Virtual Reality (2024) 28:6
https://doi.org/10.1007/s10055-023-00895-9

ORIGINAL ARTICLE

Mixed reality prototyping for usability evaluation in product design: a case study of a handheld printer

Yanbin Wang[1] · Yatong Tian[1] · Fuchang Liu[2] · Haihai Zhou[1] · Yiwen Zhang[1]

Received: 3 August 2022 / Accepted: 27 October 2023 / Published online: 3 January 2024

Abstract
Prototyping is a critical step in the usability evaluation for product design. The maturity and affordability of mixed reality technology provide an opportunity to explore its application in prototyping. This study explored a flexible solution to create the mixed reality prototype for a handheld product by employing 3D printing, interactive 3D simulation, electronic prototyping platform, and Microsoft HoloLens. A comparative experiment was conducted to validate the effectiveness of the proposed prototype solution for usability evaluation. The results demonstrated that usability testing using the mixed prototype can accurately reveal changes in user performance across different task complexities, functional attributes, and physical contexts. The subjective assessments of product usability using the mixed prototype were highly consistent with the actual product. However, the absolute value of performance obtained from usability testing with the mixed prototype may deviate from the true value. In conclusion, mixed prototypes are more suitable for comparing the usability of different design alternatives under different conditions rather than obtaining an absolute measure of usability. This study establishes a significant theoretical foundation for product design assessment utilizing mixed prototypes, while providing practical guidance to designers and developers regarding the evaluation of product usability using mixed prototypes.

Keywords Prototyping · Mixed reality · Usability · Product design

1 Introduction

Prototyping is one of the most critical activities in new product development (Wall et al. 1992). It often predetermines a large portion of resource deployment in development and influences the success of a design project (Camburn et al. 2017). A prototype is a pre-production representation of some aspect of a concept or final design. Traditionally, physical prototypes have always been used by industrial designers to develop, communicate, and validate ideas regarding the scale, proportion, CMF (Color, Materials and Finish), ergonomics, etc. (Hallgrimsson 2012). These physical prototypes are fabricated at a range of fidelities and from a range of materials, from low-fidelity foam and card models to sophisticated, high-precision as final systems (Kent et al. 2021). The disadvantages of physical prototyping include fabrication time and cost, which often scales with fidelity and material requirements (Camburn et al. 2015). Although more and more designers have turned to 3D printing technology to create physical prototypes, the challenge of flexibility remains unresolved. Whenever there is a design change, a physical prototype has to be recreated.

In user-centered design (UCD), which was first introduced by Norman and Draper (1986), prototypes are used for iterative design evaluation and validation with target users, so that designers can better understand user needs (Lewis and Sauro 2021). UCD has been widely used in software product development (Vredenburg et al. 2002; Majrashi and Al-Wabil 2018; Salinas et al. 2020), and the application of prototyping is very successful. The advancement of information and communication technologies (ICT) shifted industrial design from the notion of product as object to product as event by the need of understanding dynamic and interactive products better within the scope of human behavior (Akoğlu and Er 2015). Given the similarities between

✉ Yanbin Wang
feihongie@gmail.com; wangyanbin@nuaa.edu.cn

1 Department of Industrial Design, Nanjing University of Aeronautics and Astronautics, 29 Yudao Street, Nanjing 210016, Jiangsu, China

2 School of Information Science and Technology, Hangzhou Normal University, 2318 Yuhangtang Road, Hangzhou 311121, Zhejiang, China

software product design and smart product design, prototyping is also at the core of the iterative design process of physical products that are becoming smarter and technology-dependent. Designers have shifted their focus toward interactions rather than solely aesthetics due to the rise of smart products.

Virtual prototypes refer to the computer simulation models established for the specific elements concerned. Initially, they were introduced as replacements for physical prototypes in order to evaluate and test product designs (Wang 2002). Several research studies have shown that virtual prototypes can be effectively used to validate the design solutions, already in the early phase of product design, when the engineering of the product is also in the early phase or even not started (Bordegoni and Rizzi 2011). Virtual prototypes allow designers to check the concept design and users' acceptance of new products through testing performed with end users. Virtual prototypes can help reduce prototyping time and costs associated with design changes. However, virtual prototypes were questioned on effectiveness due to the absence of physical touch, although they are considered more useful in recognizing cognitive usability issues (Bruno and Muzzupappa 2010).

Mixed prototyping is an emerging method used to aid in the usability assessment testing for product iterative design (Zhou and Rau 2019). In addition to visual and audio feedback, mixed prototypes can provide users with a multisensory interactive experience including tangible interaction (Park and Moon 2013) and haptic feedback (Ferrise et al. 2017). This allows to address the issues about aesthetics, ergonomics, and interaction with a unified prototyping system, involving both industrial designers and interaction designers.

Mixed prototypes are created by integrating various physical and virtual elements in one model. For example, Faust et al. (2019) developed the mixed prototype of a multimedia projector by combining a physical model and augmented reality (AR) generated images. A marker was used for the correct positioning of the AR images on the physical model. Choi (2019) developed a mixed prototype for a heater using similar techniques. There have also been mixed prototypes that leverage immersive virtual environments. Maurya et al. (2019) developed a design tool that utilized a virtual reality (VR) headset for visualizing virtual scenarios, combined with a tangible user interface (TUI) for haptic interaction. Morozova et al. (2019) demonstrated a conceptual prototype that combined the hologram of a coffee machine with 2D user interfaces on a physical touch display. It is worth mentioning that most previous research on mixed prototypes has focused primarily on interaction with the user interface of the product itself. Physical manipulation or interaction with the real environment has not been extensively explored. However, as product design methods continue to evolve, there is great potential for mixed prototyping to enable more comprehensive and realistic user experiences.

The maturity and affordability of mixed reality technology have reached sufficient levels (Microsoft 2023; MagicLeap 2023), providing an opportunity to explore its application in mixed prototyping. Mixed reality enables the synchronization of digital and physical elements in near real time, allowing virtual information to be anchored into the physical world (Kent et al. 2021). However, mixed reality is still a relatively new technology for most ordinary users. The composition and optimal representation of a mixed reality prototyping system are still undetermined according to the latest research (Kent et al. 2021).

We explored a flexible solution to create the interactive-mixed prototyping system using Microsoft HoloLens. This system was capable of accommodating frequent scheme modifications in iterative product design. Additionally, it enabled us to study the interaction between the product and the real environment. To validate the effectiveness of the proposed mixed prototyping solution and gain a better understanding of its impact on usability assessment, a comparative experiment was conducted. We investigated the effects of mixed prototyping on usability assessment results, considering factors of product function attributes, test task selection, and physical contexts. The results were compared with those obtained from testing with a fully functional prototype (actual product).

2 Method

2.1 Prototype development

A handheld printer was chosen as the subject of the usability evaluation study (Fig. 1). Although it may not be a commonly used device for most individuals, users can easily develop a mental model based on their previous experience with using mobile phones and desktop printers. Furthermore, it is a typical interactive product with both physical and cognitive interactions. The handheld printer allows users to input multiple lines of text and customize the formatting for each line individually. Its dimensions are 113.5 mm × 209.0 mm × 75.0 mm, and it weighs 440 g, including the ink cartridge. The printer can be used to print text on various surfaces. The positioning plate, depicted in Fig. 2, is an optional accessory that allows users to have better control over the position of printed text. Two interactive prototypes of the handheld printer were developed for usability evaluation. One is a virtual prototype, focusing on content input and print settings, and the other one is a mixed prototype for physical operations.

The virtual prototype was developed with Unity 3D. The product functions and interactions were realized including

Fig. 1 Handheld printer used for usability evaluation study

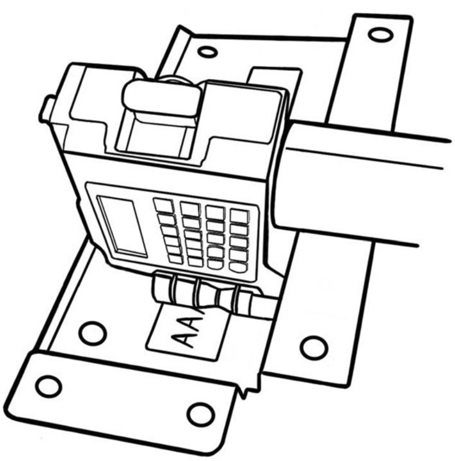


Fig. 2 Positioning plate, an optional accessory

the input of letters, numbers and symbols, input mode switch, uppercase/lowercase switch, and font and font size settings. A 7-inch touchscreen LCD display was used for testing with the virtual prototype (Fig. 3). The displayed prototype was scaled to exactly match the size of the actual product. Users input text and set the format by touching the virtual buttons, and the virtual screen on the printer displayed the user's actions in real time. Button highlighting

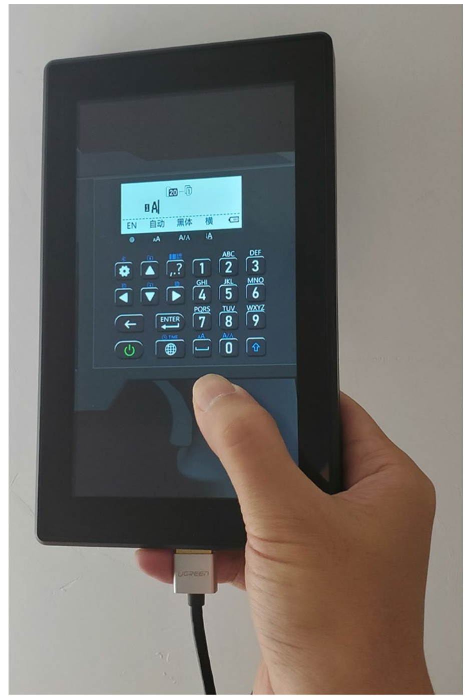


Fig. 3 Touchscreen LCD display used for testing with the virtual prototype

and sound feedback were also provided to help users confirm their actions considering the lack of force feedback.

We explored a flexible solution for mixed prototyping, which allows us to conveniently make a change in an iterative design. The mixed prototype consists of three components: a 3D-printed mockup (Fig. 4), an electronic component based on Arduino, and a holographic presentation of the product with Microsoft HoloLens (Fig. 5). Figure 6 shows the user's view of the mixed prototype while wearing the HoloLens. The mockup was 3D printed in polyamide and can be opened to put the electronic component inside. The surfaces of the interactive parts were painted with conductive ink, such as the print button, the cartridge lock knob, and the nozzle protective cover. The painted parts were connected to the Arduino board as input electrodes. When an electrode is touched, a predefined key value will be sent to the HoloLens via Bluetooth to update the simulation of the virtual product, such as opening the nozzle protective cover, unlocking the cartridge, and printing out text. The movement of the mechanism was represented by 3D simulation to simplify the physical elements of the prototype. For example, the cartridge can be locked and unlocked by rotating the knob. In our solution, the rotatable knob was replaced by two fixed parts (Fig. 7), providing users with a similar operation experience of locking and unlocking the cartridge. Meanwhile, users were able to see the hologram of the rotating animation of the knob with the HoloLens. To

Fig. 4 3D-printed mockup with cubic target

determine the position and orientation of the holographic object, Vuforia (PTC 2023) was used to recognize and track a cuboid target attached to the mockup (Fig. 4). The interactive 3D simulation of the mixed prototype was developed based on the previous virtual prototype. Vuforia engine and Bluetooth communication were integrated into the application to establish the connection between the mockup and the 3D simulation. Finally, the application was built for the Universal Window Platform to support the HoloLens.

2.2 Experimental design

To investigate the impact of prototypes on usability testing results, a learnability test and a performance test were conducted. The learnability test employed a between-subjects design, with a full-fidelity group (FF group) serving as the control group and a mixed prototype group (MP group) as the experimental group. The learnability of the product was assessed by measuring the number of mistakes and help requests (NoMHR) from participants during the learnability test. The learnability of cognitive and physical interactions was examined separately. Additionally, the learnability subscale of System Usability Scale (SUS) (Brooke 1996) was utilized as a subjective assessment.

To test the performance of input and printing, two sets of tasks were employed. The input tasks primarily focused on cognitive interaction. For the input test, a mixed factorial design was employed, with prototype category (full-fidelity prototype, mixed prototype) serving as a between-subjects factor and task complexity (simple, complex) as a within-subjects factor. Simple tasks involved entering a combination of letters and numbers, while complex tasks required inputting two lines of text with letters, numbers and symbols, and setting the two lines to the specified fonts, respectively. On the other hand, the printing tasks focused on physical interaction. Prototype category was also used as a between-subjects factor for the printing test. There were two within-subjects factors: auxiliary function and working posture. The auxiliary function was measured at two levels: without positioning plate (W/O-PP, Fig. 8a, c) and with positioning plate (W/-PP, Fig. 8b, d) to examine whether the mixed prototype could faithfully reflect the impact of auxiliary function on product usability. Considering that users often need to print on the side and top surfaces of objects using a handheld printer, two levels of the working posture factor were included: horizontal and vertical. In the horizontal condition, the user printed onto the target area on the desktop (Fig. 8a, b), while in the vertical condition, the user printed onto the target area on the wall (Fig. 8c, d).

The product usability was evaluated based on the user performance in terms of the completion time and success rate of the input tasks and the accuracy of the printing tasks. In the printing test, participants were required to print a given text on a piece of paper and center it vertically in the target area. The target area was a 10 cm × 2 cm rectangle as shown in Fig. 9. The rectangle was vertically divided into ten equal parts with a width of 2 mm for each part. To measure the accuracy of the printing tasks, the vertical offset of the printed text in the FF group (Fig. 9a) or the virtually printed text in the MP group (Fig. 9b) was counted in grids. For the MP group, the offset data were obtained from the mixed reality video recorded with the HoloLens. It was quite challenging to minimize the measurement error since the video captured was not exactly the same as what the user saw. More details about this issue will be discussed in the discussion section. Perceived usability of the product was measured using SUS. Participants rated their satisfaction with the mixed prototype on a 5-point Likert scale regarding understandability, ease of use, sense of realism, etc. The scale contains eight items and demonstrated good internal consistency (Cronbach's $\alpha = 0.85$), which was developed with reference to Park et al. (2008) and Verlinden et al. (2004)'s

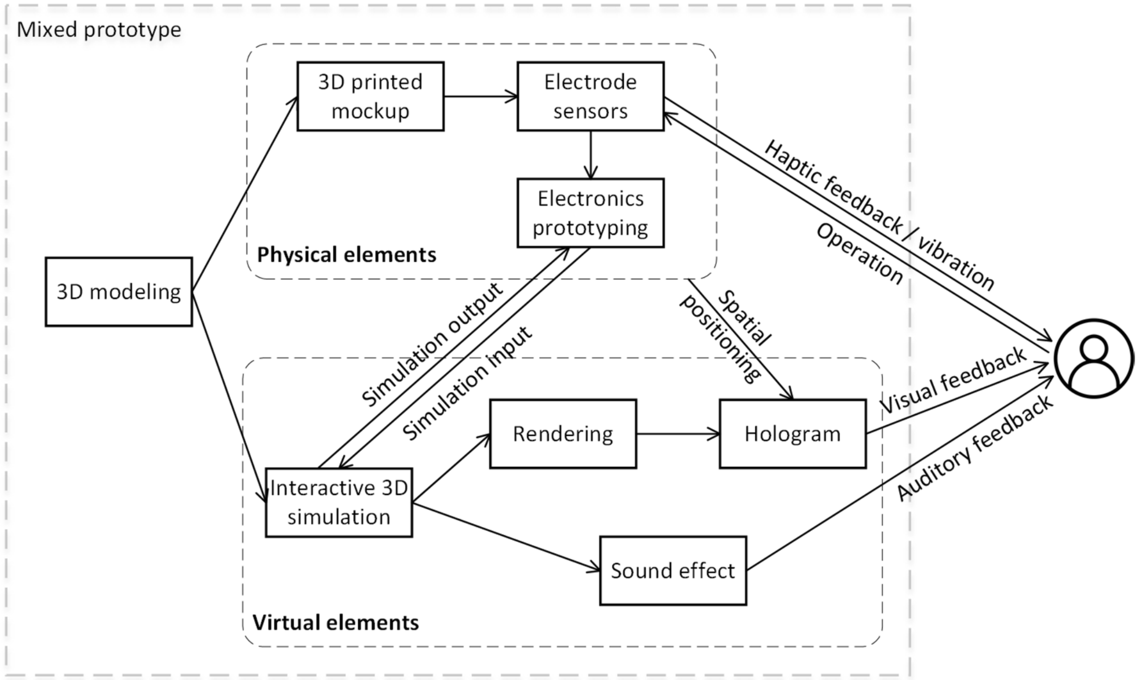


Fig. 5 System scheme of the mixed prototyping system

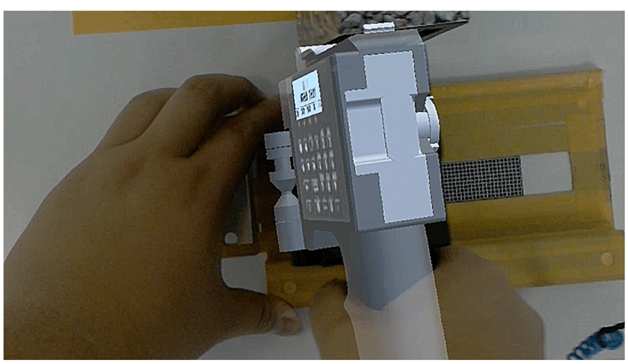


Fig. 6 User's view of mixed prototype with the HoloLens

work. Subjective workload was assessed using the NASA-Task Load Index (NASA-TLX) (Hart and Staveland 1988), which is composed of six subscales: Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Performance (OP), Effort (EF), and Frustration (FR). In addition, a questionnaire was used to assess the influencing factors that affected user performance for the MP group.

2.3 Participants

Thirty-two participants were recruited from local university campus (18 males and 14 females). The mean age of the participants was 19.6 years (SD = 1.7). Seven of them had experience in using augmented reality, while only one participant had previous experience with a handheld printer. The participants were randomly assigned to either the FF group (16 participants) or the MP group (16 participants). Thirteen participants in the FF group and 12 participants in the MP group reported that they had myopia. The myopic participants in the MP group were allowed to wear their own glasses while using the HoloLens.

2.4 Task and procedure

The participants were briefed on the research objectives and precautions before performing the experimental tasks. Afterward, they provided their consent to the experimenters and completed a demographic questionnaire.

A set of tasks (Table 1) was used to assess the learnability of the product, namely the easiness of succeeding in the basic operations at first attempt (Nielsen 1994). Before conducting the learnability test, the participants were instructed to read a concise user manual. This manual included illustrations and instructions on using the product, such as entering letters, numbers, and symbols, switching between different input methods, adjusting font and font size, and utilizing the positioning plate. Participants were allowed to refer to the user manual and search for solutions if they encountered difficulties during the test. While they were encouraged to independently complete the tasks, they were allowed to seek assistance from the experimenters if they were unable to

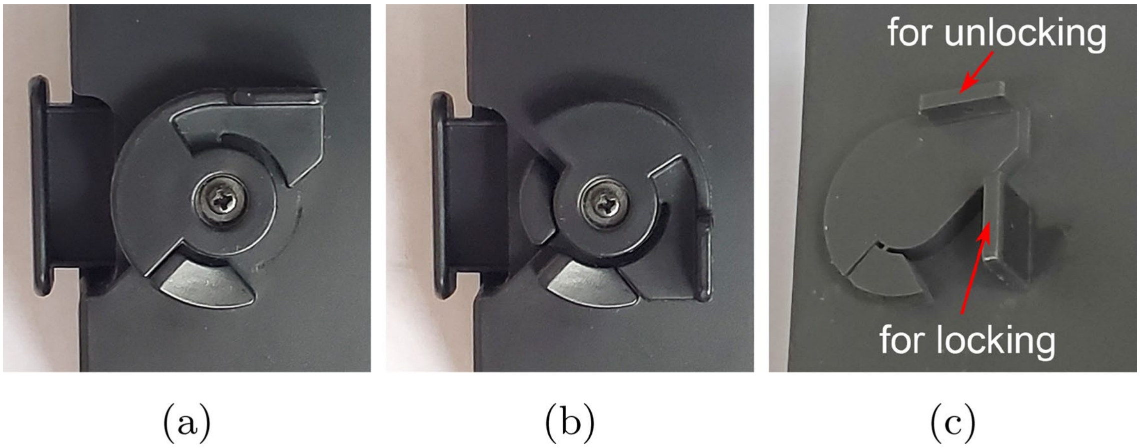


Fig. 7 Cartridge lock knob: **a** FF/locked, **b** FF/unlocked, **c** MP/simplified structure. *FF* full fidelity, *MP* mixed prototype

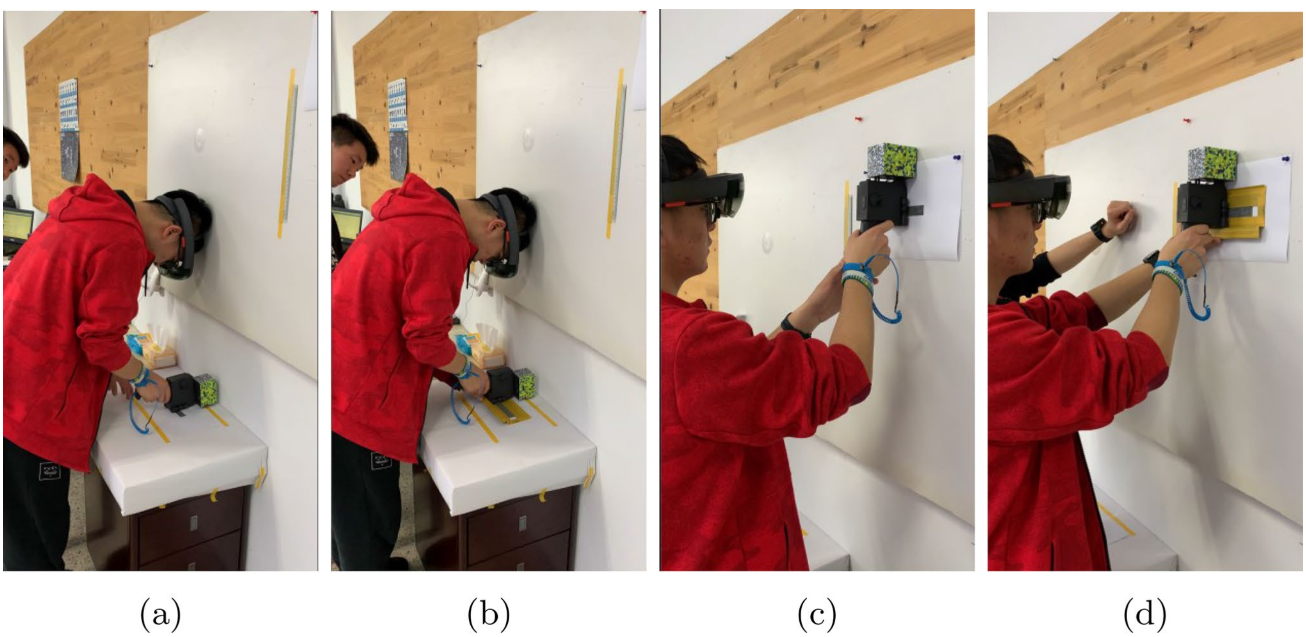


Fig. 8 Printing test conditions: **a** horizontal and W/O-PP, **b** horizontal and W/-PP, **c** vertical and W/O-PP, **d** vertical and W/-PP. *W/O-PP* Without positioning plate, *W/-PP* with positioning plate

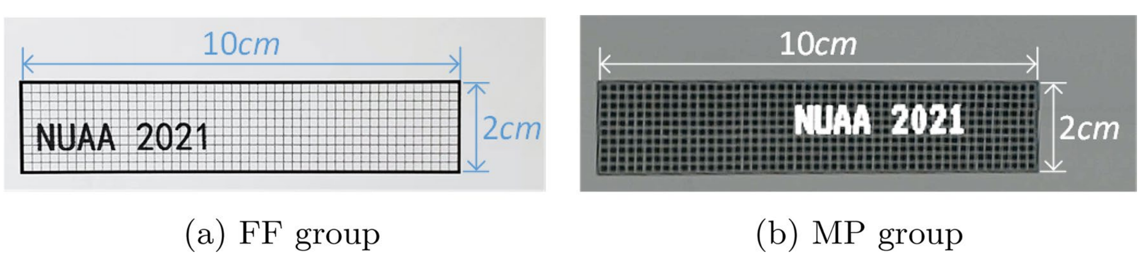


Fig. 9 Target area in the printing test

Table 1 Tasks for learnability test

Category	ID	Task
Cognitive interaction	L1	Input “2021 01 04”
	L2	Input “2c 3d 5k 7s 9w”
	L3	Input “Hello 2021”
	L4	Input “2x+3y=8”
	L5	Input “Brown Bear” and set font to “SONG”/“Medium”
	L6	Input two lines: “Blue Bird” (Line 1), “Yellow Duck” (Line 2); Set Line 1 to “HEI”/“Small” and Line 2 to “9-POINT”/“Extra small”
Physical interaction	L7	Print text to a target area without the positioning plate
	L8	Print text to a target area with the positioning plate
	L9	Reinstall the ink cartridge

succeed after attempting. The participants' operations and requests for help were recorded to measure NoMHR. The handheld printer was used in the learnability test for the FF group, while the virtual prototype was utilized for input tasks and the mixed prototype was employed for printing tasks in the MP group.

After the learnability test, participants were given the opportunity to practice using the printer or the prototypes based on their assigned groups. This practice aimed to help them become familiar with the features and usage of the product. Prior to proceeding to the performance test, they were required to pass a test to ensure that they had fully mastered the usage of the product.

In the performance test, each participant completed ten input tasks, including five simple tasks and five complex tasks (examples provided in Table 2). Participants were instructed to begin with the simple tasks. For the FF group, the handheld printer was used. To eliminate the influence of other factors such as tracking latency, the MP group used the virtual prototype instead of the mixed prototype. Although the virtual prototype had limitations in terms of affordance, participants only needed to focus on pressing buttons and monitoring the screen display for the mainly cognitive-based input tasks. Therefore, the impact of affordance was considered negligible. Participants were encouraged to complete each input task within a given time limit while minimizing errors. The time limit for each simple task was set at 18 s, while each complex task had a time limit of 55 s. These time limits were determined through a pilot test. All the trials were video recorded. Participants were asked to complete a NASA-TLX questionnaire after completing the simple task session and the complex task session, respectively. After finishing all the input tasks, participants were asked to rate the weights of the six factors of NASA-TLX based on their input experience.

After that, the printing tasks were carried out. Each participant completed two sessions (auxiliary function) × two sessions (working posture) × 4 trials = 16 trials. The order of the working posture conditions was counter-balanced, as was the order of the auxiliary function conditions within each working posture condition. The FF group used the printer to print text on paper, while the MP group used the mixed prototype to print virtual text on real paper. Before starting the test, the MP group needed to perform the HoloLens calibration procedure to ensure the best hologram viewing experience. After completing each session, participants were asked to complete a NASA-TLX questionnaire. Once all the printing tasks were finished, participants were asked to rate the weights of the six factors of NASA-TLX based on their printing experience.

The final task involved replacing the ink cartridge by following four steps: unlocking the ink cartridge by rotating the locking knob clockwise, removing the ink cartridge, checking and installing the ink cartridge, and

Table 2 Tasks for performance test

	Task	Example/description	Devices	
			FF	MP
Input	Simple complex	A15QL5 VX:ebwm (font: 7 point) M:13998822876 (font: 9 point)	Printer	Virtual prototype
Printing	W/O-PP	Print to the target area without positioning plate	Printer	Mixed prototype
	W/-PP	Print to the target area with positioning plate		
Replace ink cartridge		Rotate locking knob; remove the ink cartridge; check and install the ink cartridge; rotate locking knob	Printer	Mixed prototype

finally, locking it by rotating the knob counterclockwise. This task was only used for subjective evaluation. The operation process of the MP group was recorded with the HoloLens in the form of a mixed reality video through the device portal (Fig. 6), while a video camera was used for the FF group.

After completing all the sessions, the participants were asked to answer the SUS questionnaire regarding the usability of the handheld printer. In addition, the MP group also completed two additional questionnaires, the satisfaction questionnaire about the mixed prototype and the assessment questionnaire regarding the influencing factors of the mixed prototype. The entire experiment took around 60 min for the FF group and 90 min for the MP group.

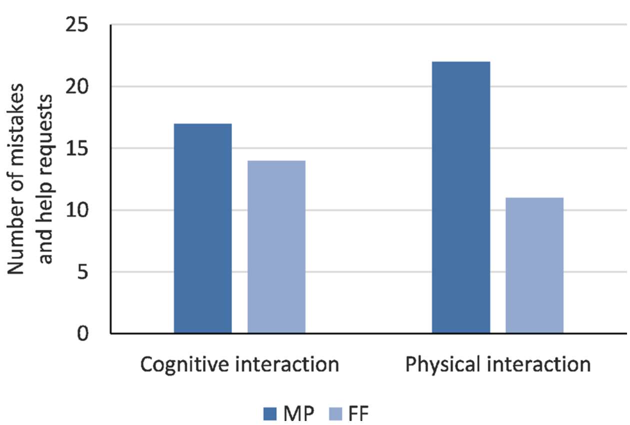


Fig. 10 Total number of mistakes and help requests in the learnability test

3 Results

3.1 Learnability

Figure 10 shows a comparison of the total NoMHR between the two groups in the learnability test. The NoMHR of cognitive interaction tasks in the MP group was higher than that of the FF group (17 vs. 14), and the difference was much larger for the physical interaction tasks (22 vs. 11). This suggests that the mixed prototype exhibits comparable effectiveness to the full-fidelity prototype when assessing the learnability through the use of cognitive interaction tasks. However, more attention should be paid to physical interaction, as mixed prototypes may present additional challenges for users during their initial attempts at physical interactions.

All the learnability issues that were observed more than once are listed in Table 3. The most common issues encountered in cognitive interaction tasks were "How to switch input methods?" and "How to switch between lowercase and uppercase?". The MP group seemed to have difficulty in finding specific keys (e.g., "Return" and "Space") and symbols that did not appear on the keyboard, such as "+". There was only a ",.?" icon displayed on the keyboard as the symbol input key. Some participants unnecessarily used "Enter" key to confirm their inputs. For physical interaction tasks, the most common issues were "Not opening the nozzle protective cover before printing," "Failing to keep the nozzle close to the paper during printing," and "Unable to use the positioning plate properly." Confusion regarding how to align the printer with the target area was only observed in the MP group. In general, the MP group encountered more difficulties in completing the tasks during the learnability test for physical interaction compared to the FF group. This could be attributed to the tracking latency and relatively small field of view (FOV) of the mixed prototype, which may have limited participants' full exploration of the product

Table 3 Learnability issues

	Issues	FF	MP
Cognitive interaction	How to switch input methods?	4	3
	How to switch between lowercase and uppercase?	4	2
	Enter unnecessary "Enter" key.	0	4
	Fail to format two lines of text separately.	2	1
	Cannot find a key/symbol (e.g., "Return," "Space," "+," etc.)	0	3
	Mistake button (e.g., mistake "Delete" for "Left," "Up" for "Shift").	1	1
Physical interaction	Not open the nozzle protective cover before printing.	3	6
	Fail to keep the nozzle close to the paper during printing.	4	2
	Cannot use the positioning plate properly.	2	4
	Fail to press the print button.	1	4
	Cannot find the location of the ink cartridge.	1	1
	How to align the printer to the target area?	0	2

use. Another possible reason could be that participants' cognitive resources were partially occupied by the use of the mixed prototype, resulting in limited attention allocation. Ratings for fidelity and task load will be presented later in this section.

3.2 Input tasks

3.2.1 Completion time

To analyze the effects of experimental factors on task completion time, a repeated-measures analysis of variance (ANOVA) was performed. Although the completion time of the MP/Complex condition did not pass Shapiro-Wilk's test for normality, a visual inspection of Q–Q plots indicated that the data were normally distributed. The homogeneity of variance was assessed using Levene's test ($p > 0.05$). The ANOVA results revealed that the main effect of task complexity ($F(1, 30) = 1021.410$, $p < 0.001$) was statistically significant. However, the main effect of prototype category ($F(1, 30) = 0.896$, $p = 0.352$) and the interaction effect ($F(1, 30) = 0.601$, $p = 0.444$) was not significant. Similar to the FF group, the MP group took significantly more time to complete complex tasks compared to simple tasks (Fig. 11). This indicates that the virtual prototype is capable of reflecting the different time requirements for input tasks of different levels of complexity.

3.2.2 Success rate and errors

The success of an input task was defined as correctly inputting the given text within the allocated time. For complex tasks, successful completion also required setting the text font correctly. Among the task failures, 14.6% of the FF group and 15.6% of the MP group were attributed to exceeding the time limit. Other input errors included missing letters/numbers/symbols, incorrect letters/numbers/symbols, adding unnecessary "Space," and selecting the wrong font settings. The success rate was analyzed using repeated-measures ANOVA for each experimental condition of the input test. No significant difference was observed in the success rate between the FF group and the MP group, $F(1, 30) = 0.109$, $p = 0.743$. As shown in Fig. 12, the success rate of the MP group was slightly lower than that of the FF group (simple tasks: 85.4% vs. 87.5%; complex tasks: 70.8% vs. 75%). The effect of task complexity on the success rate was statistically significant, $F(1, 30) = 6.677$, $p < 0.05$. The success rate of simple tasks was higher than that of complex tasks for both the FF group and the MP group, but the difference in the FF group did not reach the significant level ($p = 0.139$).

3.2.3 Task load

The overall workload score for each participant was calculated by multiplying the score of each factor by the weight the participant assigned to that factor. Figure 13 illustrates the average scores of the overall workload for simple tasks and complex tasks in both groups. The results indicate that as the complexity of the input task increases, the workload also increases when using either the mixed prototype or the full-fidelity prototype. However, the overall workload of the MP group was higher than that of the FF group, although the difference was not statistically significant. This suggests that the mixed prototype can be used to assess the differences in workload requirements for various tasks, but it may not provide accurate assessment results. Figure 14 presents the average subscale ratings for the complex tasks in the two groups. The workload of the MP group was higher than that of the FF group across all the six subscales, particularly OP and FR. This difference in workload could be attributed to

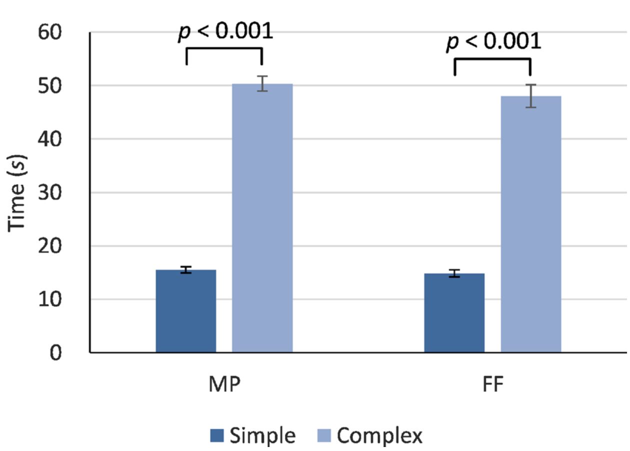


Fig. 11 Completion time of input tasks. Error bars represent ±1 standard error of the mean

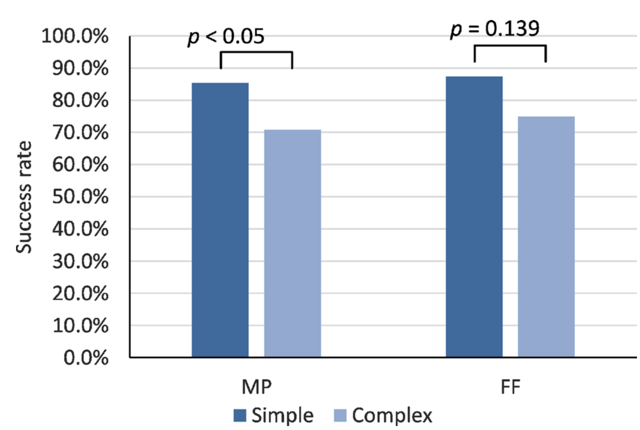


Fig. 12 Success rate of input tasks

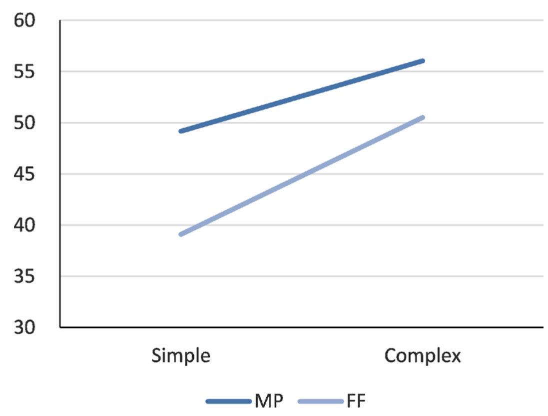


Fig. 13 Overall workload of input tasks

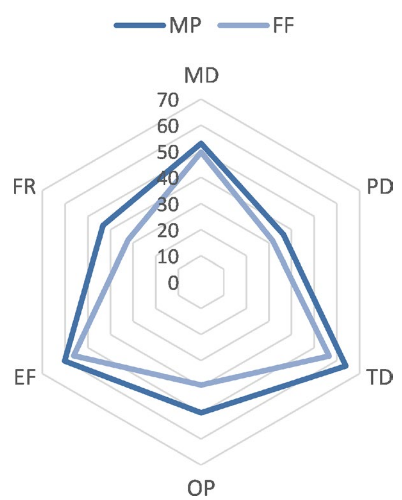


Fig. 14 Average subscale ratings of NASA-TLX for complex tasks

the higher number of task failures experienced by the MP group.

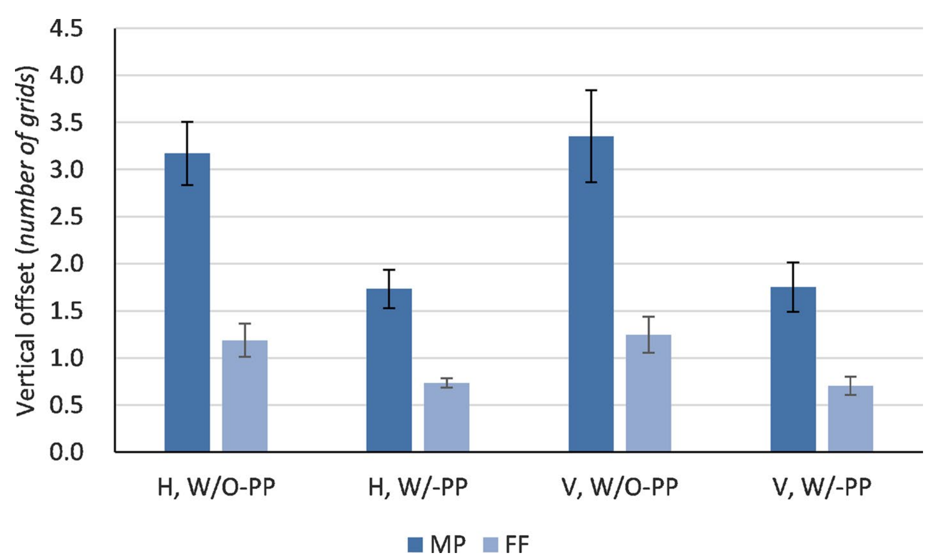
3.3 Printing tasks

3.3.1 Accuracy

Repeated-measures ANOVA was conducted to examine the effects of experimental factors on accuracy. Visual inspection of Q–Q plots indicated that the data followed a normal distribution. However, the homogeneity of variance assumption was not tenable according to Levene's test. Nonetheless, ANOVA is known to be robust against violations of homogeneity of variance, particularly when the group sizes are similar (Stevens 1996). In this study, both groups had the same sample size. Additionally, the differences in variance were approximately comparable to the differences in means between the two groups. Hence, we decided to proceed with ANOVA.

The results showed a significant main effect of auxiliary function on accuracy, $F(1, 30) = 21.799$, $p < 0.001$. Accuracy was significantly improved after using the positioning plate, with the mean vertical offset decreasing from 2.24 (SE = 0.19) to 1.23 (SE = 0.11). Furthermore, there was a significant main effect of prototype category, $F(1, 30) = 48.597$, $p < 0.001$. As shown in Fig. 15, the MP group ($M = 2.50$, SE = 0.16) had a significantly larger vertical offset compared to the FF group ($M = 0.97$, SE = 0.16). This difference will be discussed further in the next section. On the other hand, the main effect of working posture was not significant, $F(1, 30) = 0.136$, $p = 0.72$. Interestingly, a significant interaction effect between prototype category and auxiliary function was observed, $F(1, 30) = 5.570$, $p < 0.05$. The reduction in offset for the MP group due to the use of the positioning plate was significantly greater than that of the FF group (Fig. 16).

Fig. 15 Printing position accuracy under different conditions (*H* horizontal, *V* vertical, *W/O-PP* without positioning plate, *W/-PP* with positioning plate). Error bars represent ±1 standard error of the mean



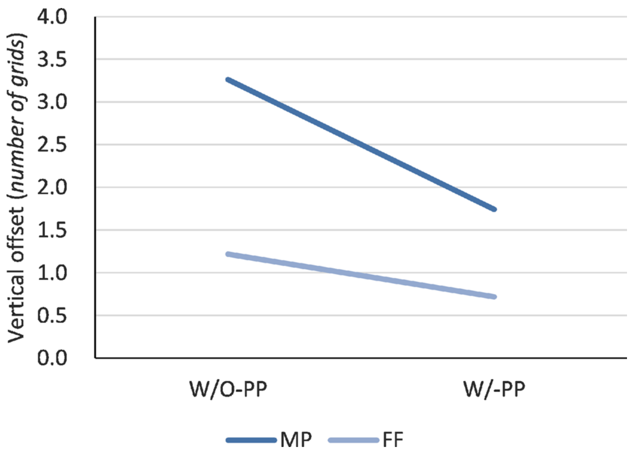


Fig. 16 Interaction effect between prototype category and auxiliary function on accuracy. One grid represents 2 mm. *W/O-PP* without positioning plate, *W/-PP* with positioning plate

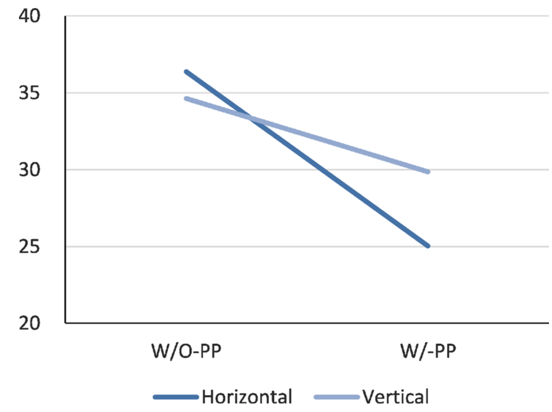


Fig. 17 Interaction effect between working posture and auxiliary function on workload. *W/O-PP* without positioning plate, *W/-PP* with positioning plate

This indicates that the effect of using the positioning plate was amplified by the mixed prototype.

3.3.2 Task load

The task load data were found to be normally distributed based on Shapiro-Wilk's test, and the assumption of homogeneity of variances was met across all groups according to Levene's test. Standard box-plot analyses were conducted to identify outliers, and one outlier was observed. This outlier was subsequently removed before conducting the ANOVA. A repeated-measures ANOVA revealed a significant main effect of auxiliary function, $F(1, 29) = 23.578$, $p < 0.001$. Additionally, the interaction effect between working posture and auxiliary function was also significant, $F(1, 29) = 18.551$, $p < 0.001$. As shown in Fig. 17, the auxiliary function of the positioning plate significantly reduced the task load, and its impact was significantly greater in the horizontal condition compared to the vertical condition. However, the interaction effect between prototype category and auxiliary function did not reach the significant level, $F(1, 29) = 4.099$, $p = 0.052$. Figure 18 illustrates the workload of the MP group and the FF group under different experimental conditions. It can be observed that the impact of the auxiliary function on workload in the MP group was generally smaller than that in the FF group. Particularly, in the vertical condition, the workload reduction in the MP group was almost negligible.

3.4 Subjective assessment

3.4.1 Perceived usability

Table 4 presents the results of perceived usability measured with SUS questionnaires, including the means and 95% confidence intervals of the global SUS scores and the two subscales, usability and learnability (Lewis and Sauro 2009). The perceived usability of the product was above average whether using the mixed prototype or the printer (a SUS score above 68 is considered above average). A t test revealed that there was no significant difference in mean SUS scores between the two groups ($t(30) = 0.644$, $p = 0.524$), although the mean score of the MP group was slightly lower than that of the FF group. We further examined the two subscales and found that the difference was mainly from the learnability subscale. Despite emphasizing to the MP group that the evaluation was on the product itself rather than the prototype system, the evaluation results were still somewhat influenced by the fidelity of the mixed prototype. The difficulty of learning to use the product increased by the mixed prototype, as indicated by the drop in the mean learnability score from 75.78 to 66.41.

3.4.2 Satisfaction

Figure 19 illustrates the results of the satisfaction evaluation of the mixed prototype in terms of product functions, features, usage, etc., as well as overall satisfaction. The mean scores for functions and operations were approximately 4.5, indicating that participants could comprehend the functions and usage of the product well through the mixed prototype. The mean scores of overall satisfaction, appearance, and physical properties were close to 4.0. The items with the lowest scores were feedback, purchase decision, and ease of use. It is likely that feedback and purchase decision were influenced by the prototype fidelity, such as the lack of force feedback from the physical buttons. The ease of use might have been impacted by the latency in target tracking.

A few participants expressed complaints regarding the tracking latency, image quality, and FOV of the mixed prototype after the experiment. Additionally, some participants

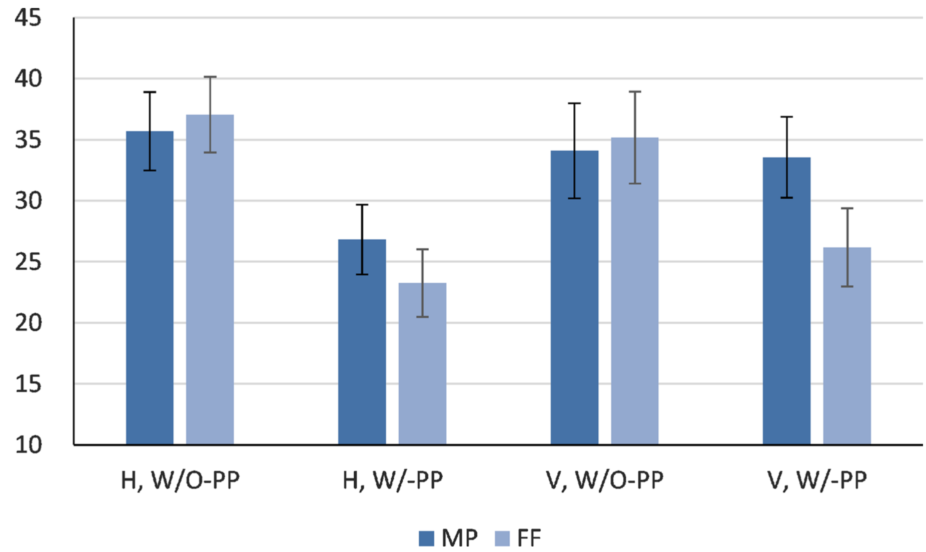


Fig. 18 Workload under different conditions (*H* horizontal, *V* vertical, *W/O-PP* without positioning plate, *W/-PP* with positioning plate). Error bars represent ±1 standard error of the mean

Table 4 Mean SUS scores of the handheld printer using the mixed prototype and the full-fidelity prototype

Group	Usability		Learnability		SUS score
	Mean	95% CI	Mean	95% CI	Mean
MP	72.46	79.46	66.41	78.98	71.25
		65.47		53.83	
FF	74.23	81.29	75.78	89.21	74.53
		67.18		62.36	

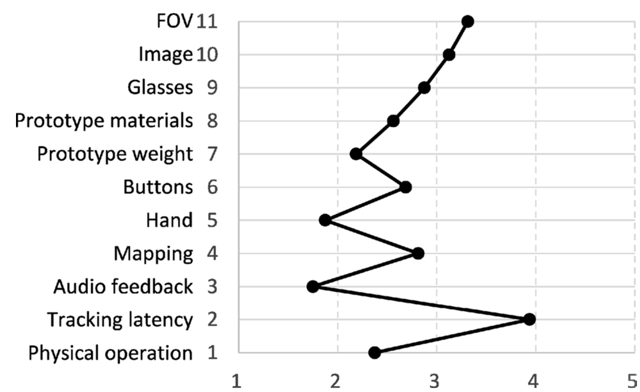


Fig. 20 Assessment of the fidelity factors of the mixed prototype affecting user performance. 1–5: very little impact–very big impact

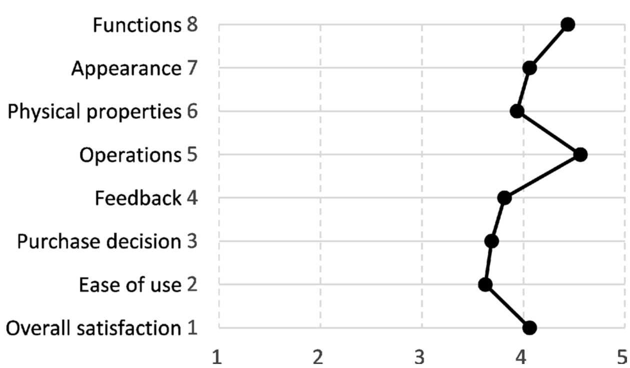


Fig. 19 Satisfaction with the mixed prototype. 1–5: very dissatisfied-very satisfied

mentioned that they occasionally touched the electrodes by mistake, which reduced the ease of use of the mixed prototype. The painted electrodes were employed to enhance the flexibility of the mixed prototype. Therefore, it is crucial to carefully differentiate between usability issues caused by the prototype itself and those related to the product whenever simplification in prototypes may introduce additional problems. One participant mistakenly perceived the cuboid tracking target as part of the product, highlighting the need for clear explanation whenever any extra component is added for prototyping purposes. Participants preferred the sound effects associated with physical manipulation, as they believed it had a positive impact on their performance.

3.4.3 Fidelity

Figure 20 shows the assessment results of the fidelity factors that affect user performance. The mean scores for tracking latency, FOV, and image quality were above 3.0, which were the top three fidelity concerns. When the mockup was quickly moved or rotated, occasionally the virtual printer would be displaced from the mockup due to the marker tracking latency. This could affect user performance in cognitive and physical interactions, as well as overall user experience. The mean scores of glasses weight, spatial mapping, buttons, and prototype materials ranged of 2.5–3.0. The means scores of the remaining factors were all below 2.5. Hand occlusion, which is typically considered a critical issue that requires addressing through techniques like Chroma key and hand-mask in previous studies (Bruno et al. 2013; Faust et al. 2019), had minimal impact in this experiment. This can be attributed to the see-through feature and adjustable display brightness of HoloLens. Furthermore, we gathered information on other issues that could affect user performance through an open-ended question. Two participants mentioned that the virtual image appeared slightly blurry even when they were wearing their myopia glasses. One participant complained about the cuboid marker obscuring the scene behind it.

4 Discussion

According to our findings, mixed prototyping has proven to be an effective tool for learnability evaluation. It helped experimenters identify almost all learnability issues that could be encountered with the actual product. However, depending on the fidelity level, it may introduce additional issues, particularly for tasks involving physical interaction. Participants experienced more difficulties when using the mixed prototype during their initial attempt to interact with the product, which aligns with the subjective measure of the learnability subscale of SUS. Therefore, it is likely to underestimate the learnability of the product when using mixed prototypes. These findings are consistent with previous studies (Choi 2019; Zhang and Choi 2015). For instance, Choi (2019) employed augmented reality (AR) and tangible augmented reality (TAR) for product usability assessment. The ease of learning scores of the AR and TAR presentations were lower than those of the actual product, although the difference between the TAR presentation and the actual product was not significant. The higher learnability scores of the TAR presentation can be attributed to its higher fidelity level compared to the AR presentation.

Cognitive interaction tasks were employed to investigate whether the prototype could accurately reflect performance differences across tasks of varying complexity. The results of the comparative experiment indicated that the prototype effectively evaluated changes in performance due to task complexity, as measured by task completion time and success rate. Our findings are consistent with previous research conducted by Faust et al. (2019), who observed a proportional relationship between task difficulty level and task completion time, as well as the number of errors and withdrawals. In our study, user performance was slightly reduced when using the prototype, resulting in higher workload for participants due to lower performance and increased task failures. The effects of auxiliary function and working posture were examined using physical interaction tasks. Usability test results demonstrated that the experimental factors had consistent impacts on user performance in both the MP group and the FF group. Additionally, the use of a positioning plate significantly improved positional accuracy in both horizontal and vertical conditions when using the mixed prototype. However, the absolute value of the performance obtained using the mixed prototype was significantly lower compared to the full-fidelity prototype. This difference in performance could be attributed to the fidelity level of the mixed prototype or the presence of systematic errors. According to Cox et al. (2022)'s fidelity taxonomy for mixed reality prototypes, operational fidelity is likely the primary influencing factor.

No significant differences were observed between the two groups in the subjective usability assessment, which is consistent with the findings of Choi (2019)'s study using the USE questionnaire (Lund 2001). However, it is important to consider the conditions under which usability testing is conducted when using mixed prototypes, as the conclusions drawn may be influenced. For instance, the presence of the auxiliary function significantly reduced the user's workload when printing on the horizontal plane, but the reduction in workload was negligible when printing on the vertical plane. This suggests that the effectiveness of the auxiliary function may be underestimated when testing in the vertical condition. To further analyze the cause of this deviation in the MP group, the subscales of the NASA-TLX were examined. Only MD showed a significant decrease of 5.7 ($p < 0.05$), while no significant decreases were observed in PD, OP, and FR. In contrast, TD and EF showed a slight increase. This indicates that the auxiliary function primarily helped reduce the mental demand associated with estimating the printing position. However, participants needed to invest more time and effort in using the mixed prototype while holding the positioning plate (no need to hold all the time for the horizontal condition). As mentioned by Kent et al. (2021), the paradox of technology (Norman 2013) must be given substantial consideration when exploring opportunities to create and integrate novel interfaces. Designers and developers should minimize extra workload for users when creating mixed prototyping systems for design evaluation purposes. It is crucial to choose appropriate test conditions to minimize the impact of new technologies on usability testing practices.

There are limitations in obtaining experimental data from the video captured with HoloLens. The mixed reality video combines the holographic output from the right eye with the photo/video (PV) camera. It has been observed that there is an offset between the rendered image and the real object due to parallax. As a result, the captured video was not exactly what the user saw. Fortunately, the deviation has been found to be relatively constant and within a certain range based on our repeated observations with different users in our pilot test. To address this issue, the estimated deviation value was subtracted from the vertical offset value measured using the captured video. The measurement error was within 1 grid unit (2 mm). Furthermore, in order to minimize errors caused by parallax, the vertical position of the printing target area was adjusted for each participant to ensure their eyes were perpendicular to the wall when completing the printing task. Additional efforts were made to minimize systematic errors. For instance, the HoloLens was calibrated for each participant before the printing test in order to ensure accurate hologram display based on their interpupillary distance (IPD). This calibration helped set the IPD correctly and prevented holograms from appearing unstable or at incorrect distances.

We chose the handheld printer as the subject of our study to develop a mixed prototype and conduct usability assessment. The proposed mixed prototyping solution and its findings are most applicable to handheld products like this. For other large products and equipment, further studies will be necessary to explore mixed prototyping solutions that effectively balance cost and fidelity while maintaining flexibility for iterative design. Furthermore, it is crucial to assess the effectiveness in terms of usability evaluation.

5 Conclusion

This study demonstrated the rapid development of a flexible mixed reality prototyping system for a handheld product. The combination of 3D printing, interactive 3D simulation, electronic prototyping platform, and MS HoloLens makes it convenient to make iterative changes to prototypes. To evaluate the effectiveness of the mixed prototype in usability assessment, we conducted a comparative experiment of usability testing. It has been proved that usability testing results using the mixed prototype accurately reflected changes in user performance across different task complexities, functional attributes, and physical contexts. However, it is important to note that the absolute value of performance obtained from usability testing using mixed prototypes may deviate from the true value. In conclusion, mixed prototypes are more suitable for comparing the usability of different design alternatives under different conditions rather than obtaining an absolute measure of usability. Nevertheless, subjective evaluation results obtained from usability testing using mixed prototypes are generally acceptable. The absolute value of usability ratings tends to be slightly lower compared to testing with a real product depending on the fidelity level of the mixed prototype. A novelty of this study is the demonstration of the application of mixed reality prototyping for handheld products to facilitate the usability assessment associated with physical manipulation and interaction with real environments beyond the scope of user interfaces. As another innovative aspect, this study revealed how factors such as product function attributes, test task selection, and physical contexts influence the results of usability assessments conducted with the mixed prototype.

The mixed reality device utilized in this study is the first-generation HoloLens, which is recognized as the world's first self-contained holographic computer based on the mixed reality platform. This device does have certain limitations in terms of system performance, including FOV, resolution, processor capacity, etc. However, with the advancements in technology, the performance of the latest generation of devices has been greatly improved. Hence, image quality and target tracking latency are expected to be enhanced with the new generation of HoloLens, which is beneficial for the efficient implementation of mixed prototyping systems and for reducing deviations in usability evaluation. Consequently, this study establishes a significant theoretical foundation for product design assessment utilizing mixed prototypes, while providing practical guidance to designers and developers regarding the evaluation of product usability using mixed prototypes.

Acknowledgements The authors are grateful to the participants for their contributions to the research. We also want to thank the anonymous reviewers for their valuable comments. This work was supported by "Research Project on Graduate Education and Teaching Reform at Nanjing University of Aeronautics and Astronautics" and "Humanities and Social Sciences Project of the Ministry of Education of China."

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflicts of interest.

References

Akoğlu C, Er HA (2015) The role of interaction design in smart product development activities. Online J Art Des 3(4):51–65. https://eresearch.ozyegin.edu.tr/handle/10679/6376

Bordegoni M, Rizzi C (2011) Innovation in product design: from CAD to virtual prototyping. Springer, Berlin

Brooke J (1996) Sus-a quick and dirty usability scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B (edS) Usability evaluation in industry. Taylor and Francis, London, pp 189–194

Bruno F, Muzzupappa M (2010) Product interface design: a participatory approach based on virtual reality. Int J Hum Comput Stud 68(5):254–269. https://doi.org/10.1016/j.ijhcs.2009.12.004

Bruno F, Angilica A, Cosco F et al (2013) Reliable behaviour simulation of product interface in mixed reality. Eng Comput 29(3):375–387. https://doi.org/10.1007/s00366-012-0293-7

Camburn BA, Sng KH, Perez KB et al (2015) The way makers prototype: principles of diy design. In: International design engineering technical conferences and computers and information in engineering conference. American Society of Mechanical Engineers, p V007T06A004

Camburn B, Viswanathan V, Linsey J et al (2017) Design prototyping methods: state of the art in strategies, techniques, and guidelines. Des Sci 3:e13. https://doi.org/10.1017/dsj.2017.10

Choi YM (2019) Applying tangible augmented reality for product usability assessment. J Usability Stud 14(4):187–200

Cox C, Hicks B, Gopsill J (2022) Improving mixed-reality prototyping through a classification and characterisation of fidelity. Proc Des Soc 2:353–362. https://doi.org/10.1017/pds.2022.37

Faust FG, Catecati T, Souza Sierra I et al (2019) Mixed prototypes for the evaluation of usability and user experience: simulating an interactive electronic device. Virtual Real 23(2):197–211. https://doi.org/10.1007/s10055-018-0356-1

Ferrise F, Graziosi S, Bordegoni M (2017) Prototyping strategies for multisensory product experience engineering. J Intell Manuf 28(7):1695–1707. https://doi.org/10.1007/s10845-015-1163-0

Hallgrimsson B (2012) Prototyping and modelmaking for product design. Hachette UK

Hart SG, Staveland LE (1988) Development of nasa-tlx (task load index): results of empirical and theoretical research. In: Advances in psychology, vol. 52. North-Holland, Oxford, pp 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Kent L, Snider C, Gopsill J et al (2021) Mixed reality in design prototyping: a systematic review. Des Stud 77:101,046. https://doi.org/10.1016/j.destud.2021.101046

Lewis JR, Sauro J (2009) The factor structure of the system usability scale. In: Kurosu M (ed) Human centered design. Springer, Berlin, pp 94–103

Lewis JR, Sauro J (2021) Usability and user experience: design and evaluation, Chap. 38. Wiley, Hoboken, pp 972–1015. https://doi.org/10.1002/9781119636113.ch38

Lund AM (2001) Measuring usability with the use questionnaire. Usability Interface 8(2):3–6

MagicLeap (2023). http://www.magicleap.com

Majrashi K, Al-Wabil A (2018) Hci practices in software-development environments in Saudi Arabia. In: Rau PLP (ed) Cross-cultural design. Methods, tools, and Users. Springer, Cham, pp 58–77

Maurya S, Arai K, Moriya K et al (2019) A mixed reality tool for end-users participation in early creative design tasks. Int J Interact Des Manuf: IJIDeM 13:163–82

Microsoft (2023) https://www.microsoft.com/en-us/hololens

Morozova A, Rheinstadter V, Wallach D (2019) Mixedux: a mixed prototyping framework for usability testing in augmented reality. In: DIS‘19 companion: companion publication of the 2019 ACM designing interactive systems conference, Assoc Comp Machinery; Adobe; Google; UC San Diego Design Lab; Virginia Tech, Sketch, pp 41–44. https://doi.org/10.1145/3301019.3325146

Nielsen J (1994) Usability engineering. Morgan Kaufmann, Burlington

Norman D (2013) The design of everyday things: revised and, expanded. Basic Books, New York

Norman DA, Draper SW (1986) User centered system design: new perspectives on human–computer interaction. CUMINCAD

Park H, Moon HC (2013) Design evaluation of information appliances using augmented reality-based tangible interaction. Comput Ind 64(7):854–868. https://doi.org/10.1016/j.compind.2013.05.006

Park H, Son JS, Lee KH (2008) Design evaluation of digital consumer products using virtual reality-based functional behaviour simulation. J Eng Des 19(4):359–375. https://doi.org/10.1080/09544820701474129

PTC (2023) https://www.ptc.com/en/products/vuforia

Salinas E, Cueva R, Paz F (2020) A systematic review of user-centered design techniques. In: Marcus A, Rosenzweig E (eds) Des User Exp Usability Interact Des. Springer, Cham, pp 253–267

Stevens J (1996) Applied multivariate statistics for the social sciences. Lawrence Erlbaum Associates, Mahwah

Verlinden J, Van Den Esker W, Wind L et al (2004) Qualitative comparison of virtual and augmented prototyping of handheld products. In: DS 32: proceedings of DESIGN 2004, the 8th international design conference, Dubrovnik, Croatia

Vredenburg K, Mao JY, Smith PW et al (2002) A survey of user-centered design practice. Association for Computing Machinery, New York, NY, USA, CHI ’02, pp 471–478. https://doi.org/10.1145/503376.503460

Wall MB, Ulrich KT, Flowers WC (1992) Evaluating prototyping technologies for product design. Res Eng Des 3(3):163–177. https://doi.org/10.1007/BF01580518

Wang GG (2002) Definition and review of virtual prototyping. J Comput Inf Sci Eng 2(3):232–236. https://doi.org/10.1115/1.1526508

Zhang X, Choi YM (2015) Applying tangible augmented reality in usability evaluation. In: Shumaker R, Lackey S (eds) Virtual, augmented and mixed reality. Springer, Cham, pp 88–94

Zhou X, Rau PLP (2019) Determining fidelity of mixed prototypes: effect of media and physical interaction. Appl Ergon 80:111–118. https://doi.org/10.1016/j.apergo.2019.05.007