




Augmented Reality Books: A User Experience Evaluation

Joana Casteleiro-Pitrez^(✉) 

Universidade da Beira Interior, LabCom-Comunicação e Artes, Covilhã, Portugal
joana.casteleiro.ferreira@ubi.pt

Abstract. The rise of information society and the profound mediatization phenomenon led to a book use decrease. On the other hand, the attempt to promote book use and widen the capacities of the traditional book led to the application of computational resources, one of these resources is Augmented Reality (AR).

This study is about AR published Books and their User Experience (UX). We want to know if publishing AR books shows sufficient UX. We chose three AR books published in Portugal and applied a quantitative empirical study to evaluate their UX and answer this research question. Sixty users read the AR books and then answered the User Experience Questionnaire (UEQ) used as the research instrument of this study. We also compare the results with a benchmark. The results show a positive evaluation in almost all the means of the scales of the three AR books. Only in one book do we find a neutral evaluation in the scales of Stimulation and Novelty. Regarding the benchmark, the scales Attractiveness, Novelty and Stimulation achieved Excellent/Good category levels. On the other hand, the pragmatic quality aspects like Perspicuity, Dependability, and Efficiency reveals a below “Good category” level, emphasizing the urgent need to improve the UX of AR books.

Keywords: Augmented reality books · User experience · User experience questionnaire

1 Introduction

The publishing market has been facing a sector crisis for some years now. Faced with the emergence of more attractive forms of leisure and learning arising from technological media coverage, the use of books decreased. This trend is particularly noticeable in younger generations. According to the results of the project “What our children read” [1], reading appears as the fifth leisure time activity with 26% of references. Multimedia uses got fourth place with 64%. Trying to remedy the situation, several formats of digital works, such as e-books, app-books, and AR books, have been created and exploited. All these products use computational resources to promote reading for the alpha generation.

The publishing sector has also encouraged these new editorial experiences by instituting prizes for digital books, such as the Bologna Ragazzi Digital Award promoted by the Bologna Children’s Book Fair or the UKLA Digital Book Award, among others. Given this scenario, AR brings new possibilities to the publishing market, allowing, on

the one hand, to maintain the traditional book format and, on the other hand, enabling the superposition of digital interactive layers over the book. Nevertheless, is this enough to increase book use? Do the users have a good experience using this product? The objective of this study is precisely to understand the UX in AR books.

The UX in AR books is often neglected due to the lack of integration between the editorial team and the AR content team (frequently contracted in an outsourcing mode). Also, the ease of developing an application using AR technology, especially by non-professional developers/designers, results in inadequate/low UX applications, which affects the overall quality of the developed applications [2]. Additionally, different practitioners have also pointed out that the research community lacks an understanding of how well AR applications have been accepted by the end-users, what kind of UX they evoke, and what the users perceive as the strengths and weaknesses of AR applications overall [3]. Thus, it is no longer sufficient to offer products focused only on novelty or functionality. Instead, users expect to learn how to use the system without much effort, solve their tasks fast and efficiently, and control the system at each point. In addition to these pragmatic, goal-oriented interaction qualities, it is also relevant that the product catches the user's attention and interest and that using the product is exciting. Consequently, hedonic, not directly goal-oriented interaction qualities have to be considered to create a successful product, system, or service [4, 5].

In the face of the situations described above, it is essential to ask, do publishing AR books shows sufficient UX? This is the research question that supports this study. To answer this question and achieve the purposed objective: a) we carried out a literature review in the domains of AR books and UX; b) we applied a quantitative methodology. We chose three of the sixteen AR books published in Portugal and evaluated their UX for the empirical work. The research instrument used was the Portuguese Version of the UEQ [6] because the main goal of the UEQ is to allow a fast and immediate measurement of UX. The UEQ also considers aspects of pragmatic and hedonic quality and results from a careful construction process [4]. To interpret scale results from the UEQ we used the UEQ Analysis Data Tool to analyze each scale item and compare it with a benchmark. Thus, the question of whether an AR book UX is sufficient can be answered by comparing its results to a larger sample of other used products, a benchmark data set. If the AR book scores high compared to the products in the benchmark, this can indicate that users will generally find this book UX satisfactory [7]. We used a sample of 60 users (20 for each book) to provide a stable measurement [8].

2 Literature Review

The vast majority of research in the fields of the augmented book is technological [9–13] or educational [14–18]. The study of UX in AR focuses on systematic literature reviews. [19–21], on mobile AR systems [22], in proposing frameworks [23] and models [24] to improve the end-user experience, and proposals for the improvement of the theoretical foundations [25]. The specific UX study in AR books focuses on understanding users' feedback [26, 28] and revealing issues that can improve AR books [20, 27, 29, 31].

Regarding the technological aspect, several studies present prototypes of AR books. One of the first studies related to the AR book is the study of Rekimoto [9], which in

1998 focuses on developing 2D matrix markers, the superposition of virtual contents, and their application on various products, one of which is a book. Although Rekimoto already had the underlying idea that augmented reality could be applied to editorial products, Billingham, Kato, and Poupyrev [10] present the first prototype of an AR book – the Magic Book.

Even today, the prototyping of augmented books motivates researchers to propose new approaches, such as creating multiple levels of interaction and the possibility of individual or shared reading [11], integration of different types of content such as videos, animations, sounds, and various interfaces, such as gesture-based interfaces [12] or the exploration of other kinds of reading like the syntactical reading experience [13].

Concerning the educational nature, the vast majority of studies aim to assess the learning impact of an AR book or AR textbook [14–17]. For example, a work [18] about students’ beliefs of AR books in teaching reinforced that motivation is a possible advantage over traditional books.

Regarding UX studies, the presented systematic reviews focus on AR applications, not specifically AR books. The work of Swan and Gabbard [19] represents one of the first systematic reviews related to usability/user experience in AR applications in general. The results indicated that three lines caught researcher’s interest:

1. Those that study low-level tasks to understand how human perception and cognition operate in AR contexts.
2. Those that examine user task performance within specific AR applications to understand how AR technology could impact underlying tasks.
3. Those that examine user interaction and communication between collaborating users.

Another systematic review [20] concludes that until 2014 there was an increase in the number of usability studies performed in AR research and a shift towards more studies on handheld displays. However, most of these studies are formal user studies, with little field testing and almost no heuristic evaluations. The most popular experimental task involves filling out questionnaires, which lead to subjective ratings being the most widely used dependent measure. This study suggests increased research opportunities in collaboration, field studies, and a more comprehensive range of evaluation methods. The study of Law and Heinz [21] on usability and UX in AR applications concludes that:

1. There is insufficient grounding in usability/UX.
2. Lack of innovative AR-specific usability/UX evaluation methods and the continuing reliance on questionnaires may hamper the advances of AR educational applications.

Many studies try to understand UX with mobile AR services [22]. For example, Olsson, Lagerstam, Kaˆrkkäˆinen, and Vainio-Mattila’s [22] study concluded that UX of mobile AR services is expected to be multifaceted and affected by various components of the underlying technology. The participants directed a set of expectations towards such services, ranging from proactivity, relevance, and the context sensitivity to social, surprising, immersive, and inspiring elements. Some works [23] proposed frameworks to enhance the UX of mobile AR, others [24] recommended a UX model that can be

implemented to provide an engaging and seamless mobile AR experience for end-users. Another study [25] advocated for theoretical foundations of the UX in Augmented and Mixed Reality and proposed several directions for more scientific research in this regard. These directions involve:

1. A revisiting generic UX theory.
2. Revisiting the theoretical foundations for AR.
3. Structuring design knowledge for the UX of content, devices, interactions, applications, and contexts of use for AR/MR.
4. Practical user studies and controlled experiments for applying UX design knowledge to AR/MR and evaluating users' experience beyond aspects of usability or acceptability of AR/MR technology.
5. Connecting to the XR Access initiative for making VR, AR, and MR more accessible.

In the specific domain of AR books and UX, the first works are limited to collecting the user's opinions, through interviews or observations. Only user feedback is measured in the Grasset, Dünser, and Billinghurst study [26]. The conclusion is that people like to discover the system and interact with the various features. They were particularly amazed by the visual effects and the animations. The Dünser and Hornecker [27] work on AR books supporting reading and working through predefined storybooks reveals that the choice of stories and integrated interactive sequences is essential. Navigation turned out to be an essential issue when combining paper and on-screen elements, in particular if these are not integrated into one visual area and deploy tangible and desktop-based input devices. Gázquez and Castro [28] introduced the AR Book System, an interactive and collaborative application for traditional books augmentation. To validate the proposed system, they designed and conducted an experimental study and obtained very positive feedback from participants confirming the usefulness of the ARBS. Cao and Hou [29] studied AR picture books for children aged 5 to 8. They designed a usability test for their prototype. The results indicated the feasibility of AR books and figured out some issues in the interaction processes. These issues are the weak guidance in the interactive operation, the incomplete essential information in the interface, and children's preference for visible interaction and touch operations. Children are also more sensitive to large objects. Polyzou, Botsoglou, Zygouris, and Stamoulis [30] presented an empirical study of AR-published books for preschool children. The conclusion indicated that AR books are not intuitive for preschool children. These children's motor skills are not yet ready to use the fine touch screen movements needed to handle an AR object, and AR books turned out to be a fun and exciting way to capture preschool children's attention. Children perceived AR books as an excellent way to have fun. Even though the AR book appeared to have a better chance of keeping children's attention for longer, this was not practicable due to the highly delicate screen manipulation limitations imposed. Another study [31] reported a user test with 136 children examining the impact of content length and presentation in a digitally-augmented comic book. The results reveal that authors and designers need to balance physical and digital mediums while designing digital augmentation for comic books. In order to achieve the holistic experience and equal engagement with both mediums, the experience needs to be carefully designed by:

1. Avoiding or reducing duplication, especially when duplication means the reader gets no real benefit from engaging with both mediums.
2. Integrating augmented content only if it has high relevance to a particular frame in the comic book.
3. Using appropriate frequency and length of the augmented content to maintain the interest in both digital and physical medium.
4. Selecting the technology carefully as it still presents an interaction barrier (e.g., the access method).

Our brief literature review shows that the first studies related to AR books and UX focused on hedonic qualities and captured the users' feedback. This feedback is usually positive where the motivation and the capacity of the AR book to capture attention are highlighted. More recent studies focused mostly on pragmatic qualities and some guidelines to improve AR books result from the findings. Still, there is a need for studies that join hedonic and pragmatic qualities.

3 Materials and Methods

3.1 Quantitative Experimental Design

The objective of our study is to research whether three AR-published books have enough UX. Therefore, the research question driving this study is: RQ. Do published AR books have sufficient UX? Furthermore, have a positive evaluation on Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty? Moreover, fit in the category of good compared to the benchmark values?

We elaborated an experiment with three groups of users that read and tested the three selected books to answer the research question. The three groups responded to the UEQ.

3.2 Procedure

We assigned each one of the three books to a group of users. Then, between 8 to 11 November 2021, we ask users to go to a university laboratory where they read and experiment with the AR book assigned to them and, after the experience, respond to the online UEQ questionnaire. The UEQ was accessed easily with a link placed in University Beira Interior Moodle, and Google forms support it.

3.3 Participants

We recruited sixty participants ($N = 60$) voluntarily. The participants were undergraduate students of the 1st Cycle Degree in Multimedia Design, Art Department, University Beira Interior, Portugal. The 60 participants were randomly assigned to a group of 20 users. According to the UEQ handbook, there should be at least 20–30 respondents to get reliable results [8].

3.4 Data Collection

Data were collected through the Portuguese version of the UEQ questionnaire [6]. The UEQ enables the analysis of the entire UX beyond mere usability. The questionnaire considers the respondents' feelings, impressions, and attitudes and creates a format that supports the direct expression of these elements. The scales collect usability aspects (Efficiency, Perspicuity, and Dependability) and UX aspects (Stimulation, Novelty) and thus offer a comprehensive impression of the UX of product/system touchpoints [7]. The UEQ consists of 26 items distributed among six scales. The scales are the following: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. The UEQ uses a 7-point Likert-type scale. The UEQ's items form a semantic differential; two contradictory terms represent every item. Terms are randomly ordered per item; half of the scale items begin with the positive term, and the other half begin with the negative one. UEQ uses a seven-stage scale to decrease the central tendency bias for the types of items. The items are scaled from -3 to $+3$. Accordingly, -3 denotes the most negative answer, 0 a neutral answer, and $+3$ the most positive answer [7]. The reliability and validity of the UEQ was investigated in several studies [4, 8].

3.5 Materials

1) Independent variables

The independent variables of this study were the three AR-published books. The chosen books are “Star Wars Galaxy Explorer Guide” (see Fig. 1) from Leya/D. Quixote, published in 2019. “Frozen II the Ice Kingdom: An Enchanted Adventure” (see Fig. 2) from Leya/D. Quixote, published in 2019 and “Toy Story 4 Woody the Adventure of Augmented Reality” (see Fig. 3) from Leya/D. Quixote, published in 2019.

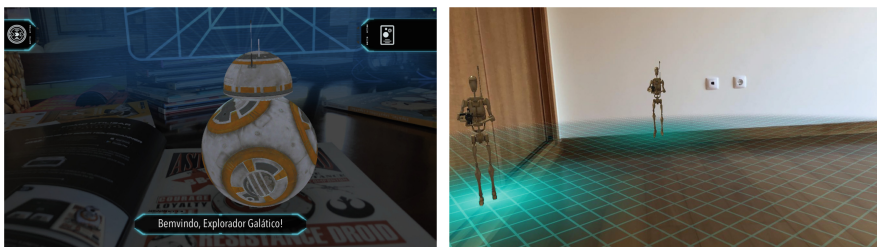


Fig. 1. Star Wars Galaxy Explorer Guide, AR book.

The book *Star Wars Galaxy Explorer Guide* has a hybrid interface. The digital contents are 3D and 2D animations, sounds, videos, and text. The access mode is through an app, and the book is classified as juvenile literature.

Frozen II the Ice Kingdom: An Enchanted Adventure is a juvenile literature book with a hybrid interface. The digital contents are 3D animations and sounds, and an app accesses these contents.



Fig. 2. Frozen II the Ice Kingdom: an enchanted adventure, AR book.

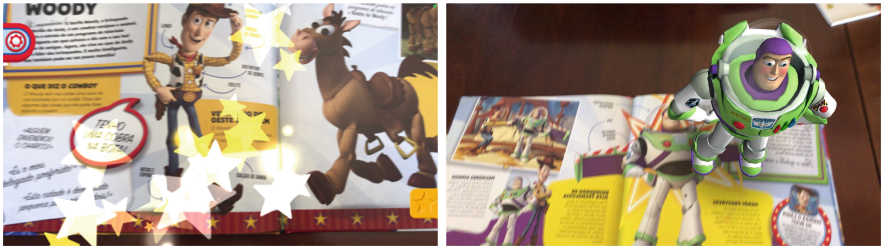


Fig. 3. Toy story 4 woody the adventure of augmented reality, AR book.

Toy Story 4, Woody, the Adventure of Augmented Reality has a hybrid interface. The digital contents are 3D animations and sounds. The access mode is through an app and is classified as children's fiction book.

2) Dependent variables

The dependent variables used in this study were the six scales that consist of the UEQ:

- a) Attractiveness reflects the overall impression of the product and shows whether users like the product.
- b) Efficiency shows whether users can solve their tasks without unnecessary effort.
- c) Perspicuity indicates whether it is easy to become familiar with the product or learn how to use it.
- d) Dependability shows whether users can feel control over the interaction.
- e) Stimulation shows whether it is exciting and motivating to use the product.
- f) Novelty indicates whether the product is innovative and creative and whether the product catches the user's interest.

Attractiveness is a pure liability dimension. Efficiency, Perspicuity, and Dependability are aspects of pragmatic quality, while Stimulation and Novelty are aspects of hedonic quality [4].

3.6 Data Analysis

The quantitative data are obtained from a UEQ. The data was then processed to investigate the UX level of the chosen AR Books. The system’s UX is measured in six scales: Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty. First, we Calculate the level of UX for each scale by processing statistical data using UEQ Analysis Data Tool. Values for the single items between -0.8 and 0.8 represent a neutral evaluation of the corresponding scale, values $> 0,8$ represent a positive evaluation, and values $<-0,8$ represent a negative evaluation (see Table 1). After obtaining the score for each scale, the data is analyzed using a benchmark graph to know the quality of AR Books compared with other products contained in the data set UEQ Analysis Data Tool. The data set contains 21175 persons from 468 studies concerning different products (business software, web pages, webshops, social networks). Next, we consider the benchmark intervals (see Table 2) for the UEQ presented in Schrepp, Hinderks and Thomaschewski [7]. The feedback is limited to five categories:

- Excellent: The evaluated product is among the best 10% of results.
- Good: 10% of the results in the benchmark are better than the evaluated product, 75% of the results are worse than the evaluated product.
- Above Average: 25% of the results in the benchmark are better than the evaluated product 50% of the results are worse.
- Below Average: 50% of the results in the benchmark are better than the evaluated product 25% of the results are worse.
- Bad: The evaluated product is among the worst 25% of results.

To test the viability of the questionnaire scales, we also used Cronbach’s Alpha. A Cronbach’s Alpha between 1–0.90 indicates an excellent internal consistency, between 0.70–0.90 a good internal consistency, between 0.60–0.70 an acceptable consistency, between 0.50–0.60 a poor consistency, and less than 0.50 an unacceptable consistency.

Table 1. Interpretations of the UEQ scales means

Positive evaluation	$>0,8$
Neutral evaluation	-0.8 and 0.8
Negative evaluation	$<-0,8$

Table 2. Benchmark intervals for the UEQ scales (Schrepp, Hinderks & Thomaschewski; 2017b)

	Attractiveness	Efficiency	Perspiciuity	Dependability	Stimulation	Novelty
Excellent	≥ 1.75	≥ 1.78	≥ 1.9	≥ 1.65	≥ 1.55	≥ 1.4
Good	≥ 1.52 < 1.75	≥ 1.47 < 1.78	≥ 1.56 < 1.9	≥ 1.48 < 1.65	≥ 1.31 < 1.55	≥ 1.05 < 1.4

(continued)

Table 2. (continued)

	Attractiveness	Efficiency	Perspiciuity	Dependability	Stimulation	Novelty
Above average	≥ 1.17 < 1.52	≥ 0.98 < 1.47	≥ 1.08 < 1.56	≥ 1.14 < 1.48	≥ 0.99 < 1.31	≥ 0.71 < 1.05
Below average	≥ 0.7 < 1.17	≥ 0.54 < 0.98	≥ 0.64 < 1.08	≥ 0.78 < 1.14	≥ 0.5 < 0.99	≥ 0.3 < 0.71
Bad	< 0.7	< 0.54	< 0.64	< 0.78	< 0.5	< 0.3

4 Results

Considering the completion of the UEQ carried out by 60 participants, we can see that the average age of our sample is 21 years old, with 22 (36.7%) male respondents and 38 (63.3%) female respondents.

4.1 The Star Wars Galaxy Explorer Guide, AR Book

The reliability analysis of the scales Attractiveness ($\alpha = .96$), Perspiciuity ($\alpha = .93$), Efficiency ($\alpha = .93$) Stimulation ($\alpha = .97$) indicated an excellent internal consistency, the analysis of the scales Novelty ($\alpha = .88$) and Dependability ($\alpha = .74$) indicated a good internal consistency. The experience with The Star Wars Galaxy Explorer Guide book produced two types of results. First, we can observe the value of each UEQ item (see Fig. 4), where the average reveals a positive evaluation of the UX for each scale. The value of the Attractiveness scale is 1,865; Perspiciuity 1,393; Efficiency 1,238; Dependability 1,012; Stimulation 1,536; and Novelty 1,667.

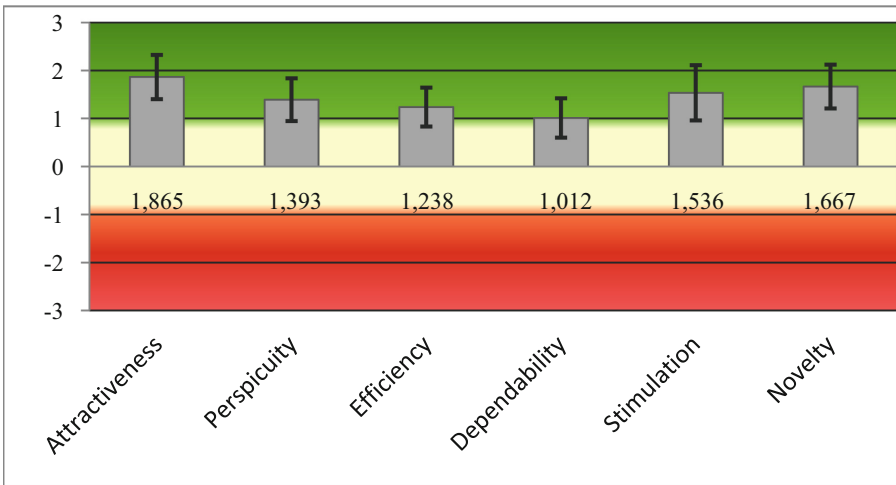


Fig. 4. Average UEQ scale values of the Star Wars Galaxy Explorer Guide, AR Book.

We obtained the other results about the UEQ benchmark (see Fig. 5). We analyze the UX of the AR book Star Wars Galaxy Explorer Guide compared to other products. The diagram shows that the scale value of Attractiveness and Novelty are in the Excellent category. This AR book corresponds to the 10% best results in these scales. The scale Stimulation is in the Good category, in the 25% best results. The scale Perspicuity and Efficiency is in the Above Average category, which means that 25% of the benchmark products are better than this AR book. Finally, the scale Dependability is Below the Average, which means that 50% of the benchmark products are better than this AR book.

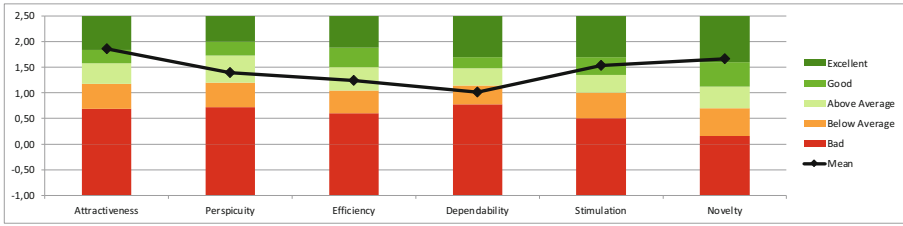


Fig. 5. UEQ benchmark diagram on the Star Wars Galaxy Explorer Guide, AR book

4.2 Frozen II the Ice Kingdom: An Enchanted Adventure, AR Book

The reliability analysis of the scales Attractiveness ($\alpha = .91$), Perspicuity ($\alpha = .90$), Stimulation ($\alpha = .94$) and Novelty ($\alpha = .91$), indicated an excellent internal consistency, the analysis of the scales Efficiency ($\alpha = .89$) and Dependability ($\alpha = .71$), indicated a

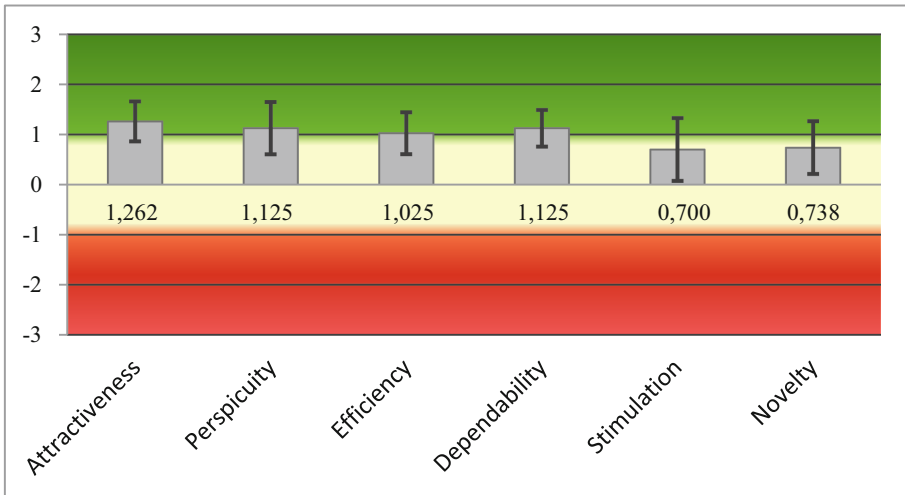


Fig. 6. Average UEQ scale values of the Frozen II the Ice Kingdom: an enchanted adventure, AR book.

good internal consistency. With the AR book Frozen II experience, we can observe the value of each UEQ item (see Fig. 6). The analysis of the average reveals a positive evaluation of the UX for the scales of Attractiveness (1,262), Perspicuity (1,125), Efficiency (1.025), and Dependability (1,125). It also reveals a neutral evaluation for the scales of Stimulation (0,700) and Novelty (0,738).

The results of the UEQ benchmark (see Fig. 7) for this AR book show that the scale value of Attractiveness and Novelty are in the Above Average category, which means that 25% of the benchmark products are better than this AR book. On the other hand, the scale Perspicuity, Efficiency, Dependability and Stimulation are in the Below category, which means that 50% of the benchmark products are better than this.

Average category, which means that 50% of the benchmark products are better than this.

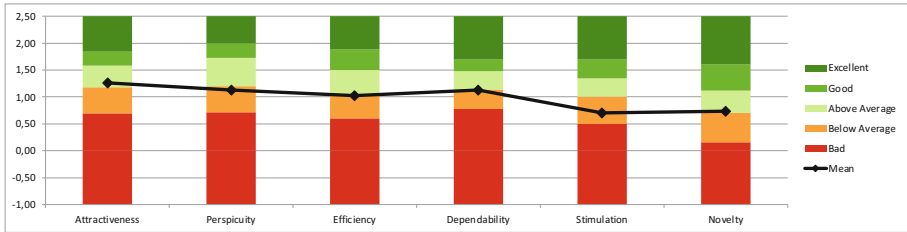


Fig. 7. UEQ benchmark diagram on the Frozen II the Ice Kingdom: an enchanted adventure, AR book.

4.3 Toy Story 4 Woody the Adventure of Augmented Reality, AR Book

The reliability analysis of the scales Attractiveness ($\alpha = .95$) and Stimulation ($\alpha = .94$) indicated an excellent internal consistency, the analysis of the scales Perspicuity ($\alpha = .86$), Efficiency ($\alpha = .79$), Dependability ($\alpha = .76$) and Novelty ($\alpha = .87$) indicated a good internal consistency. The experience with the AR book Toy Story 4 Woody, the Adventure of Augmented Reality, also produced two types of results. First, we can observe the value of each UEQ item (see Fig. 8), where the average reveals a positive evaluation of the UX for each scale. The value of the Attractiveness scale is 2,000; Perspicuity 1,475; Efficiency 1.475; Dependability 1,325; Stimulation 1,613; and Novelty 1,438.

We obtained the other results about the UEQ benchmark (see Fig. 9). The diagram shows that the scale value of Attractiveness is in the excellent category. This AR book corresponds to the range of 10% best results in this scale. The scale Stimulation and Novelty are in the Good category, meaning this book is in the 25% best results compared to the other benchmark products. Finally, the scale Perspicuity, Efficiency, and Dependability are in the Above Average category, which means that 25% of the benchmark products are better than this AR book in these scales.

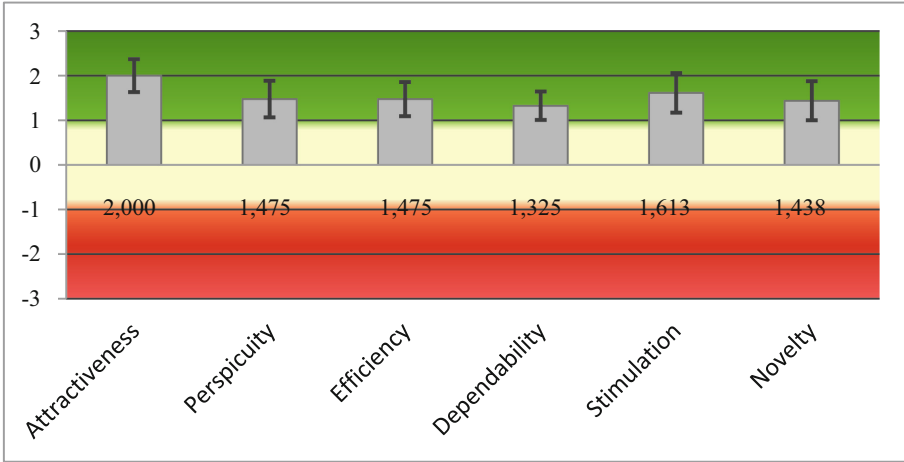


Fig. 8. Average UEQ scale values of the Toy Story 4 woody the adventure of augmented reality, AR book

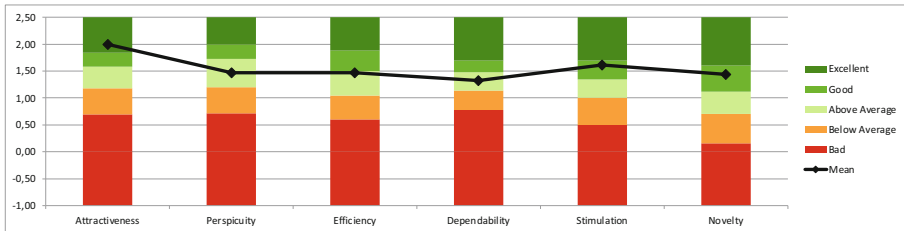


Fig. 9. UEQ benchmark diagram on the Toy Story 4 woody the adventure of augmented reality, AR book.

5 Discussion

The hypothesis: do AR published books show sufficient UX by having a positive evaluation in the scales of Attractiveness, Perspicuity, Efficiency, Dependability, Stimulation, and Novelty and fit in the Good/Excellent category in the benchmark analysis guided this fieldwork study. However, the fieldwork produced empirical evidence that does not entirely support this hypothesis. The same publisher publishes the three AR books in the same year. They have the same interface, and an app makes access to the contents. The essential differences are the themes of the books. Also, the book *Star Wars Galaxy Explorer Guide* has more content (text and 2D Animations). The book’s classification also represents a difference. *Toy Story 4, Woody the Adventure of Augmented Reality*, is a children’s fiction book, and the other two are juvenile literature.

Through the UEQ, the AR book *Star Wars Galaxy Explorer Guide*, and the AR book *Toy Story 4, Woody the Adventure of Augmented Reality* evaluated all the scales positively. The scale with the higher mean in the three books is Attractiveness, which means users liked the product. On the other hand, AR book *Frozen II the Ice Kingdom: An Enchanted Adventure* obtained a positive evaluation only in Attractiveness, Perspicuity,

Efficiency, Dependability, and a neutral evaluation in the scales of Stimulation and Novelty. These results show that the Frozen II AR book is a little exciting and motivating product, and that the product is less innovative and creative and catches less of the user's interest.

The benchmark analysis reveals curious results. In the Star Wars Galaxy Explorer Guide AR book, Attractiveness and Novelty are in the Excellent category. This result means that this book is in the range of 10% best results in the benchmark, meaning users like it and think it is more innovative and creative than 90% of the products. In the Good category is the Stimulation scale. Meaning this book is in the best 25% of the products in terms of the excitement and motivation it is to use it. Above Average are the Perspicuity and Efficiency scales, which means 25% of the products in the benchmark perform better in solving tasks without unnecessary effort and learning to work with the product efficiently. The scale Dependability is in the Below Average category. It is to say that 50% of the products in the benchmark are better in making users feel control over the interaction.

In Frozen II the Ice Kingdom: An Enchanted Adventure AR Book, 25% of the products in the benchmark perform better in the Attractiveness and Novelty scales. These two scales are in the Above Average category. However, users' response after experimenting with this book indicates that 50% of the products in the benchmark perform even better in the scales of Perspicuity, Efficiency, Dependability, and Stimulation. These scales are in the Below Average category.

In Toy Story 4 Woody the Adventure of Augmented Reality, Attractiveness is in the Excellent category. The scales Novelty and Stimulation are in the Good category. The Perspicuity, Efficiency, and Dependability scales are in the Above-average category.

Has said by Schrepp, Hinderks, & Thomaschewski [7, p43] "the general UX expectations have grown over time. Since the benchmark also contains data from established products, a new product should reach at least the Good category on all scales".

As we showed before, the three AR books do not reach the Good category in all the scales compared to the benchmark. The scales that in all the books are in the categories below the Good category are Perspicuity, Efficiency, and Dependability (the goal-oriented categories also named pragmatic quality aspects). It is to say that, despite our groups of users thought AR books are exciting, motivating to use, innovative, creative, could catch user's interest, and liked the overall product, our users also thought AR books could be improved. This improvement concerns solving tasks without unnecessary effort, making learning how to use the product more accessible, becoming familiar with it, and making users feel control over the interaction.

This study contributes towards the emergent body of quantitative studies whose aim is to evaluate the UX/usability of AR books [27, 29, 31] the UX of AR systems [22–24, 32] with both hedonic and pragmatic qualities [33] and comparing it with other products [30, 32]. This study can assess that our users liked the presented AR books, considered them an innovative product, and felt motivated to use them. In this concern, the results we obtained from our study are consistent with the results of other studies [18, 26, 28]. However, we cannot just depend on the hedonic qualities of AR books. As users feel more familiar with this type of product, these qualities tend to have less impact on users' opinions. Also, hedonic qualities are not sufficient to make a good UX of a product,

meaning that users will use it for the first time because they think it is exciting to use it but will not use it continuously if the pragmatic qualities of the product are not ensured. In this sense, it is necessary to improve pragmatic aspects to achieve a better UX. The resulting guidelines of other studies [27, 29–31] and the results we obtained in this experience prove that AR books need improvements.

Future studies should also include long-term UX evaluations of AR books. Also, it would be interesting to create from scratch the measuring instruments specifically for AR products bringing novelty and usefulness for the Human-computer interaction study area. As in any study, there were limitations to the present study. We recognize we use a UEQ for this evaluation on AR books. This study could be completed by an ergonomic inspection or a usability test. Despite our users reading the AR books as we observed, we did not use that information in this study.

6 Conclusion

This research is the first study to evaluate the UX of AR published books that includes pragmatic and hedonic qualities to the best of our knowledge. Thus, this study is a first step towards improving AR published books through a holistic UX approach. Although books obtained positive evaluation from users in almost all the scales, the comparison with the benchmark showed that the pragmatic qualities need improvement. The analyzed books failed to meet the UX goals requirements, in what respects to Efficiency, Perspicuity, and Dependability.

As we know, AR can widen the capacities of the traditional book, and it offers the opportunity to create new layers of information, including visual contents. If, in the beginning, AR was treated more like a gimmick than an essential part of the book, today's readers expect more integration and equilibrium between real and virtual contents and a better UX. So, we cannot rely anymore only upon the freshness aspect of this product. Our findings indicate that users liked the overall product, though it was innovative and creative, caught users' interest, and thought it was exciting and motivating to use it. However, the obtained results indicate that not all UX dimensions have a good enough level. The pragmatic dimensions need improvement so AR books can be used as an editorial product capable of facing the demands and expectations of the new generations.

References

1. Gomes, A., Esteves, M., Duarte, S.: Hábitos de leitura de livros. Relatório. MC'Donald's - PNL2027 – Expresso (2019)
2. Arifin, Y., Sastria, T., Barlian, E.: User Experience metric for augmented reality application: a review. *Procedia Comput. Sci.* **135**, 648–656 (2018). <https://doi.org/10.1016/j.procs.2018.08.221>
3. Irshad, S., Awang, D.: A UX oriented evaluation approach for mobile augmented reality applications. In: *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia (MoMM 2018)*, pp. 108–112. Association for Computing Machinery, New York (2018) DOI:<https://doi.org/10.1145/3282353.3282357>

4. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 63–76. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-89350-9_6
5. Schrepp, M., Hinderks, A., Thomaschewski, J.: Design and evaluation of a short version of the user experience questionnaire (UEQ-S). *IJIMAI* **4**(6), 103–108 (2017)
6. Cota, M., Thomaschewski, J., Schrepp, M., Goncalves, R.: Efficient measurement of the user experience: a portuguese version. *Procedia Comput. Sci.* **27**, 491–498 (2014). <https://doi.org/10.1016/j.procs.2014.02.053>
7. Schrepp, M., Hinderks, A., Thomaschewski, J.: Construction of a benchmark for the user experience questionnaire (UEQ). *Int. J. Interact. Multimedia Artif. Intell.* **4**(4), 40–44 (2017)
8. Schrepp, M., Hinderks, A., Thomaschewski, J.: Applying the user experience questionnaire (UEQ) in different evaluation scenarios. In: Marcus, A. (ed.) DUXU 2014. LNCS, vol. 8517, pp. 383–392. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07668-3_37
9. Rekimoto, J.: Matrix: a real time object identification and registration method for augmented reality. In: 3rd Asia Pacific Computer Human Interaction, Shonan, 17 July 1998, pp. 63–68 (1998)
10. Billinghurst, M., Kato, H., Poupyrev, I.: The magicbook—moving seamlessly between reality and virtuality. *IEEE Comput. Graph. Appl.* **21**(3), 6–8 (2001)
11. Vanderschantz, N., Hinze, A., AL-Hashami, A.: Multi-level engagement in augmented reality children’s picture books. In: Lamas, D., Loizides, F., Nacke, L., Petrie, H., Winckler, M., Zaphiris, P. (eds.) INTERACT 2019. LNCS, vol. 11749, pp. 558–562. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29390-1_37
12. Hita, Z., Lee, Y.: Interactive spatial augmented reality book on cultural heritage of Myanmar. *J. Inf. Commun. Conv. Eng.* **18**(2), 69–74 (2020). <https://doi.org/10.6109/JICCE.2020.18.2.69>
13. Shinhyo, K., Jihyun, P., Jusub, K.: E-mmersive Book: the AR book that assists the syntopical reading. In: Proceedings of the 9th Augmented Human International Conference (AH 2018) , Article 37, pp. 1–2. Association for Computing Machinery, New York (2018). <https://doi.org/10.1145/3174910.3174942>
14. Garcia-Sanchez, J.C.: Augmenting reality in books: a tool for enhancing reading skills in Mexico. *Publ. Res. Q.* **33**(1), 19–27 (2017). <https://doi.org/10.1007/s12109-017-9499-2>
15. Cheng, K.: Reading an augmented reality book: an exploration of learners’ cognitive load, motivation, and attitudes. *Austr. J. Educ. Technol.* **33**(4), 53–69 (2017). <https://doi.org/10.14742/ajet.2820>
16. Bakri, F., Oktaviani M., Mulyati, D.: Textbooks equipped with augmented reality technology for physics topic in high-school. *Jurnal Penelitian Pengembangan Pendidikan Fisika* **5**(2), 113–122 (2019). <https://doi.org/10.21009/1.05206>
17. Weng, C., Otanga, S., Christianto, S., Chu, R.: Enhancing students’ biology learning by using augmented reality as a learning supplement. *J. Educ. Comput. Res.* **58**(4), 747–770 (2020). <https://doi.org/10.1177/0735633119884213>
18. Swan, J., Gabbard, J.: Survey of user-based experimentation in augmented reality. In: Proceedings of 1st International Conference on Virtual Reality, vol. 22, pp. 1–9 (2005)
19. Koutromanos, G., Mavromatidou, E.: Augmented reality books: what student teachers believe about their use in teaching. In: Tsiatsos, T., Demetriadis, S., Mikropoulos, A., Dagdilelis, V. (eds.) Research on E-Learning and ICT in Education, pp. 75–91. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-64363-8_5
20. Billinghurst, M., Lindeman, R., Swan, J.: A systematic review of usability studies in augmented reality between 2005 and 2014. In: 2016 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), pp. 49–50 (2016). <https://doi.org/10.1109/ISMAR-Adjunct.2016.0036>

21. Law, E., Heintz, M.: Augmented reality applications for K-12 education: a systematic review from the usability and user experience perspective. *Int. J. Child-Comput. Interact.* **30**, 100321 (2021)
22. Olsson, O., Lagerstam, E., Kärkkäinen, T., Väänänen-Vainio-Mattila, K.: Expected user experience of mobile augmented reality services: a user study in the context of shopping centres. *Pers. Ubiqu. Comput.* **17**(2), 287–304 (2013). <https://doi.org/10.1007/s00779-011-0494-x>
23. Irshad, S., Rambli, D.: Multi-layered mobile augmented reality framework for positive user experience. In: *Proceedings of the 2nd International Conference in HCI and UX Indonesia 2016 (CHiUXiD 2016)*, pp. 21–26. Association for Computing Machinery, New York (2016). <https://doi.org/10.1145/2898459.2898462>
24. Irshad, S., Rambli, D., Sulaiman, S.: Design and implementation of user experience model for augmented reality systems. In: *Proceedings of the 18th International Conference on Advances in Mobile Computing & Multimedia (MoMM 2020)*, pp. 48–57. Association for Computing Machinery, New York (2020). <https://doi.org/10.1145/3428690.3429169>
25. Pamparău, C., Vatavu, R.: A research agenda is needed for designing for the user experience of augmented and mixed reality: a position paper. In: *19th International Conference on Mobile and Ubiquitous Multimedia (MUM 2020)*, pp. 323–325. Association for Computing Machinery, New York (2020) DOI:<https://doi.org/10.1145/3428361.3432088>
26. Grasset, R., Dünser, A., Billinghamurst, M.: Edutainment with a mixed reality book: a visually augmented illustrative childrens’ book. In: *Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology (ACE 2008)*, pp. 292–295. Association for Computing Machinery, New York (2008). <https://doi.org/10.1145/1501750.1501819>
27. Dünser, A., Hornecker, E.: Lessons from an AR book study. In: *Proceedings of the 1st International Conference on Tangible and embedded interaction (TEI 2007)*, pp. 179–182. Association for Computing Machinery, New York (2007). <https://doi.org/10.1145/1226969.1227006>
28. Gazcón, N., Castro, S.: ARBS: an interactive and collaborative system for augmented reality books. In: De Paolis, L.T., Mongelli, A. (eds.) *AVR 2015*. LNCS, vol. 9254, pp. 89–108. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22888-4_8
29. Cao, R., Hou, W.: Research on the interaction design of AR picture books via usability test. In: Tang, Y., Zu, Q., Rodríguez García, J.G. (eds.) *HCC 2018*. LNCS, vol. 11354, pp. 524–534. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-15127-0_53
30. Polyzou, S., Botsoglou, K., Zygouris, N., Stamoulis, G.: Interactive books for preschool children: from traditional interactive paper books to augmented reality books: listening to children’s voices through mosaic approach. *Education*, 3–13 (2022). <https://doi.org/10.1080/03004279.2021.2025131>
31. Kljun, M., et al.: Augmentation not duplication: considerations for the design of digitally-augmented comic books. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI 2019)*, Paper 103, pp. 1–12. Association for Computing Machinery, New York (2019). <https://doi.org/10.1145/3290605.3300333>
32. Stumpp, S., Knopf, T., Michelis, D.: User experience design with augmented reality (AR). In: *Proceedings of the ECIE 2019 14th European Conference on Innovation and Entrepreneurship*, pp 1032–1040 (2019)
33. Alenljung, Z., Lindblom, J.: User experience in augmented reality: a holistic evaluation of a prototype for assembly instructions. In: Soares, M.M., Rosenzweig, E., Marcus, A. (eds.) *HCII 2021*. LNCS, vol. 12781, pp. 139–157. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78227-6_11