# Lessons Learned in Human Evaluation: A Case Study of Augmented Reality in Books

Bruno Hideki Adachi
Universidade Federal de Ouro Preto
Ouro Preto, Brasil
bruno.adachi@aluno.ufop.edu.br

Silvia Amélia Bim
Universidade Tecnológica Federal do Paraná
Curitiba, Brasil
sabim@utfpr.edu.br

Saul Delabrida
Universidade Federal de Ouro Preto
Ouro Preto, Brasil
saul.delabrida@ufop.edu.br

## Abstract

This paper presents an experiment performed to analyze, in the context of Augmented Reality usage combined with physical books, the problems related to usability and ergonomics. Although the intention of this work is not to discuss the results obtained, but to report the process of preparing and executing the experiment, as well as the lessons learned.

## Keywords

Augmented Reality, Evaluation with Humans, Books, Lessons Learned

**How to cite this paper:**
Bruno Hideki Adachi, Silvia Amélia Bim, and Saul Delabrida. 2025. Lessons Learned in Human Evaluation: A Case Study of Augmented Reality in Books. In *Proceedings of ACM IMX Workshops, June 3 - 6, 2025*. SBC, Porto Alegre - RS, Brazil, 4 pages. https://doi.org/10.5753/imxw.2025.8905

## 1 Introduction

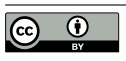
The usage of immersive technologies as Augmented Reality, is widely used in the entertainment field. Although not so much explored and used in other areas that could benefit from the interactive and playful nature that this technology can offer.

In a time where there is rapid information dissemination, the need for instant feedback, and the competition for user attention, books are no longer an option for children and young people. A study conducted by a center for education research, Interdisciplinaridade e Evidências no Debate Educacional (Iede[1]), in conjunction with the reading platform Árvore [2], found that between 66,3% of Brazilian students aged between 15 and 16 years old, the most extensive reading did not exceed 10 pages.

The main objective of this work is to investigate the possibility of using Augmented Reality combined with physical books to make the learning and reading process more exciting and interactive, captivating students into reading more books.

It is important to emphasize that this work is a continuation of an article published in IHC 2023 [3] and does not aim to report the results obtained with the experiment carried out, but rather the lessons learned and the challenges that were identified when planning and executing the experiment.

[1]Interdisciplinarity and Evidence in the Educational Debate



This work is licensed under a Creative Commons Attribution 4.0 International License.
*ACM IMX Workshops, June 3 - 6, 2025*
© 2025 Copyright held by the author(s).
https://doi.org/10.5753/imxw.2025.8905

## 2 Relevance of the Theme

In the following subsections, we will discuss the importance of the use of technology in schools and how it can contribute to a better learning experience. As well as the importance of evaluation with real users.

### 2.1 Relevance of Technology for Teaching

Nowadays, technology plays a fundamental role in people's daily lives, and this is no different in the case of education. Using different technologies in classrooms offers a series of benefits that can transform the way content is taught and how it is received by students [7].

Through computers and mobile devices, teachers and students can access information quickly, ensuring a wide range of educational materials at any time, enriching the learning experience. In addition, tools such as educational games can make the experience more dynamic and engaging, encouraging more active students participation.

The aspects raised previously also favor a more personalized learning experience, where content is taught at the pace and based on the needs of each student, making classrooms a more friendly and inclusive environment.

### 2.2 Relevance of Human-Centered Development

User-centered development uses multiple investigative methods such as interviews and analysis to understand user needs. There are several approaches in the literature that can be used according to each context. In general, the approaches aim to fulfill the following steps: 1) Understand the context of use 2) Specify the requirements 3) Design solutions 4) Evaluate. This process can generally occur with the direct participation of end users in all or most of the steps through Participatory Design [5] or Collaborative Design [9] approaches. Or, users may be involved only in the evaluation phase.

It is important to highlight ethical aspects of research involving human beings. Although each country may have legislation with considerations based on their contexts, these must be guaranteed according to local legislation concerning the people volunteering in the study.

The user-centered development process is relevant because all development time is focused on the user, which opens up opportunities for identifying requirements and solutions, improving usability and user experience once the product is ready.

In this work, we performed the steps of understanding the context, specifying requirements, and designing the solution based

on data collected in the literature. Finally, we presented evaluations made by users of the developed solution and from which we presented lessons learned as the focus of this work.

## 3 Experiment Performed

This section presents the evaluation carried out, as well as the motivation and its objective. This paper main point of discussion are the aspects that should be taken into consideration when carrying out an evaluation with people involving Augmented Reality technologies, as well as the lessons learned. The results will not be discussed in this text.

The experiment carried out in this work had as its main reference a previous study done by Ze Dong et al. [8], where the authors evaluated user interaction with Augmented Reality applications to verify people's perception of different forms of user interaction with such applications. Although the mobile application market offers a wide range of Augmented Reality software, most of them still use a form of interaction focused on 2D applications, through touches on a surface (in this case, the mobile device's screen). The authors then question whether interaction through movement gestures, using the devices' sensors, might be more intuitive and more relevant to AR applications.

As previously stated, this is a continuation of an ongoing work [3]. In the first part a test was conducted, where different AR SDKs were selected and tested in terms of detection and tracking capacity of illustrations in physical books not designed for the use of Augmented Reality. Difficulties were identified for people to interact with the application and the book simultaneously. To investigate the reasons for these difficulties, which may be related to usability and ergonomic factors, we used the [8] protocol (volunteers must perform different tasks, while the observer takes notes about how they act in these various scenarios) in a very similar way.

To this end, we developed an application using the Unity tool combined with the Vuforia SDK, which worked as follows: when pointing the smartphone camera at an illustration from the book we chose as the object of study (Alan Turing: His Machines and Their Secrets [1]), a 3D object appeared (as can be seen in Figure 1) and, with that, the evaluation volunteer had to perform an interaction based on a series of tasks. The creation of this application was intended to facilitate the evaluation, since, in the study by [8], different applications were selected to test different tasks and gestures.

To describe the experiment carried out, two stages will be presented, the first being the preparation phase, where all ethical issues of the research were raised and the experimental design was developed, while the second phase is related to the execution of the evaluation.

### 3.1 Ethical Conduct

Conducting evaluations involving the participation of volunteers must be very well structured, and all risks must be considered [6]. Therefore, before starting the experiment, a series of documents were written and evaluated by the Ethics Committee of the institution in question (in this case, the Universidade Federal de Ouro Preto), following the legislation on research involving human beings in Brazil.

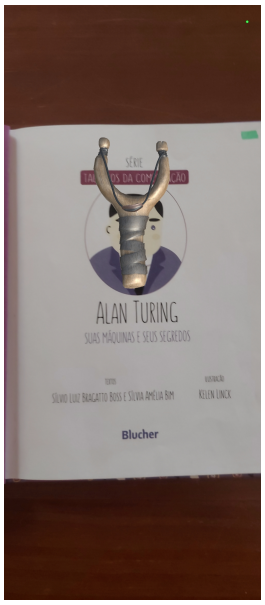



**Figure 1: Detection Screen**

Among the documents evaluated by the Ethics Committee, there are two that are extremely important: the Research Brochure and the Evaluation Project. The Research Brochure is a document that contains information about scientific research and is used by the committee's evaluators to verify the validity of the study that will be conducted. The Evaluation Project is a normative document that must describe the main information about the experiment to be conducted:

- the project design (how the evaluation will take place);
- the hypothesis that the authors intend to answer;
- the primary objective;
- the methodology (the evaluation methods to be used);
- data related to the recruitment of volunteers (criteria for inclusion and exclusion);
- the risks to which participants will be subject;
- the benefits of participating in the research;
- and sample size.

The research documents reported in this article were then submitted to the ethics committee of the Universidade Federal de Ouro Preto. The research was approved under the number of Certificado de Apresentação de Apreciação Ética (CAAE[2]) 76992024.2.0000.5150.

The entire evaluation took place in a testing laboratory within the premises of the Universidade Federal de Ouro Preto (as can be seen in Figure 2). A specific area of the room was used and another area related to evaluations with Virtual Reality technologies was isolated.

The area where the experiments were carried out is part of a room specifically designed for user evaluations, a private space where noise and distractions are avoided. In the case of this study, only the observer and the volunteer were allowed in the room, so

[2]Certificate of Presentation of Ethical Appreciation

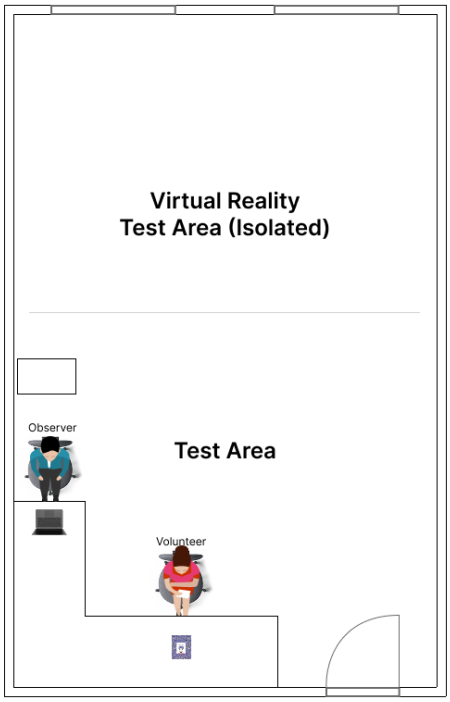



**Figure 2: Top View of Evaluation Room**

that any embarrassment on the part of the volunteer would not be an obstacle, and they could express themselves freely.

The positioning of those involved was also designed with the aim of avoiding possible embarrassment on the part of the volunteer, since direct eye contact could, in some way, inhibit their behavior, and important information for the study could be disregarded.

It is important to emphasize that the observer's point of view allowed all actions performed by the volunteer to be observed, facilitating note-taking.

## 3.2 Call for Volunteers

The next step was to call for volunteers to carry out the experiment; the main method used to do this was sending messages to online student communication groups. At the end of the evaluation, 18 volunteers contributed to the study.

The volunteers were welcomed and the evaluation protocol began, where the reason for conducting the study was explained and how their participation would help the research. Next, the Termo de Consentimento Livre e Esclarecido (TCLE[3]) was presented, which contained relevant information about the evaluation, such as the risks presented and the possibility of the participant withdrawing.

Considering the case study work and how the results would be examined in order to create a better understanding of the problem, it was agreed between the authors to collect data on the experiment volunteers, data observed during the experiment and a general assessment by the volunteers of the proposals presented to try to solve the problems addressed.

Immediately after the acceptance of the volunteer to contribute to the study, the evaluation began. In order to carry out a thorough analysis of the results, we considered it necessary to collect some information about the participants, such as age, gender, level of education, course (if the person was enrolled in higher education), and whether the volunteer was already familiar with Augmented Reality technologies.

After collecting the sociodemographic data, an order in which the tests would be performed was chosen. This was necessary because the evaluation included a total of 6 different scenarios, and if we started all the evaluations with the same scenario, factors such as fatigue could affect the results obtained. Each scenario used different equipment, so when starting a new scenario, the objects were presented to the volunteers, as well as the possibilities of interacting with them. The following objects were used: a fixed and an articulated support for smartphones, and a book support. By combining these objects, we tested the following scenarios: hands-free (the volunteer held the smartphone with one hand and the other held the book), fixed support (the smartphone was fixed to the support), articulated support (the smartphone was held by the support), hands-free and book support, fixed support and book support, and finally, articulated support and book support.

During the evaluation, volunteers were asked to perform a series of tasks and, for each of them, a gesture had to be chosen. And, using the Think Aloud methodology, volunteers also made comments about the scenario in question; this data was filled in a spreadsheet using a laptop by the observer. At the end of each scenario, each person was asked to fill out a System Usability Scale (SUS) form [4] and, upon completing all 6 scenarios, a Technology Acceptance Model (TAM) form [10] and was also asked for each task what their opinion was about the best scenario. The evaluation was then finished.

# 4 Lessons Learned and Challenges Identified

This section presents the main lessons learned and challenges identified with the experiment, aspects that worked, and points for improvement for future studies.

## 4.1 Evaluation Duration

The biggest problem we faced during the study was related to the length of the evaluation. During the preparation phase, we anticipated that each of the six scenarios would last an average of 10 minutes. We were concerned that the process would become tiring and that volunteer fatigue could interfere with the results.

In an attempt to get around this problem, at the beginning of each evaluation, we randomly selected an order for the tests, that is, instead of always starting with the same scenario. In that way, it is possible to verify in the analysis of the results whether, over time, the volunteer's fatigue affected the perception of the different cases.

Another point to be raised is related to the time we allowed the volunteers to familiarize themselves with the application. Since this is a technology that has not yet been explored much, we noticed that some of the volunteers felt somewhat dazzled by the possibilities

[3]Free and Informed Consent Form

of practical applications of the technology. Thus, consuming a good part of the time foreseen for the tests, a possibility that was thought of later was to define a time limit for the "training" of the volunteers and to explain orally and in detail the functionalities of the application.

The point raised above is directly related to our concern about potential volunteer fatigue. We cannot allow the evaluation to go on for an indefinite period of time and allow fatigue to compromise the results obtained.

### 4.2 Materials and Infrastructure

The entire evaluation process took place with only one observer (which in this case has only a few academic experiences in evaluations involving humans, so the lack of experience must be considered) and one volunteer at a time, where the observer was responsible for receiving the volunteer, explaining the reason and objective of the experiment, as well as the TCLE, collecting socio-demographic information, and taking notes on all six different scenarios.

The way the procedure was carried out meant that some information could have gone unnoticed by the observer, since some volunteers made many comments in a short space of time, which made it difficult to take notes. Therefore, one way to get around this problem would be to use cameras or sound recorders to capture all the information given by the volunteer. As well as the presence of more observers on site to not only assist in the process of explaining the experiment, but also to speed up the entire process of filling out basic data, such as the TCLE form and the collection of socio-demographic data.

### 4.3 Volunteers' Characteristics

After the evaluation period, we observed that, in order to carry out a more robust analysis of the results obtained, we could have included in the socio-demographic data form information relevant to the proposed study. Data such as the volunteer's dominant hand, the use of glasses or contact lenses, and other aspects may interfere in some way with the evaluation responses.

It is also important to highlight that the evaluation conducted had the participation of many volunteers belonging to the academic community of the Universidade Federal de Ouro Preto. Considering this, a broader sample and in different contexts (not only university students, different locations, etc.) can bring more insights to the research.

### 4.4 Pilot Tests

All the problems reported in the previous subsections could have been predicted by carrying out more pilot (or experimental) tests, as this is a way to identify errors and improvements in the experiment even before the final evaluation is applied.

The point to be considered is the difficulty of carrying out such pilot tests without employing the participation of potential volunteers for the final experiment. Since one of the problems faced is attracting people and organizing a schedule capable of meeting the needs of both parties involved.

## 5 Conclusions

The study reported in this text, as well as the evaluation carried out, explored in a practical way the main problems faced by users of Augmented Reality technologies. And, despite the specific context of the work (use of AR combined with physical books), the lessons learned here are generally useful for evaluations with people.

Based on our experience, we recommend good experimental planning so that possible errors can be avoided, which could invalidate the results obtained. The main lesson we learned was the need to carry out more experimental tests, since these tests can help identify potential areas for improvement and refine the process in general.

Overall, we addressed the main difficulties in conducting an evaluation involving people and the errors made in the study. Thus, highlighting the need for studies that offer guidelines capable of assisting in the planning of studies of this type. Improving the study area and facilitating the understanding of the importance of the topic.

## Acknowledgments

The content of this work was written by the authors. The authors used AI tools only for grammatical proof.

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) finance code 306101/2021-1, and the Universidade Federal de Ouro Preto (UFOP). The authors wrote the paper, and AI tools were utilized to improve grammar.

## References

[1] 2022. *Alan Turing: Suas Máquinas e Seus Segredos.* https://www.blucher.com.br/alan-turing

[2] 2023. *Estudo do Iede e da Árvore aponta relação do hábito leitor com indicadores educacionais e socioeconômicos.* https://portaliede.org.br/estudo/estudo-do-iede-e-da-arvore-aponta-influencia-do-habito-leitor-em-indicadores-educacionais-e-socioeconomicos/

[3] Bruno Hideki Adachi, Júlia Carlos Gonzaga, Paula Cintra Fernandes, Saul Emanuel Delabrida Silva, Silvia Amélia Bim, and Silvio Luiz Bragatto Boss. 2024. Augmented Reality in Books: An Evaluation of Alan Turing Book. In *Proceedings of the XXII Brazilian Symposium on Human Factors in Computing Systems* (Maceió, Brazil) *(IHC '23)*. Association for Computing Machinery, New York, NY, USA, Article 12, 10 pages. doi:10.1145/3638067.3638124

[4] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[5] Bronwyn Cumbo and Neil Selwyn. 2022. Using participatory design approaches in educational research. *International Journal of Research & Method in Education* 45, 1 (2022), 60–72.

[6] Maria Candida Soares Del-Masso, Maria Amélia de Castro Cotta, and Marisa Aparecida Pereira Santos. 2014. Ética em pesquisa científica: conceitos e finalidades. *São Paulo: UNESP* (2014).

[7] Eleni Demitriadou, Kalliopi-Evangelia Stavroulia, and Andreas Lanitis. 2019. Comparative evaluation of virtual and augmented reality for teaching mathematics in primary education. *Education and Information Technologies* 25, 1 (Jul 2019), 381–401. doi:10.1007/s10639-019-09973-5

[8] Ze Dong, Jingjing Zhang, Xiaoliang Bai, Adrian Clark, Robert W. Lindeman, Weiping He, and Thammathip Piumsomboon. 2022. Touch-Move-Release: Studies of Surface and Motion Gestures for Mobile Augmented Reality. *Frontiers in Virtual Reality* 3 (2022). doi:10.3389/frvir.2022.927258

[9] Sharon Friesen and Michele Jacobsen. 2021. Collaborative Design of Professional Graduate Programs in Education. *International Journal of Designs for Learning* 12, 1 (2021), 64–76.

[10] Qingxiong Ma and Liping Liu. 2005. *The Technology Acceptance Model.* doi:10.4018/9781591404743.ch006.ch000