Luxembourg

---

## Abstract

Augmented Reality (AR) systems have seen a rapid adaptation in both training and in virtual assembly instructions. AR systems assist the operator by enhancing user perception of reality, reducing the defects, and lead-time. However, there is a significant lack in the existing literature to compare AR systems and conventional work instructions. The aim of this research is to provide an empirical evidence of comparing task completion time, number of errors, workload index and system usability of AR and conventional paper-based work instructions. For this purpose, we designed an experiment where participants use paper based instructions and AR instructions to assemble a planetary gearbox. The task was assessed using NASA-TLX and System Usability Scale (SUS) tests, which allowed further analysis using descriptive statistics. Moreover, the qualitative interview at the end of the experiment gave more insights about participant's overall experience.

© 2021 The Authors. This is an open access article.

Peer Review statement: Peer-review under responsibility of the scientific committee of the 11th Conference on Learning Factories 2021.

*Keywords:* Augmented Reality; Digital instructions; Assistance systems; Head Mounted Devices; Hand Held Device; Smart glasses

---

## 1. Introduction

Conventionally, industries provided instructions on paper for assembly and training tasks. As the demand for product variants is increasing, it is becoming increasingly difficult to provide work instructions on paper in manual assembly workplaces [1]. Moreover, searching for the correct manual for a specific task can be cumbersome and increases the lead-time of the process. Secondly, training the new employees on these complex assembly tasks is often time consuming and requires a senior colleague's involvement [2], and this could be resource intensive. In today's scenario, operators need more than simple text and picture aids such as animations, videos, or projections for complex assembly tasks. These capabilities are difficult to include in the traditional paper instructions.

Industries and research institutions proposed virtual instructions as an alternative to the traditional paper based instructions. Furthermore, the digital revolution also known as industry 4.0 [3], enabled the development of new devices that cannot only display virtual instructions but also assist the operator with interactive cues. Latest technologies such as Augmented Reality (AR), which overlays digital information in the real world without compromising the awareness of reality [4], has potential to provide interactive worker assistance. Currently existing AR devices are classified into four categories [5]: Hand Held Devices (HHD), Head Mounted Devices (HMD), mobile devices, projector-based systems. These devices create value across all the industrial sectors, especially in manufacturing, AR is used in assembly work instructions [4,6–8] remote maintenance [9,10], quality assurance [11,12], and training [7].

AR systems are beginning to appear in the manufacturing landscape; however, there is a significant lack in the research about the effectiveness of these systems when compared to traditional paper instructions. Existing

---

\* Corresponding author. Tel.: +352 661 486 328

E-mail address: sri.kolla@uni.lu / sri.no1991@gmail.com

comparative studies (*section 2*) between conventional and AR systems have either chosen simple tasks or assembly of Lego bricks for the case studies. The authors of this paper believe that, to relate to a real manufacturing scenario, the experiments must include diversity of tasks such as fitting, screwing, fastening etc.

This paper provides two major results for the scientific and industrial community: (1) Reviews existing comparative studies between physical and virtual instruction systems, and (2) Presents an empirical evidence with a user study to compare conventional paper instructions with two AR systems. The first AR system is designed in a HMD (HoloLens®) and the second one uses an android mobile device (Samsung Galaxy A7). A comparison of Task Completion Time (TCT), number of errors, total workload, and system usability between paper and AR instructions are presented. This helps to understand advantages and disadvantages of new instruction systems replacing the legacy instruction systems.

With this paper, we aim to compare three instruction systems for an assembly-training task. The chosen task is to assemble a planetary gearbox contains both 3D printed and conventional parts and require diversity of tasks such as fitting, screwing, fastening etc. This is relatable to real shop floor like scenario. The three systems consists of conventional paper instructions and two AR systems, a HHD (mobile phone) and a HMD (HoloLens). The applications are developed in-house at the University of Luxembourg's industry 4.0 learning factory.

### 1.1. Structure

Section 2 presents a brief overview of existing literature in the context of comparative studies of AR vs paper instructions. Section 3 describes the methodology, experimental set-up, experimental procedure, and information regarding the participants of the user study. The results of the study are described in section 4, followed by a brief discussion in section 5. Finally, section 6 concludes the research outcomes. The experiments and user study were conducted in the industry 4.0 learning factory at University of Luxembourg.

## 2. Existing Literature

The origins of AR for instructions is not very new to the industry. Back in 1990s, researchers from Boeing, presented a heads-up see through display to project virtual cues on to the aircraft wing [13]. This helped to enhance the operator perception of the real work environment by displaying the needed boreholes on the work piece. However, not until 2006, the AR applications started to see some traction in the automotive industry and manufacturing maintenance [5]. Since then, several works have been published [14–18] related to virtual instructions using any of the four AR systems presented in section 1. However, we found that there is a lack of publications in the context of empirical evidence with user studies to measure the effectiveness of AR instructions compared with the traditional paper instructions. This section provides a brief review of five selected publications, which attempted to compare paper instructions with AR instructions.

The suitability of AR devices for an application is not only depending on the task complexity but also on the devices used and interaction mechanisms (modality). In a study [19], the authors compared the AR based assistance using smart phone (HHD), HoloLens (HMD), Epson (HMD), with conventional paper based instructions. They found that the participants finished the task quickest in paper-based instructions but made less errors in HoloLens. The perceived cognitive load was also lower in paper-based instructions. However, the cognitive load is same in HoloLens and the smart phone. The researchers encountered two challenges, one is handling the smart phone and the other is displaying the instructions in the middle of Field of View (FoV), as the instructions were interfering with the assembly task. In section 3, we explain how we tackled these challenges. The operators need a special way to adapt to the changes on actual shop-floor scenario of dealing with increasing number of product variants. As the traditional paper instructions do not fulfill the new industrial requirements such as animated instructions, video assistance etc. To tackle this issue, the authors in [1] developed in-situ projection AR system and compared it with HMD, HHD and paper instructions. Moreover, the authors in [1] claim that the developed AR systems can substantially reduce the training time and resources spent on new employees. In their empirical study, they found that on an average, the participants were quicker to finish the task in-situ projection AR system followed by paper instructions, tablet, and HMD. These results contradict with the results in [19], where participants were quicker with a HMD. However, the designing of the application, modality, and user interface of the application affects the results across experiments. The participants made fewer errors and experienced lower cognitive load with in-situ projection AR system. In their research, the least successful instruction system was a HMD device mainly because of occlusion.

In warehouse operations, picking an order is the most important task at the same time picking could be a bottleneck process if managed poorly. Researchers in [20] developed a first ever hands-free order picking system called OrderPickAR (OPA) and compared its effectiveness with existing systems such as pick-by-paper (PbP), pick-by-voice (PbVo), and pick-by-vision (PbVi). OPA's superior performance with respect to TCT, error rate, and workload is validated by descriptive statistics. A large-scale study is necessary when evaluating the effectiveness of various systems. In [21], the researchers compared the relative effectiveness of printed manual, computerized instructions, and instructions on a HMD device with 75 participants hired from university. In their experiments, they found for TCT, printed manual was the fastest followed by computerized instructions, instructions on a HMD. However, participants made fewer errors in computerized HMD instructions followed by printed manual and computerized digital instructions. In today's market, HMDs are available in two variant technologies: peripheral eye-wearable display (pHMD) and central eye-wearable displays (cHMD). The authors in [22] compared the two HMD variants with tablet and paper instructions. The experiments revealed that, for TCT, instructions on tablet is the fastest followed by cHMD, paper, and pHMD. Moreover, HMDs yielded longer finishing times on ambiguous instructions. In our research, we gave a special attention to design for AR while preparing the AR instructions to avoid ambiguity, occlusion, and collision.

### 3. Methodology

The experiments were designed to replicate the real shop floor scenario with three different instruction systems to guide the participant. The instruction system is the only independent variable. The following instruction systems were used: a conventional paper based instruction system, a HoloLens AR system (HMD), and a mobile device AR system (HHD). We chose multiple dependent variables to compare the effectiveness of the above systems. The dependent variables are, Task Completion Time (TCT), number of errors ( $\eta$ ), NASA-Task Load Index (NASA-TLX) ( $\sigma$ ) [23], and System Usability Score (SUS) ( $k$ ) [24].

In order to counterbalance the biases such as learning effects raised due to the order in which participant use different systems, 18 participants ( $3 \times 3!$ ) were invited for the user study. 3 indicates the number of independent variables. All the participants used every instruction system to complete the assembly task of a planetary gearbox. We asked the first participant to start with the paper instructions, second participant with instructions on HoloLens, and the third participant with the instructions on a mobile phone. The fourth participant will start with the paper instruction and so on.

#### 3.1. Experimental set-up

The assembly environment has three workbenches each dedicated to one instruction system as shown in Fig. 1. Each workbench consists of three main areas, 1) the assembling area, 2) the pre-assembled parts area, 3) the parts and tools area. These areas are pre-defined before the AR instructions were designed in order to facilitate the design for AR. There are two major differences in our approach in the application design compared to existing studies. The first one concerns the design of application in HoloLens. We defined separate frames for assembly tasks, pre-assembled parts, parts, & tools and made them available up on request using speech recognition. This also enabled smoother user interface. The second difference is making the mobile device stationary using a mobile holder. The advantage is that the user do not need to carry the device while assembling the parts. The disadvantage of this approach is limited Field of View (FoV) to a specific frame. However, we prioritized the fixed FoV in the mobile device to enable the user dexterity with both hands. The users stand in front of each workstation while assembling the gearbox.

#### 3.2. Procedure

At the beginning, the participants were introduced to the user study. Basic demographic information such as name, age, and gender is collected. Participants were given training on all the three instruction systems. At each workstation, the participants familiarize themselves with the working areas, the task, and the instruction system. They get first hand on experience with each system during this stage. Participants were given time to get comfortable with HoloLens and mobile device because they are inexperienced with these devices. The introduction phase takes around 15 minutes for each participant. At the end of the introduction phase, the user study begins with few verbal instructions. They were asked to finish the task as quick as possible and at the same time focus on not making errors.

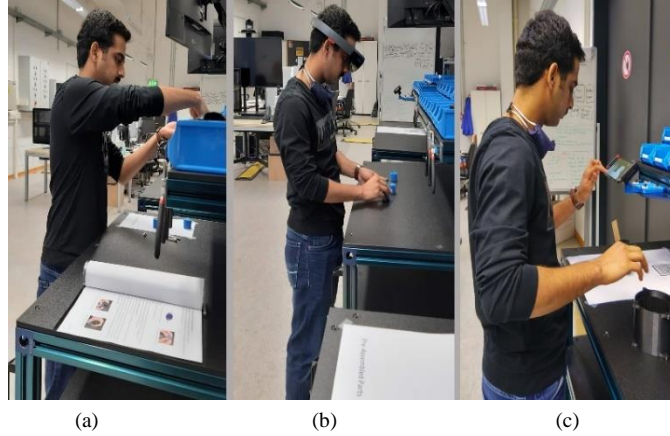


Fig. 1. Workbenches (a) with paper instructions; (b) with HoloLens AR instructions; (c) with mobile phone AR instructions.

When the participants finish the assembly-training task at one station, they were asked to fill the NASA-TLX questionnaire including pairwise comparison. Furthermore, the facilitator counted the number of errors and documented them after each workstation. The NASA-TLX questionnaire contains six subjective subscales: mental demand, physical demand, temporal demand, performance, effort, and frustration. The scales are defined from 5 (very low) to 100 (very high) with the increment of 5. When the participants finish all the three workstations and related NASA-TLX questionnaire, they were asked to fill the System Usability Scale (SUS) questionnaire comparing their usability experience with the three systems. The SUS questionnaire has 10 standard questions as mentioned in [24]. Finally, the participants were asked semi-formal qualitative questions where they suggested, commented, and gave feedback on the three systems. During the user study, the facilitator collect the information on TCT, number of errors and document user's responses for qualitative questions. On an average, it took 1.5 hours per participant to finish the user study. The next section presents the results of our experiments carried between Nov-Dec 2020 under strict hygiene and sanitary precautions.

#### 4. Results

We statistically compared Task Completion Time (TCT), number of errors ( $\eta$ ), NASA-TLX workload index ( $\sigma$ ), and system usability score ( $k$ ). Based on the relatively low number of participants ( $n=18$ ) and ambiguity of normality of data, we used a non-parametric test like Wilcoxon to measure the statistical significance of the results. In Table 1, we show the results for the mean and standard deviation (SD) of four variables calculated for the three methods of instructions. Additionally, in Fig. 2. We give a graphical comparison of the results through boxplots.

Table 1. The mean and standard deviation for four dependent variables.

|                      | Task Completion Time (TCT)(secs) |     | Number of errors ( $\eta$ ) |       | NASA-TLX workload index ( $\sigma$ ) |       | System usability score ( $k$ ) |       |
|----------------------|----------------------------------|-----|-----------------------------|-------|--------------------------------------|-------|--------------------------------|-------|
|                      | Mean                             | SD  | Mean                        | SD    | Mean                                 | SD    | Mean                           | SD    |
| <b>Paper</b>         | 1003                             | 344 | 1.39                        | 1.33  | 34.31                                | 1.2   | 44.31                          | 16.06 |
| <b>HoloLens</b>      | 912                              | 243 | 0.611                       | 0.77  | 29.1                                 | 10.69 | 76.25                          | 76.25 |
| <b>Mobile device</b> | 883                              | 240 | 0.66                        | 1.029 | 28.07                                | 12.69 | 73.75                          | 14.28 |

Results showed that the mobile device with AR instructions is the fastest method with a shorter average TCT by 2 minutes compared to paper instructions which was the slowest method. However, Wilcoxon non-parametric test revealed that there is no statistically significant difference among three proposed systems. For the second dependent variable  $\eta$ , 11 participants made no errors using mobile device, while 10 participants made zero errors using HoloLens. However, 13 participants made 1 or more errors using paper instructions. The Wilcoxon non-parametric test revealed a statistically significant difference between paper and HoloLens AR instructions (test stat  $33 <$  critical value of 40), and between paper and mobile device AR instructions (test stat  $36 <$  critical value of 40). However, there is no statistically significant difference between HoloLens and mobile device AR instructions. Therefore, we can conclude that the participants made fewer errors using AR instruction systems when compared to traditional paper instructions. Considering  $\sigma$ , the participants experienced less overall workload when using mobile device followed by HoloLens and finally paper instructions. However, the Wilcoxon non-parametric test at 95% confidence level, suggested that there is no statistically significant evidence to conclude the significant difference between the instruction systems. For the final dependent variable,  $k$ , The Wilcoxon non-parametric test has found statistically significant difference between paper and HoloLens (test stat  $4 <$  critical value of 40) and

between paper and mobile device (test stat  $2 <$  critical value of 40). However, there is no statistically significant difference in usability score between HoloLens and mobile device AR instructions. Therefore, we can conclude that the AR devices have superior usability scores than the paper instructions. At the end of the experiments, we asked participants what was their favorite instruction system based on overall experience. The participants rated an average score of 4.19/5 for HoloLens followed by 3.98/5 for mobile device, and 2.33/5 for paper instructions. It is clear that the participants preferred the experience with AR devices compared with the paper instructions.

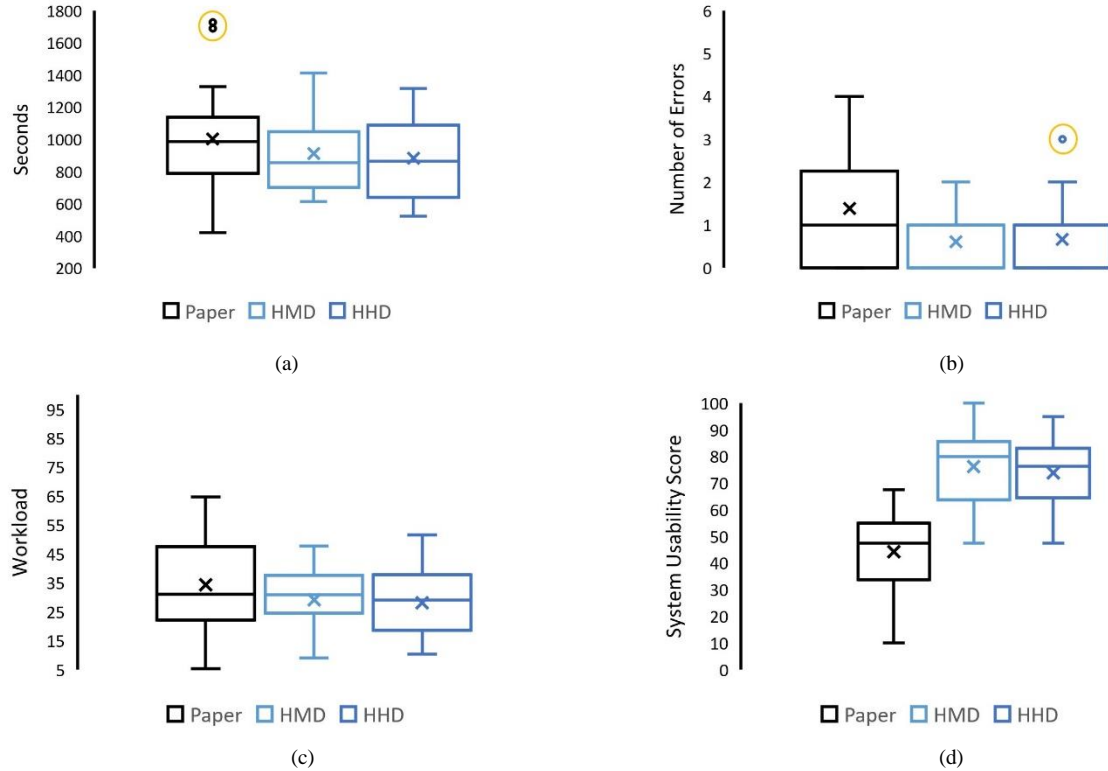


Fig. 2. The results of our study (a) TCT in seconds; (b) number of errors ( $\eta$ ); (c) NASA-TLX workload ( $\sigma$ ); (d) system usability score ( $k$ ).

## 5. Discussion

The results suggest that the participants made fewer errors using AR instruction systems. Moreover, these systems are rated high for their superior usability compared to paper instructions. The reasons for AR instruction systems to have higher acceptance rate, fewer errors, and better usability score is because of the careful design of the applications. The limitations of the existing studies are well understood before designing our AR applications. To counter the limitations of AR devices such as occlusion and collision, we adapted spatial mapping techniques, used frames, and arrows to guide the participants wherever necessary. These adaptations resulted in smoother user interface. For HoloLens we also used speech modality instead of cumbersome gesture modality. We fixed the position of mobile device AR system to enable the user dexterity of using both hands. This resulted in better results than the existing literature. Qualitative interviews conducted in the scope of learning factory research revealed the following observations from participants. Paper instructions take more time to read leading to higher TCTs. The wear-ability of HoloLens is an issue as it is non-ergonomic and heavy for extended period of wearing. Furthermore, the FoV of HoloLens is limited, so, the users need to move their head a lot causing neck pain in some participants. Nevertheless, these limitations are going to get better as the technology gets better. However, we did not encounter problems with HoloLens wear-ability for participants with prescription glasses. We also observed that the speech recognition of MixedRealityToolKit encountered issues with regional accents. It was suggested by participants to use a tablet instead of mobile phone to have a bigger display unit for mobile AR instructions. We plan to test these systems with a higher number of participants to have statistically significant results for TCT and overall workload.

## 6. Conclusion

In this paper, we compared the effectiveness of traditional paper instructions with two AR instructions systems: HoloLens and an android mobile device for an assembly-training task. The results showed that AR instructions are superior to paper based instructions in terms of number of errors and usability. However, we did not find statistically significant difference between paper and AR instructions concerning task completion time and overall

workload. We believe that the smoother experience and enhanced user perception of reality using AR instructions systems will significantly improve the lead times as people get used to these systems. We eliminated the mobile AR instruction's major problem of holding it in one hand by fixing its position with a mobile holder at the expense of FoV. In future studies, we want to investigate further with larger set of participants and to do field-tests on real assembly or training tasks in an enterprise. We also want to investigate the effectiveness of HoloLens instructions with screen dwelling interaction mechanism, as speech recognition might be ineffective in noisy shop floor scenarios.

## Acknowledgements

The research project PRODPILOT is supported by INTERREG V program. The program aims at intensifying cross-border cooperation. This goal is to be achieved by means of local and regional projects between partners from the various areas in the Greater Region of Luxembourg. The authors observe no conflicts of interests concerning this paper. Find more details about the project PRODPILOT at <https://prodpilot.eu/de>

## References

- [1] M. Funk, T. Kosch, A. Schmidt, Interactive worker assistance: comparing the effects of in-situ projection, head-mounted displays, tablet, and paper instructions, in: Proc. 2016 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., 2016: pp. 934–939.
- [2] G.I. McCalla, J.E. Greer, V.S. Kumar, P. Meagher, J.A. Collins, R. Tkatch, B. Parkinson, A peer help system for workplace training, B. d. Boulay, R. Mizoguchi (Eds.), AI-ED. 97 (1997) 183–190.
- [3] H. Kagermann, J. Helbig, A. Hellinger, W. Wahlster, Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group, Forschungsunion, 2013.
- [4] S.S.V.K. Kolla, A. Sanchez, M. Minoufekar, P. Plapper, Augmented Reality in Manual Assembly Processes, *Augment. Real. Man. Assem. Process.* (2020) 121–128.
- [5] P. Fraga-Lamas, T.M. Fernández-Caramés, Ó. Blanco-Novoa, M.A. Vilar-Montesinos, A Review on Industrial Augmented Reality Systems for the Industry 4.0 Shipyard, *IEEE Access.* 6 (2018) 13358–13375. <https://doi.org/10.1109/ACCESS.2018.2808326>.
- [6] M. Funk, T. Kosch, S.W. Greenwald, A. Schmidt, A benchmark for interactive augmented reality instructions for assembly tasks, in: Proc. 14th Int. Conf. Mob. Ubiquitous Multimed., 2015: pp. 253–257.
- [7] P. Hořejší, Augmented reality system for virtual training of parts assembly, *Procedia Eng.* 100 (2015) 699–706. <https://doi.org/10.1016/j.proeng.2015.01.422>.
- [8] F. Loch, F. Quint, I. Brishtel, Comparing video and augmented reality assistance in manual assembly, *Proc. - 12th Int. Conf. Intell. Environ. IE* 2016. (2016) 147–150. <https://doi.org/10.1109/IE.2016.31>.
- [9] D. Mourtzis, V. Zogopoulos, F. Xanthi, Augmented reality application to support the assembly of highly customized products and to adapt to production re-scheduling, *Int. J. Adv. Manuf. Technol.* 105 (2019) 3899–3910.
- [10] R. Masoni, F. Ferrise, M. Bordegoni, M. Gattullo, A.E. Uva, M. Fiorentino, E. Carrabba, M. Di Donato, Supporting remote maintenance in industry 4.0 through augmented reality, *Procedia Manuf.* 11 (2017) 1296–1302.
- [11] S.K. Ong, M.L. Yuan, A.Y.C. Nee, Augmented reality applications in manufacturing: A survey, *Int. J. Prod. Res.* 46 (2008) 2707–2742. <https://doi.org/10.1080/00207540601064773>.
- [12] H. Wuest, T. Engekle, F. Wientapper, F. Schmitt, J. Keil, From CAD to 3D Tracking - Enhancing & Scaling Model-Based Tracking for Industrial Appliances, *Adjunct. Proc. 2016 IEEE Int. Symp. Mix. Augment. Reality, ISMAR-Adjunct 2016.* (2017) 346–347. <https://doi.org/10.1109/ISMAR-Adjunct.2016.0114>.
- [13] P.C. Thomas, W.M. David, Augmented reality: An application of heads-up display technology to manual manufacturing processes, in: *Hawaii Int. Conf. Syst. Sci.*, 1992: pp. 659–669.
- [14] D. Mourtzis, V. Zogopoulos, E. Vlachou, Augmented reality application to support remote maintenance as a service in the robotics industry, *Procedia Cirp.* 63 (2017) 46–51.
- [15] R. Woll, T. Damerau, K. Wrasse, R. Stark, Augmented reality in a serious game for manual assembly processes, in: 2011 IEEE Int. Symp. Mix. Augment. Reality-Arts, Media, Humanit., IEEE, 2011: pp. 37–39.
- [16] A. Syberfeldt, O. Danielsson, M. Holm, L. Wang, Dynamic operator instructions based on augmented reality and rule-based expert systems, in: *CIRP C. 2015, 48th CIRP Conf. Manuf. Syst. Res. Innov. Manuf. Key Enabling Technol. Factories Futur. 24-26 June 2015, Ischia (Naples), Italy, Elsevier*, 2016: pp. 346–351.
- [17] O. Danielsson, A. Syberfeldt, M. Holm, L. Wang, Operators perspective on augmented reality as a support tool in engine assembly, *Procedia CIRP.* 72 (2018) 45–50. <https://doi.org/10.1016/j.procir.2018.03.153>.
- [18] O. Sand, S. Büttner, V. Paelke, C. Röcker, smart. assembly—projection-based augmented reality for supporting assembly workers, in: *Int. Conf. Virtual, Augment. Mix. Real.*, Springer, 2016: pp. 643–652.
- [19] J. Blattgerste, B. Strenge, P. Renner, T. Pfeiffer, K. Essig, Comparing conventional and augmented reality instructions for manual assembly tasks, in: Proc. 10th Int. Conf. Pervasive Technol. Relat. to Assist. Environ., 2017: pp. 75–82.
- [20] M. Funk, A.S. Shirazi, S. Mayer, L. Lischke, A. Schmidt, Pick from here! An interactive mobile cart using in-situ projection for order picking, in: Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., 2015: pp. 601–609.
- [21] A. Tang, C. Owen, F. Biocca, W. Mou, Comparative effectiveness of augmented reality in object assembly, in: Proc. SIGCHI Conf. Hum. Factors Comput. Syst., 2003: pp. 73–80.
- [22] X.S. Zheng, C. Foucault, P. Matos da Silva, S. Dasari, T. Yang, S. Goose, Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation, in: Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst., 2015: pp. 2125–2134.
- [23] S.G. Hart, NASA-task load index (NASA-TLX); 20 years later, in: Proc. Hum. Factors Ergon. Soc. Annu. Meet., Sage publications Sage CA: Los Angeles, CA, 2006: pp. 904–908.
- [24] T. Nathan, How To Use The System Usability Scale (SUS) To Evaluate The Usability Of Your Website, (2020). <https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html> (accessed October 1, 2020).