Mobile Augmented Reality Design Evaluation

Steven Ford[1], Michael Miuccio, Luke Shirley, Fiona Duruaku, & Pamela Wisniewski
University of Central Florida

Augmented reality (AR) usage has increased substantially in recent years due to the proliferation of smart phone technology. Retail companies have leveraged augmented reality by allowing users to view products in their space before they buy. Because usability is related to return on investment (Marcus, 2005; Nielsen et al., 2013), ensuring the usability of these products is vital. One way to accomplish this is through heuristic analyses (Tan et al., 2009). While there are many time-tested sets of heuristics (Nielsen, 1994; Schneiderman & Plaisant, 2004), these often fail to address many of the interactions that are unique to AR (Derby & Chaparro, 2021). In this paper, we demonstrate the viability of supplementing traditional heuristic analyses (Nielsen, 1994) with a new heuristic checklist designed specifically for augmented and mixed reality (Derby & Chaparro, In Press) by evaluating a retail AR program.

INTRODUCTION

Augmented reality (AR) applications have gained increasing popularity in recent years. Since the inception of modern, augmented reality in the early 1990s (Rosenberg, 1992), usage has skyrocketed, due in large part to the proliferation of smart-phone technology. This technology has not only enabled high-profile social and gaming applications such as Snapchat filters and Pokémon Go but has also ushered in a breadth of other use cases. These include a variety of areas such as vehicle heads up displays, navigation, education, healthcare, and military operations (Snap Inc, 2022; Niantic, 2022; Google LLC, 2022; Bach, 2021; Tang et al., 1998; Antonioli et al., 2014; Maroto et al., 2018).

Of particular interest is the adoption of mobile AR by retail companies. Many retail companies have successfully leveraged AR by allowing users to virtually sample an item. Users are often able to tailor their experience by changing colors and other variables which enables them to try out a virtual version of the product before purchasing. This process makes the purchase seem less risky (Kim & Forsyth, 2008). Adding this type of AR functionality to traditional, retail practices enhances user satisfaction and willingness to buy (Poushneh & Vasquez-Parraga, 2017) because customers feel more certain about their decision (Oh et al., 2008).

In the retail setting, if the goal of an application is to enhance customer satisfaction and generate revenue, then the user experience must be at the forefront of the development process. User interface usability is often strongly related to satisfaction (Bevan & Mcleod, 1994) and return on investment (Marcus, 2005; Nielsen et al., 2013). Therefore, ensuring usability should be a primary goal of developers hoping to leverage AR in a retail context.

A common way of assessing the usability of products is through heuristic analyses. Heuristic analyses are a method of usability inspection that relies on practitioners evaluating interfaces based on their expertise with a set of user-centered design principles, rather than gathering information from actual users (Tan et al., 2009). While heuristic analyses often work best in conjunction with traditional usability testing (Nielsen & Mack, 1994; Ford et al., 2021), heuristic analyses tend to uncover more unique usability problems than usability testing (Jeffries et al., 1991).

Although undoubtably beneficial, difficulties arise in applying traditional heuristic analyses to AR applications. For example, both Nielsen's 10 Usability Heuristics for User Interface Design (Nielsen & Molich, 1990; Nielsen, 1994; 2006) and Schneiderman's Eight Golden Rules of Interface Design (Schneiderman & Plaisant, 2004) fail to address many interactions that are unique to AR, such as integrating physical and virtual objects (Derby & Chapparo, 2021). Considering that these are the most widely adopted sets of heuristics for design evaluation, this highlights a clear deficit in the practice of heuristic evaluation. To address this, several research groups have developed heuristics and design principles specifically for AR (Ko et al., 2013; Franklin et al., 2014; de Paiva Guimarães & Martins, 2014; Gale et al., 2015; Kalalahti, 2015; Santos et al., 2016; Endsley et al., 2017; Aultman et al., 2018)by adapting traditional heuristics to the AR environment (Labrie & Cheng, 2020). Most recently, Derby and Chaparro (In Press) developed and validated a checklist specifically for augmented and mixed reality. Checklists ensure that each evaluator uses the same criteria and uncovers a higher volume of problems than traditional heuristic evaluation (Khajouei et al., 2018).

PRACTICE INNOVATION

The current project assesses the usability of a mobile AR application developed for American Plant Exchange in Clearwater, Florida, while also contrasting a checklist based on the classic Nielsen's 10 Usability Heuristics for User Interface Design (Batagoda, 2019; Nielsen, 1994; 2020) with Derby and Chaparro's Heuristic Evaluation Checklist for Augmented Reality and Mixed Reality applications (Derby & Chaparro, In Press).

Technology

The current evaluation focused on the web-based AR program developed for American Plant Exchange. American Plant Exchange started off as a wholesale plant distributor that sold primarily to nurseries and large government projects. In recent years, however, they have made a name for themselves as one of the first companies to sell plants online via Amazon, Etsy, and their own website. Their AR program is accessed via the American Plant Exchange retail website

(https://americanplantexchange.com/) and was designed so that customers can view plants in their space before they buy. The AR program allows users to select and inspect a plant from different angles. They then can use their tablet or mobile device to project the plant in their surrounding area to see how the plant would look in different locations of their choosing.

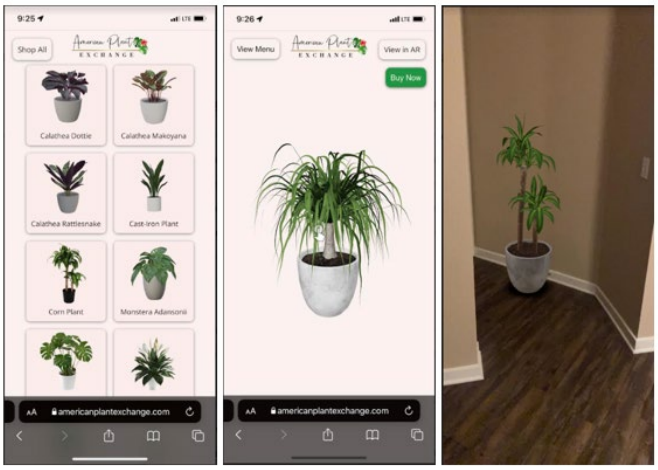

Figure 1. American Plant Exchange Augmented Reality Plants

Heuristic Evaluation Checklists

Checklist Based on Nielsen's 10 Usability Heuristics.

Nielsen's 10 Usability Heuristics for User Interface Design (Nielsen, 1994; 2020) are among the most widely adopted for heuristic evaluations. The checklist used for this study was developed by Muditha Batagoda (2019). Nielsen's 10 Heuristics are as follows:

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation

The checklist breaks down these heuristics into 123 individual checklist items, posed as questions, that can be rated as "yes", "no", or "not applicable." However, in order to match this checklist as closely as possible to Derby and Chaparro's (In press) checklist, we included the option to use "Somewhat" as a rating. Additionally, space is given to include comments about each rating.

Heuristic Evaluation Checklist for Augmented Reality and Mixed Reality.

To address the unique usability issues found in augmented and mixed reality applications and devices, Derby & Chaparro (In Press) developed a set of 11 heuristics (94 checklist items). The evaluator is given the option to omit any heuristics that are not applicable to the application or device being assessed. The included heuristics are:

1. Unboxing & Setup
2. Help & Documentation
3. Cognitive Overload
4. Integration of Physical and Virtual Worlds
5. Consistency & Standards
6. Collaboration
7. Comfort
8. Feedback
9. User Interaction
10. Recognition Rather than Recall
11. Device Maintainability

The checklist includes examples of each checklist item and summarizes the data by tallying up ratings on for each heuristic and graphing the results. The possible ratings are "Yes", "No", "Somewhat", and "N/A".

Procedure

Four researchers evaluated the American Plant Exchange AR platform using Derby and Chaparro's (In Press) Heuristic Evaluation Checklist for Augmented Reality and Mixed Reality and Batagoda's (2019) checklist based on Nielsen's 10 Usability Heuristics (Nielsen, 1994; 2020). The number of evaluators is consistent with Nielsen and Mack's (1994) recommendation of using 3-5 evaluators with general usability experience. After familiarizing themselves with the platform, each researcher used both checklists to indicate the validity of each checklist item (ex. "For mobile devices are both landscape and portrait mode supported?" and "can users easily reverse their actions?"). For each checklist item, the researchers had the option to input any comments about the rating. When completed, the data was aggregated and reported. Finally, the researchers were invited to provide comments comparing each of the heuristic checklists.

FINDINGS

Batagoda (2019) Checklist

The Batagoda (2019) checklist based on Nielsen's 10 Usability Heuristics uncovered a total of 22 usability issues, and another 11 potential issues. To visualize the results, we generated a graph similar to the one automatically generated in Derby and Chaparro's (In Press) checklist (see Figure 2).

It is worth noting that there were several issues with the Batagoda (2019) checklist itself, ranging from nuisance to critical. The most critical problem with the checklist itself is that it does not accurately represent the 10 heuristics, which makes it difficult to organize usability issues by heuristic. For example, flexibility and efficiency of use questions are incorporated along with aesthetic and minimalist design questions and listed under one heading. Similarly, error prevention is included in the help users recognize, diagnose, and recover from errors section. Finally, it does not address help and documentation. Minor errors are presented in the discussion section below.

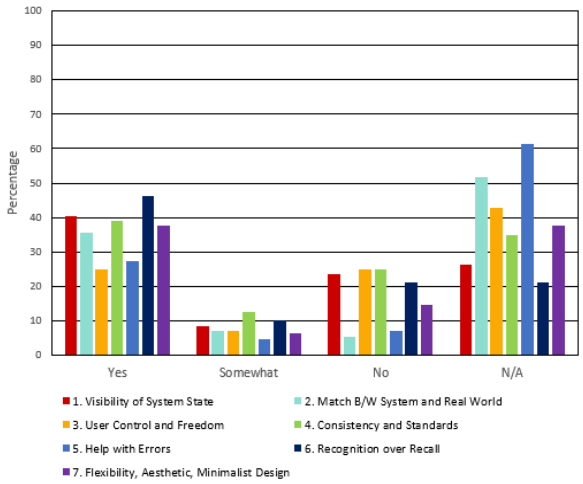

Figure 2. Percentage by heuristic for Batagoda's (2019) checklist

Some highlighted findings from the Batagoda (2019) checklist are listed below. Heuristics are listed in the manner they appear in the checklist.

Visibility of system state.
The operational state of the program should be apparent to the user at all times (Nielsen, 1994; 2020). There were several issues uncovered related to visibility of system state. When viewing the 3D object and while in AR, there is no title or header. This would be helpful, as non-experts would not know what plant they are examining. It is also difficult to tell where the plant will be placed when first opening the AR application.

Match between system and the real world. The system should follow real-world conventions, rather than creating new conventions (Nielsen, 1994; 2020). This is one area that the app excels, primarily because it uses conventions followed by other phone applications. For instance, the snapshot button is a shutter logo similar to that used in the iPhone camera application.

User control and freedom. Once a plant is placed in the environment, users are unable to undo that placement and revert to a previous location.

Consistency and standards. Features that perform similar roles should avoid using different terms or icons (Nielsen, 1994). In this case, the button width changed on some pages. This could introduce the possibility of selection errors.

Help users recognize, diagnose, and recover from errors.
Errors should be prevented when possible, but when errors occur, error messages should indicate the problem and offer a solution (Nielsen, 1994). Because of the simplicity of the current program there are few chances for errors. However, if the user places the plant in the wrong location, there is no way to reverse that error.

Recognition rather than recall.
Users should rely on recognition processes while navigating the interface, rather than being forced to remember the location of items (Nielsen, 1994). As mentioned before, there is no label for the plant when viewing the 3D image or when using AR. This requires the user to remember which plant they are viewing. Because the naming conventions of plants can be difficult for novices, this could be problematic and require extra steps to verify which plant they are examining.

Flexibility, Aesthetic, and Minimalist Design.
Flexibility refers to implementing features that speed up efficiency for experienced users (Nielsen, 1994; 2020). Likewise, interfaces should limit irrelevant information. There are currently no features that speed up efficiency for experienced users. However, the design does limit the use of irrelevant features.

Derby & Chaparro (In Press)

Derby and Chaparro's (In Press) checklist for usability uncovered 62 usability issues and another 47 potential issues. The checklist automatically generates graphs to visualize the percentage of issues per heuristic (see Figure 3). Highlights from selected heuristics are presented below.

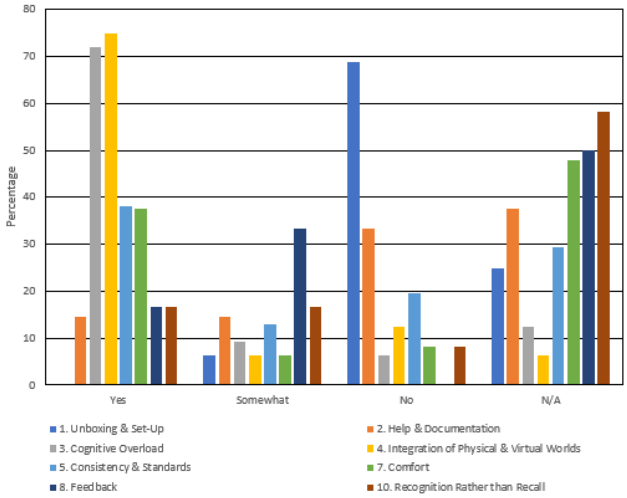

Figure 3. Percentage by heuristic for Derby & Chaparro's (In Press) checklist.

Help and documentation. Help and documentation accounts for items related to learnability and instructions such as an introduction, tutorial, error messages, etc. In the case of the American Plant Exchange program, there are few instructions given or available. While the app is fairly easy to learn, some simple instructions would be helpful, especially for special populations that have limited experience with AR, such as older adults.

Cognitive overload. The cognitive overload heuristic addresses the organization and quantity of information in the system. For instance, developers should avoid clutter, when possible, and organize information in a way that makes sense to the user. This is one area of the system that is executed well, likely because of the simplicity of the program.

Integration of physical and virtual worlds. This heuristic deals with how the virtual items interact with the physical (real) world. In this case, evaluators observed issues with obstruction, where the virtual object does not interact naturally with the real world.

Consistency and standards. Some aspects of this heuristic are not applicable to the present AR program, such as the use of audio. However, some uncovered issues included the impossibility of determining where the item will originally be rendered, no environmental requirements are given to the user (e.g., "avoid bright light"), and the application does not remind users to be aware of their surroundings.

Comfort. The comfort heuristic addresses multiple types of comfort such as physiological comfort, eye strain, and adaptability. While some of the checklist items in this section refer to physical AR devices, several issues were identified. For instance, users are not reminded to take breaks at any point to avoid eye strain and they might have to move quite a bit to see the plant from the desired angle.

User Interaction. The user interaction heuristic addresses both user control and application performance. For example, in the American Plant Exchange program, the user cannot control where the plant is initially rendered. Additionally, there is no way to rotate the plant in AR so it can be seen from different angles.

DISCUSSION

We used two heuristic checklists to evaluate an AR program developed for American Plant Exchange that enables customers to view plants in their environment before they buy. The app is still in early-development stages. We conducted these evaluations for two reasons. The first was to provide actionable feedback to American Plant Exchange so that they may improve the AR program in future iterations. The second was to evaluate the viability of using Derby and Chaparro's (In press) AR/MR checklist to either supplement or replace more traditional heuristics when evaluating AR applications. To match the processes as closely as possible between the two sets of heuristics, we used a checklist based on Nielsen's 10 Usability Heuristics (Batagoda, 2019) and compared that to Derby and Chaparro's (In Press) AR/MR heuristic checklist. To allow for a fair comparison, we adjusted the possible ratings on the Batagoda (2019) checklist to include a "somewhat" option. Following completion of the evaluations, each researcher was asked to compare and contrast the two checklists.

Three out of the four researchers expressed a definitive preference for the Derby and Chaparro (In Press) AR/MR heuristic checklist. Those who preferred this checklist gave several reasons for this. One researcher explained that they felt the questions were more targeted to the product we were testing and that the categories were better defined and separated. All three of these researchers said that they felt the examples and wording were much easier to apply, and one explained that they felt the checklist did exactly what they felt a heuristic checklist should do. One of these researchers also pointed out that the instructions and automatically calculated summary information made the checklist especially user friendly and useful. Also, there were several specific issues with the Batagoda (2019) checklist. There were many grammatical and spelling mistakes within the checklist, and some questions were not designed well. A good example of this is the question "Is the naming conventions are gramatically correct and consistant though the application?" Not only is this question double-barreled, it also contains spelling and grammar errors that impact its interpretability.

The researcher who preferred the Batagoda (2019) checklist said that they felt it was more focused on general usability that was important to the design, while Derby and Chaparro's (In press) AR/MR checklist was too specific and had many items that did not apply to the AR application we examined. This observation also highlights an important point: even though Batagoda's (2019) actual checklist might have been lacking in many ways, the actual heuristics represented in the checklist (Nielsen, 1994) are robust and deserving of their time-tested legacy.

Given the information gleaned from each researcher, a good case can be made that rather than replacing traditional heuristics with new AR specific heuristics, a better strategy might be to supplement traditional evaluations with more AR specific heuristics when applicable. This is especially apparent when considering that the two checklists each revealed unique usability issues, even on overlapping categories. A perfect example of this is that under "Recognition Rather Than Recall", the Batagoda (2019) checklist identified that there are no labels on the AR app or 3d object, so the user is forced to remember the name of the plant they are looking at. For non-experts, these names can be somewhat confusing. On the other hand, the Derby and Chaparro (In press) AR/MR heuristic checklist uncovered many issues specifically related to AR that would not have been addressed using traditional heuristics.

PRACTITIONER TAKE-AWAYS

- Mobile AR applications often contain features outside the scope of traditional heuristic evaluations.
- Derby and Chaparro's (In Press) Heuristic Evaluation Checklist for Augmented Reality and Mixed Reality is useful as a supplement to traditional heuristic analyses when evaluating AR applications.
- Checklists are a way of ensuring consistency in criteria used to evaluate AR systems.

REFERENCES

Antonioli, M., Blake, C., & Sparks, K. (2014). Augmented reality applications in education. *The Journal of technology studies*, 96-107.

Aultman, A., Dowie, S., Hamid, N. A. (2018, April). Design Heuristics for Mobile Augmented Reality Game User Interfaces. In Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1–5)

Batagoda, M. (2019). *A New Usability Heuristic Evaluation Checklist*. Medium.

Bevan, N., & Macleod, M. (1994). Usability measurement in context. *Behaviour & Information Technology*, 13(1-2), 132-145.

Guimaraes, M. D. P., & Martins, V. F. (2014, May). A checklist to evaluate augmented reality applications. In *2014 XVI Symposium on Virtual and Augmented Reality* (pp. 45-52). IEEE.

Derby, J. L., & Chaparro, B. S. (in press). The Development and Validation of an AR/MR Usability Heuristic Checklist. *HCII 2022. Lecture Notes in Computer Science*. Springer, Cham.

Derby, J. L., & Chaparro, B. S. (2021). The Challenges of Evaluating the Usability of Augmented Reality (AR). In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 65, No. 1, pp. 994-998). Sage CA: Los Angeles, CA: SAGE Publications.

Endsley, T. C., Sprehn, K. A., Brill, R. M., Ryan, K. J., Vincent, E. C., & Martin, J. M. (2017, September). Augmented Reality design heuristics: Designing for dynamic interactions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 61, No. 1, pp. 2100-2104). Sage CA: Los Angeles, CA: SAGE Publications.

Ford, S., Schotz, J., & Montalvo, F. (2021) Individuation-based usability of a customized e-commerce platform. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 65(1), 908-912.

Franklin, F., Breyer, F., Kelner, J. (2014, May). Usability heuristics for collaborative augmented reality remote systems. In *2014 XVI Symposium on Virtual and Augmented Reality* (pp. 53–62). IEEE.

Gale, N., Mirza-Babaei, P., Pedersen, I. (2015, October). Heuristic guidelines for playful wearable augmented reality applications. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play* (pp. 529–534).

Google LLC. (2022). Google AR & VR. Retrieved from arvr.google.com/ar/.

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. (1991, March). User interface evaluation in the real world: a comparison of four techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 119-124).

Kalalahti, J. (2015). Developing usability evaluation heuristics for augmented reality applications

Khajouei, R., Gohari, S. H., & Mirzaee, M. (2018). Comparison of two heuristic evaluation methods for evaluating the usability of health information systems. *Journal of biomedical informatics*, *80*, 37-42.

Kim, J., & Forsythe, S. (2008). Sensory enabling technology acceptance model (SE-TAM): A multiple-group structural model comparison. *Psychology & Marketing*, *25*(9), 901-922

Ko, S. M., Chang, W. S., & Ji, Y. G. (2013). Usability principles for augmented reality applications in a smartphone environment. *International Journal of Human-Computer Interaction*, 29(8), pp. 501-515

Labrie, A., & Cheng, J. (2020). Adapting usability heuristics to the context of mobile augmented reality. In *Adjunct Publication of the 33rd Annual ACM Symposium on User Interface Software and Technology* (pp. 4-6).

Marcus, A. (2005). User interface design's return on investment: Examples and statistics. In *Cost-justifying usability* (pp. 17-39). Morgan Kaufmann.

Maroto, M., Caño, E., González, P., & Villegas, D. (2018). Head-up Displays (HUD) in driving. *arXiv preprint arXiv:1803.08383*.

Niantic. (2022). Pokémon GO. Retrieved from https://pokemongolive.com/en/.

Nielsen, J. (1994). Usability Engineering. San Diego: Academic Press. pp. 115–148.

Nielsen, J. (2020). Ten usability heuristics. Retrieved from: https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*, 37(4), 66-75.

Nielsen, J., & Mack, R. L. (Eds.) (1994). Usability inspection methods. New York, NY: John Wiley & Sons.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249-256).

Nielsen, J., Berger, J. M., Gilutz, S., & Whitenton, K. (2013). *Return on Investment (ROI) for Usability*.

Oh, H., Yoon, S. Y., & Shyu, C. R. (2008). How can virtual reality reshape furniture retailing?. *Clothing and Textiles Research Journal*, *26*(2), 143-163.

Poushneh, A., & Vasquez-Parraga, A. Z. (2017). Discernible impact of augmented reality on retail customer's experience, satisfaction and willingness to buy. *Journal of Retailing and Consumer Services*, *34*, 229-234.

Rosenberg, L. B. (1992). *The Use of Virtual Fixtures as Perceptual Overlays to Enhance Operator Performance in Remote Environments*. Stanford Univ Ca Center for Design Research.

Santos, C., Miranda, B., Araujo, T., Carneiro, N., Marques, A., Mota, M., & Meiguins, B. (2016). Guidelines for graphical user interface design in mobile augmented reality applications. In *International Conference on Virtual, Augmented and Mixed Reality* (pp. 71–80). Springer, Cham.

Shneiderman, B. & Plaisant, C., (2004). Designing the user interface: Strategies for effective Human-Computer Interaction (4th ed.). Boston, MA: Addison Wesley.

Snap, Inc. (2022). Lens Studio. Retrieved from lensstudio.snapchat.com.

Tan, W. S., Liu, D., & Bishu, R. (2009). Web evaluation: Heuristic evaluation vs. user testing. *International Journal of Industrial Ergonomics*, *39*(4), 621-627.

Tang, S. L., Kwoh, C. K., Teo, M. Y., Sing, N. W., & Ling, K. V. (1998). Augmented reality systems for medical applications. *IEEE engineering in medicine and biology magazine*, *17*(3), 49-58.