Mobile Augmented Reality Apps in Education: Exploring the User Experience Through Large-Scale Public Reviews

Jessica Lizeth Domínguez Alfaro(✉) and Peter Van Puyvelde

Department of Chemical Engineering, KU Leuven, 3001 Heverlee, Belgium
jessicalizeth.dominguezalfaro@kuleuven.be

Abstract. Augmented reality (AR) is considered one of the top technologies that will revolutionize the future of education. Real-time interaction, different formats of visualization, and the merge of the real and digital world may open up new opportunities for teaching and learning. Although AR is easily accessible via mobile phones, the extent to which this technology will be adopted greatly depends on the user experience. The user reviews of mobile applications or so-called "apps" are a potential source of information for designers, software developers, and scholars interested in understanding the user experience. This study investigates the current state of the user experience of augmented reality apps by extracting and classifying the information from reviews published in the Google Play Store. A set of 116 educational mobile AR apps were mined from the Google Play Store, and a total of 1,752 user reviews were retrieved and classified. Results suggest developers of educational mobile AR apps need to solve technical problems, improve certain features, and provide more explicit instructions to users. Regardless of these needs, users recognize that these apps have great potential as educational tools. Future developments should focus on tackling these shortcomings, expanding the use of AR apps to more fields of education, and targeting specific audiences to extend the technology adoption.

Keywords: Augmented Reality · App reviews · Data mining · Mobile learning · User experience

1 Introduction

With the advances in mobile computing, Augmented Reality (AR) capabilities have been combined with wireless devices such as smartphones and tablets [31], this implementation is called Mobile Augmented Reality (MAR) [27]. As a result, the use of AR has spread over many sectors such as entertainment, industry, games, tourism, and education [34]. In education, AR is considered

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-87595-4_32) contains supplementary material, which is available to authorized users.

L. T. De Paolis et al. (Eds.): AVR 2021, LNCS 12980, pp. 428–450, 2021.
https://doi.org/10.1007/978-3-030-87595-4_32

one of the technologies that are revolutionizing teaching and learning. Unlike other technologies, it often only requires a handheld device [13]. Several studies have documented via systematic reviews the different uses, advantages, and disadvantages of AR in educational settings [6,11,21,26,46]. However, the state of this technology for educational purposes in the public market has not been analyzed.

Current examples of user experience research of AR public apps have shown that 1) users like AR applications, 2) technical features strongly influence user satisfaction, and 3) the risk-benefit ratio of using a certain AR app influence the rate of adoption. Nonetheless, none of these researchers focused on public educational apps [19,27,32,38].

This study analyzes educational MAR applications from the Google Play Store and synthesizes user reviews into insights for improvement. Due to the number of applications and user reviews, the analysis is a challenging task. This study takes advantage of an automatic tool derived from data mining techniques to aid in qualitative and quantitative analysis.

In addition, to user reviews, this study analyses some of the indicators that users consider before downloading an app, such as ratings, number of installs, release dates, app sizes, and AR-type. Since the user experience starts from the moment the user selects and downloads an application [9], the study of such metrics can provide valuable insights.

Additionally, the levels and fields of education are analyzed using the International Standard Classification of Education (ISCED) [48] used by other researchers [11,20]. This serves to identify the educational sector that an AR application targets.

Overall, this study aims to answer the following research questions:

- RQ1: What are the main problems, feature demands, and opinions expressed by the users of educational MAR apps published in the Google Play Store?
- RQ2: What are the main characteristics in terms of price, size, rating score, installs, release dates, and AR types of educational MAR apps?
- RQ3: Which are the levels and fields of education that educational MAR apps published in the Google Play Store target?

2 Background and Related Work

2.1 MAR and Education

Mobile Technology has opened up opportunities for learners to access ubiquitously to learning content. Today, this access has been enriched by blending other technologies such as Augmented Reality (AR). Initially developed for the aerospace industry [14], AR is described as a system that "supplements the real world with virtual (computer-generated) objects that appear to coexist in the same space as the real world", and its main features are its capability to combine the real and virtual environments, the real-time interaction, and the alignment of digital content with the real world [10].

In general, AR systems can be divided into three types: marker-based, marker-less, and location-based. Marker-based AR uses artificial markers (e.g., QR codes, fiducial markers, or images) placed in the real environment to detect the position and display the virtual objects. Marker-less AR tracks the natural features of physical objects in the environment. Whereas location-based AR employs the position data of the devices, determined by Global Positioning System (GPS) or Wi-Fi to deploy the virtual content [16,43,51].

MAR is getting more attention in education as recent evidence suggests that AR has several affordances that allow supporting meaningful learning. For example, AR can enrich learning experiences with 3D objects, which can help visualize abstract or unobservable phenomena. It can annotate spaces with overlaid information or combine digital and physical objects, creating hybrid learning environments [6,11,21,52]. These affordances have been linked to several learning effects which has proven some of the main advantages of implementing this technology as a new educational tool. Some of these effects are related to the improvement of spatial abilities and memory retention, the decrease in cognitive overload, and the increase in motivation [6,11,21,52] .

Despite the evident evolution of AR, there exist some limitations regarding its use in educational settings. These limitations are related to both pedagogical and technical aspects. Pedagogically, previous research has established that AR can be unwieldy, distracting, and cognitively demanding for students [6,11,21,52]. Concerning technological aspects, most reported limitations are related to the usability and design of AR apps. For instance, AR systems can have difficulties maintaining the virtual content, detecting the user's location, and can be technically challenging to implement by teachers as the virtual content is not easy to create.

Given these challenges, some authors have argued that for AR to be successfully integrated into educational settings, well-designed user experiences are needed [16,20]. This paper attempts to explore the user experience in more detail by examining the technological challenges that users of public educational MAR apps are experiencing directly from the user reviews, which may serve to inform instructional designers on how to improve the user experience of their systems.

2.2 User Experience of MAR Applications

According to the ISO 9241-210, user experience (UX) is defined as the "person's perceptions and responses resulting from the use and anticipated use of a product, system or service" [28].

Previous studies on UX of MAR applications have found that users value the practical application or usefulness, the relevant content, the ease of use, and the absence of technical difficulties [19,38]. Furthermore, the emotional perception of users can influence the experience as observed by Dirin and Laine [19] where users have positive emotions such as encouragement, excitement, and interest. Similarly, Li et al. [32] found that the satisfaction of users can be a motivation to continue or stop using a MAR application.

Although not prescriptive, this work aims to add to the efforts of other researchers by analyzing the user experience of mobile AR users within the context of educational apps.

2.3 User Reviews and App Data as a Source of Information

Application distribution platforms or apps stores contain rich information that can help software engineers and designers during the design process [24,39]. They contain public reviews reflecting users' experience and app metadata such as price, number of installs, release dates, rating, app size, etc., which provides an overview of the app characteristics.

User review analysis can be used to quantitatively summarize the topics contained in the reviews and to gain insights from a qualitative examination of the content [49].

Nicholas et al. [35] used the user reviews of 48 apps for bipolar disorder to analyze the user perspectives of these apps qualitatively. Similarly, Alqahtani and Orji [7] analyzed reviews of mental health apps to reveal what users of such apps like and dislike.

However, using app reviews as sources of information comes with some challenges. App reviews can be short, unstructured, grammatically incorrect, and depending on the app; they can be abundant, which can make it a complex and time-consuming process to extract information from them [24,41].

Regardless of these challenges, when compared to other resources such as focus groups or surveys, the reviews provide direct, and numerous feedback for the stakeholders involved in the app development [25]. Moreover, the topics included in such reviews contain relevant information from the user experience such as the ideas, the needs, and the experience with a certain app that may potentially inform scholars who are interested in understanding the context of a particular app [39].

In this study, user reviews are used as a source of information (e.g., identify problems of existing apps, requested features from the users, etc.), and unlike other scholars, we used an automatic and public tool developed by Panichella et al. [41] to classify the content of the reviews into relevant categories that allow the detailed analysis of the requests and information in a procedural way with low processing time.

3 Methodology

The general methodology of data processing consists of three main phases: app extraction, review extraction, and review classification (Fig. 1). The Google Play Store was chosen as the data source. This decision was made due to a relatively larger amount o f apps in Google' s app store when compared to the Apple Store [17].

3.1 App Extraction

The first step was to extract a list of apps from Google Play using an open-source scraping tool, called google-play-scraper (version 7.1.2) [37]. The "search" function provided by this library allows the extraction of the app identifiers corresponding to the results from a search in the Google Play store using any term. The search terms used are: "augmented reality education", "augmented reality education apps","augmented reality educational apps", "ar education", "ar education app" and"ar application for education". These were chosen by using a keyword tool [3].

For each search term, a maximum of 250 apps was retrieved, which gave 1500 apps from all search terms. Out of these 1500 apps, only 440 were unique, i.e. not repeated in the list. The metadata from each app were retrieved using the "app" function from the scraping tool mentioned earlier from the unique list.

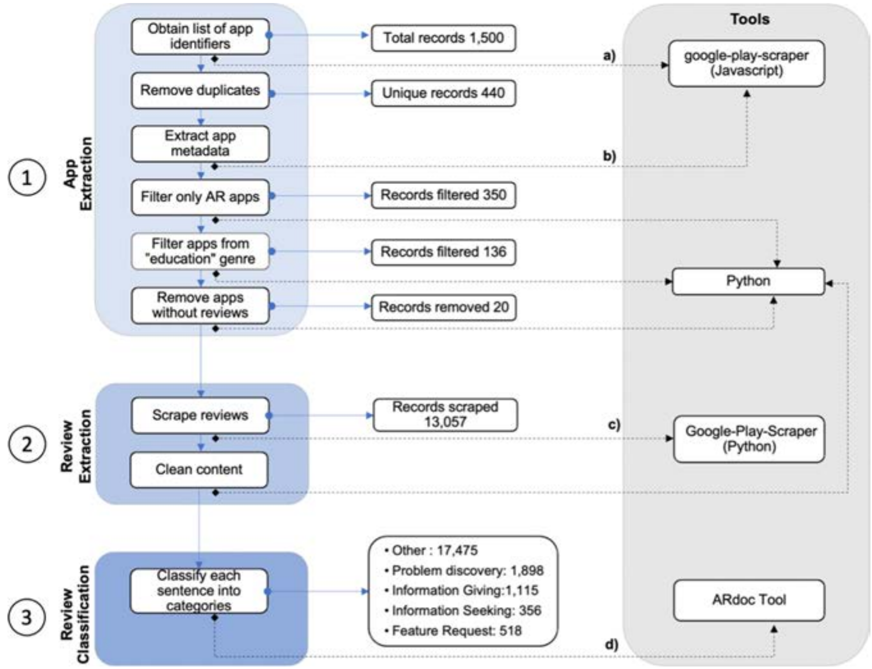

Fig. 1. The General methodology for data processing. The main phases are 1) app extraction, 2) review extraction, and 3) review classification. This figure summarizes the specific steps followed in each phase, the results, and the tools used. a) "search" function [37]; b) "app" function [37]; c) "reviews" function [30]; d) automatic data mining tool [41]

Finally, an additional cleaning step to remove non-AR apps was performed by assuming that AR apps must contain in their title or description the keywords "ar" or "augmented reality". Applications that did not contain any of these non-case-sensitive keywords in the title or the description were removed from the database. As a result, the data-set was reduced to 350 apps.

3.2 Review Extraction

From the previously cleaned app's database, a subset was created from apps that fall under the app store "Education" class; only 136 apps satisfied this criterion. The user reviews were mined using a different open source scraping tool called Google-Play-Scraper (version 0.0.2.2) [30] coded in Python. The reviews from each app were extracted using the function "reviews". In total, 13,057 single reviews were obtained from 116 apps. It is important to mention that even though 136 apps were mined, only 116 apps contained reviews.

3.3 Review Classification

To automatically classify the content of the reviews, the App Reviews Development Oriented Classifier (ARdoc) developed by Panichella et al. [41] was used. This tool combines three feature extraction techniques: Natural Language Processing (NLP), Text Analysis (TA), and Sentiment Analysis (SA). Using a machine learning (ML) algorithm, ARdoc automatically classifies the sentences of the reviews into a predefined taxonomy. The taxonomy includes five categories: information giving, information seeking, feature request, problem discovery, and other. The taxonomy emerged from previous research, which combined common topics from user reviews and comments given by developers [40,41]. In this study, the Java API version of the ARdoc tool was used. The feature combination of text structures and sentiment analysis ("NLP+SA") was employed since, according to Panichella et al. [41], this combination of features allows a higher classification precision, recall, and F-measure (all ranging between 84–89%).

In total 21,362 sentences from 13,057 reviews of 116 apps were classified. It is important to note that most of the reviews contained more than one sentence; therefore, the number of reviews is smaller than the number of sentences.

The extraction of informative reviews was performed to analyze the content of each sentence and derive conclusions. For this purpose, it was assumed that informative reviews are those which contain sentences classified into "problem discovery", "information giving", "information seeking", and "feature request". As a result, the filtering step consisted of the removal of reviews with sentences that had all been classified as "other" by the ARdoc tool; the total informative reviews were 3,065. It is important to note that not all the apps contained informative reviews; therefore, the reviews used for further analysis only come from 86 apps. The complete data-set and classification can be found in the Supplementary Information.

A manual inspection of the informative reviews was performed to derive more specific and detailed conclusions. The application "Sky View Lite" was excluded due to the high number of reviews (1,313). In total, 1,752 reviews were further analyzed (Table 1). It is important to mention that all reviews were divided manually into mutually exclusive groups corresponding to the app categories from ARdoc. Most of the reviews contained more than one sentence that could belong to different categories. If a review had sentences classified into different classes, the priority was given based on the review context. In addition, another class, denominated "Irrelevant", was added to reviews with vague or unclear statements.

Finally, each of the informative reviews was manually analyzed and assigned to a broad topic. These topics were derived from those defined by Pagano and Maalej [39] and subsequently used by Panichella et al. [40] to define the taxonomy of ARdoc [41]. If the content of the review was not related to any of these topics, a new broad topic was created.

3.4 App Characteristics Extraction

All the metadata of each application obtained in Sect. 3.1 was formatted in a table for further analysis (see Supplementary Material).

3.5 Classification Under Fields and Levels of Education

To classify the apps according to the levels of education, the International Standard Classification of Education (ISCED) [48] was used in its aggregated level. The app developers' description was manually examined to identify the level and field of education targeted by the app. Each app can be classified into several levels. If the target level was not explicitly expressed, the app was assigned considering other features such as images, videos, and educational content reported in the app description. If this information did not provide clear indications, the apps were labeled as "Not Classified". On the other hand, if the fields of education were not explicitly expressed, the apps were classified as "N/A" (see Supplementary Material).

4 Results and Discussion

This section summarizes the findings related to the user reviews and their classification using the ARdoc tool, the characteristics of the AR apps obtained from the metadata of each application and the survey of the educational levels and fields.

4.1 What Are the Main Problems, Feature Demands, and Opinions Expressed by the Users of Educational MAR Apps Published in the Google Play Store?

As described in Sect. 3.3, the reviews extracted from 116 educational apps were classified by the ARdoc tool. As a result, 1,752 reviews were analyzed.

From Table 1, it can be seen that the largest number of reviews correspond to problem discovery (1,142/1,752). These reviews come from 70 apps, meaning that

Table 1. Classification of informative reviews as described in Sect. 3.3. This classification contains the number of reviews per category, the percentage in relation to the total of reviews, the average app score rating per category, and the number of apps on each category

Category	Number of reviews	Percentage of total reviews ($N = 1,752$)	Average score per category	Number of apps
Problem discovery	1,142	65%	1.6	70
Information giving	317	18%	3.8	56
Feature request	135	8%	3.9	29
Information seeking	125	7%	3.0	38
Irrelevant	33	2%	2.9	17

from the total of apps with informative reviews, 81% (70/86) of apps have reviews that report problems. Additionally, according to these reviews, the average score rating of the app is 1.6 out of 5 stars, which is the lowest between the categories.

Problem Discovery. Among the category of "problem discovery", six broad topics were identified. Specifically, it was identified that users have problems with functionality of the app, specific features, general service, hardware, software, and content.

In this category, most problems with AR apps were found to be related to the functionality of the apps. Although some reviews were not descriptive, for example, some users provide reviews that indicate a problem such as "app is not working" or "app did not work", the high number of reviews related to this topic suggests that some users are not able to use the AR apps, install them or even use basic functionality. Commenting on AR functionality, one of the users said,"only concern is that ar does not work it is just freezes or shows a black screen" (from the Physics Lab app).

Hardware-related problems were found on 14 apps. The most common problems related to hardware are camera autofocus, battery consumption, and phone overheat. Users from different apps expressed these observations. Commenting on phone overheat, one user said: "it's pretty good and awesome for children even for adult it's fun! but it makes my phone overheat and shutdown always. . ." (from the Animal 4D+ app). Regarding battery consumption, one user commented: "expeditions is dramatically draining my phones battery. I am wondering why my phone is dead. looked at history and found that expeditions was responsible for using up 40% of my battery even though I have not used it in several days..." (from the Expeditions app).

Software-related problems were classified as those reviews in which users explicitly express the appearance of an error message, issues with the mobile data, or software compatibility. For example, one user stated: "tried in 3 devices. failed to initialize Vuforia" (from Human Anatomy AR app).

Another recurrent problem expressed by users was content-related. Either the AR content was not available, it was not working, or the access to it was absent. A user said: “this is not a good app the animations are really too short, and, in some topics, the audio is not available...” (from Cambridge Explore app).

Thus, from the analysis of the “problem discovery” category, users of educational MAR apps experience issues of diverse nature which can be related to several factors. However, the technical hurdles related to the AR technology such as power consumption, AR functionality, and real-time interaction or unavailability of content must be a high priority to developers as these problems will cause users to stop using the applications [32,38]. For example, power consumption can be solved through energy-saving techniques, reduce power consumption from GPU, and the use of active cooling methods to control the temperature of handheld devices [15]. To improve AR functionality, developers can leverage improved tracking methods. For instance, hybrid tracking technology tackles the computational inefficiency of handheld devices and network performance [44]. Finally, Web services, together with the implementation of emerging 5G networks, can help to reduce the issues related to the content and real-time interaction by improving the data transmission rate. As mobile device computational efficiency tends to be limited, especially in low-end devices, apps that rely on web services may be able to provide a better experience.

Information Giving. The reviews in this category account for 18% (317/1,752, see Table 1) of the total informative reviews coming from 45 apps, and the corresponding average score rating according to these reviews is 3.8 out of 5 stars.

Five broad topics were identified from the content analysis of this group of reviews. Some of these topics were also identified by [39,40]. The most common issues are praise, dispraise, helpfulness, shortcoming, and other feedback.

The following reviews illustrate some examples of praises found in the dataset. One user of the AR mondly app stated: “unique ar experience! mondly ar proves augmented reality can be the future of language learning!”. Another user from the Expeditions app commented: “I love using this app with my students. I used a google expeditions kit from @aquilaeducation and my students absolutely loved traveling all over the world!”.

On the other hand, users also expressed their discontent. It is important to note that most of the criticisms made by users mostly come together with a “problem discovery”. Therefore it was not usual to find dispraise sentences alone but combined with a specific issue or a dissuasion statement. For example, one user expressed: “I didn’t even liked to give one star it is not a good app don’t download it I hate it” (from Space 4D+ app). Another example of reviews is “the worst app ever. Waste of time. Do not download this app” (from Physics Lab app).

Another common topic among the “information giving” category was the “helpfulness” topic. In general, these reviews where a particular application was practical or helpful. Among the total of apps, 29 apps have reviews related to this topic. Commenting on the learning experience, one of the users said: “a great app to learn and practice chemistry for grade 10 students! you should try using it” (from Dat Thin Pone app). And another user stated: “thanks a lot for the app its awesome and very helpful for the students who cannot afford physical lab" (from Physics Lab app).

Some users provide information about concrete aspects of apps they are not satisfied with. These types of reviews were classified under the "shortcoming" topic. It is important to mention that even though shortcomings seemed to be similar to problems classified under the "problem discovery category", they differ on the intention expressed by the user. While a problem is a fact that the user experience, the shortcoming derives from subjective opinion. For example, users unsatisfied with markers expressed, "maybe works but need to print markers. no thanks! only free trial" (from the Predators AR app). This review not necessarily implies that the marker does not work but only that the user is not happy with printing it. In the same fashion, another user said, "please provide instructions. ok, I downloaded the file and yes it works and works well. but having to have the target image is not practical" (from the Augmented Reality Solar System app).

The most common shortcomings include payment options, privacy concerns, and instructions provided by developers. For example, a user of the Quiver 3D coloring app shared:"I really like your app, but I wish you did not have to buy everything separately and you could buy a membership or whatever. I understand that you have to make money, but I would also like if there were more free pages. thanks!" . Regarding privacy concerns, a user said: "it requests to spy on you", and another expressed: "the app wants to access my contacts and to make and manage phone calls. This is a coloring app. I do not see why it needs the ability to make calls. . ." (from the Quiver 3D coloring app). Finally, related to the app instructions, a user from the Physics lab app expressed "it is alright if you understand physics but for people for do not understand it needs more instructions for people who don't understand. The instructions that you currently have a very hard to understand..." .

Overall, the results indicate that most users tend to write reviews when they are extremely satisfied or highly dissatisfied (praise or dispraise), but they do not provide specific insights unless the user gives specific details. In addition, users share the helpfulness of certain apps and some shortcomings that they experienced, such as lack of instructions, dislike towards markers, payment issues, and privacy concerns.

Taking into account these observations, developers can improve the user experience by providing clear explanations about why the AR systems require certain permissions. This information can be provided before and during the use of each application, and although not all users may understand the justification of the system requirements, providing the data will allow them to choose with an informed view [23].

Users also provide information about the real-life scenarios in which MAR educational apps were found helpful. According to these findings, MAR apps can be useful when students cannot be in a particular place e.g., classroom, laboratory, or historical place. Therefore, developers and practitioners can concentrate efforts on developing experiences for users that cannot be physically present in the same learning environment of their interest. Nowadays, this will not be hard to realize after the recent pandemic where these scenarios are more "normal" [22].

Feature Request. The "feature request" category accounts only for 8% (135/1,752) of the reviews analyzed coming from 29 apps (Table 1). Interestingly, the users providing these reviews gave a higher rating score (3.9 stars) when compared to users sharing other types of reviews. In general, three types of requests were identified among this group of reviews: 1) content, 2) feature, and 3) general improvements.

The content request was one of the main demands of users. A user stated: "could you please add more information in the app when you peel off the layers? students will learn even more about it. nicely done!" (from the Human eye-Augmented Reality app). Commenting on the addition of features, one user said "fun app. for kids and adults alike their comments are: would be good if we could draw our own pictures! maybe for a future update?..." (from the Quiver 3D coloring app). Finally, concerning the request for improvements, one user stated: "the first app to use ar core and does it in a brilliant way. Would be great if you could expand the app to include placement options (placing on a surface) and have the atoms react to light". (from the Atom Visualizer for ARCore app).

The current request of users of AR apps may be partly explained by the fact that the developers are exploring the technology; thus, they might be showing only some examples showcasing how the technology/app can be used but not necessarily offering a "complete" experience. Therefore, future developments may dive into further educational content integration while maintaining good design, which can be achieved through a multidisciplinary collaboration of scholars. For instance, app developers can leverage multimedia instructional design techniques to enhance learning, and practitioners can validate content and learning outcomes of students through evaluation studies.

Information Seeking. The "information seeking" category is defined by sentences expressed by users attempting to obtain information [41]. One hundred twenty-five reviews from 38 apps were classified under this category which accounts for 7% of the sum of informative reviews. The average rating score provided by users in this category was 3.0 out of 5 stars (Table 1).

Among the reviews in this group, two main topics were identified. Users asking about "how to use" and "how or where to get content/marker" were the most popular questions. Regarding the question of "how to use?" a user expressed: "do not understand how to use it the screen say scan the flat surface. what flat surface? do I need one of those printed qr cards or what someone help me" (from the Apollo's Moon Shoot AR app). Content-related, one user asked, "from where I can get the image?" (from the AR Dino Roar app).

Other questions were related to AR elements. For example, a user asked, "what is an AR marker" (from the iBugs AR app). Another one said "i do not have an ar maker why is it that you need a AR maker. Improvement I want for my child that you do not need a AR maker." (from the Predators AR app).

Although less common, privacy concerns were also found under this category. This topic was identified under "information giving", however, some users express this concern as questions rather than as statements. For example, one user said: "why do I need to give permissions to make calls?" (from the AR animals app) and another stated: "a great way to make your pictures come alive and an underused technology. my kids (and me) love it. but why is there a permission to know my precise location?" (from the Quiver 3D coloring app).

In summary, these results show the common queries that users of MAR apps have. Interestingly, the most popular questions correspond to the function and the use of AR. It is not clear for some users how to use certain apps, what a marker is, or where to get the content. A possible explanation for this might be that there is a lack of information provided by developers, which confuses users, or they are still unfamiliar with the concept of AR. Thus, clear explanations, more intuitive interactions, or familiar cues can be integrated into the systems to familiarize users with the technology.

4.2 What Are the Main Characteristics in Terms of Price, Size, Rating Score, Installs, Release Dates, and AR-types of Educational MAR Apps?

To answer this question, we used the metadata extracted in Sect. 3.1. The data includes price, app size, ratings, number of installs, and release dates.

Price. In terms of price, most apps were reported as "free" except for the app called "Quiver Education". This result suggests that most users can access AR apps for free. However, from the data extracted, it was not possible to determine whether extra paid content was available to the user after download.

For educational purposes, this is a positive result as students would have access to the AR content. Even if the "in-app purchases" are present, the students may have a higher chance to access the content if the practitioners or institutions are able to assess the content for free and provide students with the resources.

App Size. Regarding the app size, which is the amount of memory in megabytes (MB) that each app occupies in the phone, only 105 apps declared this data. The other 11 apps contained the legend "varies with device". This information, as noted by D'Heureuse et al. [18], does not include the "in-app downloads", which refers to additional data (e.g., music, videos, 3D models).

The estimated average app size is 51.6 MB with a variance of 25.6 MB. This result is larger than the app size of "all apps" inside the Google Play Store, which was reported to be 15 MB in 2017 [12]. A possible explanation for this result may be due to multimedia content such as video, images, and 3D models. These common resources, in regular apps, account for most of their size, which may suggest that for AR apps, this also applies. Generally, when apps are optimized by developers, the size is reduced. However, in the case of MAR educational apps, the optimization process may not be a priority due to the novelty of the technology.

The app size is a factor that developers consider since it can be related to how often an app is installed [5] or as Mahmood [33] suggested, it may impact how the app is rated by users. For users, this information is vital to assess how much memory does their device needs to run AR apps. Furthermore, app size can also be relevant for scholars and practitioners looking for the successful use of AR apps among their audience. This issue may be overcome by the use of cloud services and Web-based AR applications. These approaches can be suitable since users will not have to download any application to their phones, and the digital content can be display in the cloud. However, these applications are still scarce, and experience other challenges related to mobile networks [45].

Rating Score. With respect to the ratings, Google Play Store uses a five-star scale ranging from low (1 star) to high (5 stars). The average rating of all apps ($N = 116$) is 3.2 stars with a standard deviation of 1.5 stars. According to the stats reported in AppBrain [8], the average rating for all apps in the Google Play store is 4.1 stars. Therefore, the educational MAR apps from the data-set are on average rated lower than other apps. It is difficult to explain this result since the rating can be associated with multiple factors. According to Tian et al. [50] the rating of an app is mainly influenced by the size of the app, the number of promotional images on the app store, and the target software version. In general, this finding reflects that the user experience of educational MAR apps is not optimal when compared to conventional apps.

Number of Installs In the case of the number of installs in the Google Play Store, they are expressed in terms of ranges rather than exact numbers. The ranges are expressed as follows: (0–5), (5–10), (10–50), (50–100), (100–500), (500–1000), etc. To explore if the rating score is correlated with how much an app is downloaded, the install range was compared with the weighted average score rating. Figure 2 shows the relationship between the weighted average score versus the number of installs or download ranges.

As shown in Fig. 2, apps with a small number of installs (e.g., 50–100) have a high rating score, then the score of the apps starts to drop with respect to the number of installs until the range (10,000–50,000), from that install range, the app score appears to increase in value again. As noted by other authors [18], this behavior can be possibly explained by the following hypothesis. Apps with few downloads have a rating score which is likely influenced by a small number of users (probably from people known by the developer) whereas, apps in larger install ranges receive feedback from more users. For apps that go beyond 10,000 downloads, the rating starts increasing respectively with more downloads as a result of user satisfaction and popularity.

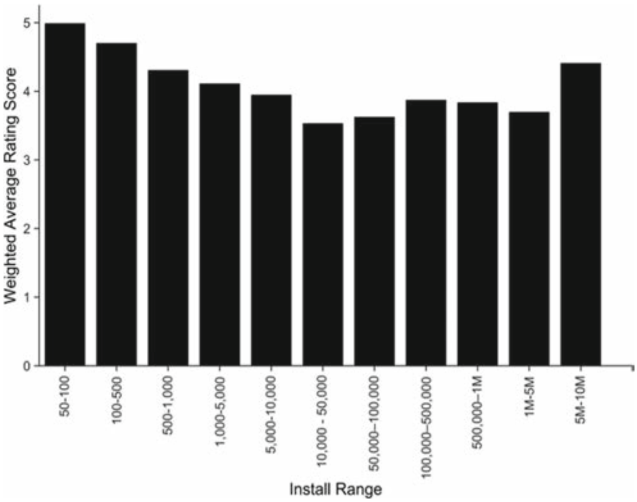

Fig. 2. Weighted average score per install range

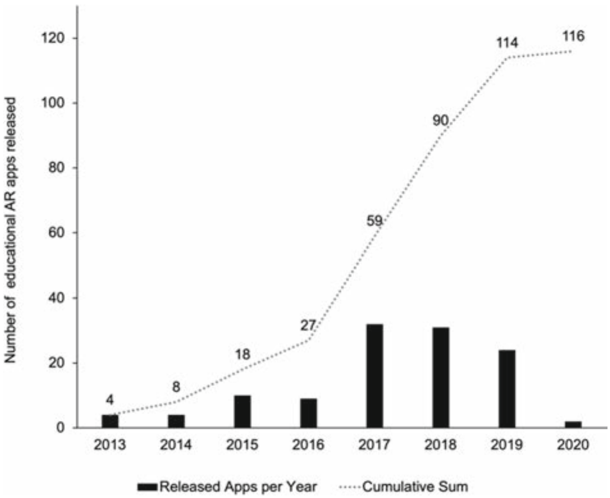

Fig. 3. Total educational MAR apps released per year since 2013 and its cumulative sum

For educational MAR apps, the number of apps in the large install range (1M–5M and 5M–10M) is very low. Just four apps in the 1M–5M range were found (Animal 4D, Expeditions, Devar, and Quiver), and only one app (Skyview) in the 5M–10M range. A possible explanation for this might be that the novelty of AR among users and developers is still present. Likewise, the use of the technology as an educational tool has not reached its maturity, thus leading to limited use, disinterest in installing the apps, or dissatisfaction of the users.

Additionally, considering that the number of installs can be related to the success of the apps [18], these findings suggest that overall, educational MAR apps are still not very successful.

Release Dates. The release dates were also analyzed to explore the evolution of MAR in the public market. According to Fig. 3, the apps in the data-set have been released since 2013, having a higher number of releases between 2017 and 2019. Since 2017, Google LLC and Apple Inc. have released their application programming interfaces or software development kits (SDKs): ARCore [2] and ARKit [1] respectively. These platforms allow developers and users of iOS and Android to create AR apps for mobile devices [36]. Other SDKs such as Vuforia have eased the implementation of AR capabilities regardless of the main operating system of the device [4]. Overall, the release of these SDKs may be a possible explanation for the increase in the number of app releases.

AR-Types. Finally, to explore the type of AR technology used by the mobile applications, the description given by the developers was examined manually. In case that the technology was not explicitly expressed, the application was downloaded and subsequently assigned to one of the AR types.

As it can be observed from Table 2, marker-based apps account for the majority of the apps (66%). Marker-less apps account for 23% of the apps and hybrid types only for 11%. Interestingly, nine apps were found to mix marker-less technology and virtual reality. Finally, no applications that used location-based AR were found in the data-set.

As mentioned, the majority of educational apps were classified as marker-based apps. This result may be explained by the fact that marker-based technology is easier to implement, the tracking is more robust, and it is less computational expensive when compared to marker-less AR [11,47]. Altogether, these results suggest there are opportunities for scholars and developers to create more marker-less apps and apps using combinations of AR-types. For educational purposes, each type of AR may provide different outcomes; therefore, it would be interesting to have more explicit guidelines on the different learning affordances that each tracking technology has to offer per field and level of education so that user and practitioners can choose the application according to their needs.

Table 2. The number of apps per type of AR technology

AR-type	Number of apps	Percentage of total ($N = 116$)
Marker	76	66%
Marker-less	27	23%
Marker-less and VR	9	8%
Marker and marker-less	4	3%
Irrelevant	33	2%

4.3 Which Are the Levels and Fields of Education that Educational MAR Apps Published in the Google Play Store Target?

To understand the educational target (levels and fields of education) of the MAR apps in the Google Play Store, each application was assigned to a field and level of education of the corresponding ISCED classification according to the information identified in the developer's description.

Table 3. The number of apps on each aggregated level of education according to the ISCED [48]

Aggregated level of education	Number of apps	Percentage of total ($N = 116$)
Levels 0–2	78	67.2%
Levels 3–4	4	3.4%
Levels 3–4 & or 5–6	22	19%
Not classified	12	10.3%
Irrelevant	33	2%

Table 3 shows that most apps target the aggregated levels 0–2 (67.2%). This observation is in line with previous studies of educational apps, which have found that most developers target preschool and primary school-age children [42]. Interestingly, this finding is contrary to that of Bacca et al. [11] who found that most AR studies target higher levels of education. One possible explanation for this is that the app market locates the "kids app market" into several industries, and developers use the term "educational" or tag words like "kindergarten", "middle school" , etc. to sell apps for the children audience, but they do not necessarily develop the apps with an educational purpose, or this specific audience [29].

It is important to note that for some apps (10%), the target level was not explicitly expressed. This outcome may be relevant for future developers who aim to get students interested in their apps. Adding this information may also allow practitioners to get more confidence in using a certain app in a formal setting.

Table 4 shows that most of the apps were classified in the broad fields of "Generic programs and qualifications" and "Natural Sciences, Mathematics and Statistics". Overall, these results reflect those of Bacca et al. [11], who also found that science was the most popular field for AR studies suggesting that this is due to the advantages that AR provides when teaching abstract concepts. However, further studies and app development in different fields of education are recommended to expand and leverage the advantages that AR may bring in various educational areas. For example, the public market of educational MAR apps can exploit more apps in the context of science and engineering but for higher levels of education with validated educational content. Also, new apps can target adult learners who may not be considered under the International

Table 4. Classification of the apps according to the aggregated level of ISCED of broad fields of education [48] in descending order and the corresponding general topics

ISCED broad fields	General topics	Number of apps	Percentage of total ($N = 116$)
00 Generic programs and qualifications		79	68%
	Natural science, animals	16	14%
	Multiple areas	15	13%
	Natural science, space	14	12%
	Literacy	12	10%
	Art	7	6%
	Natural sciences	6	5%
	History	3	3%
	Natural science, biology	2	2%
	Foreign language	1	1%
	Geography	1	1%
	Mathematics	1	1%
	Social skills	1	1%
05 Natural sciences, maths & statistics		13	11%
	Chemistry	7	6%
	Physics	4	3%
	Mathematics	2	2%
N/A		12	10%
	Visualization platform	6	5%
	Museum	4	3%
	Game	1	1%
	Marketing	1	1%
09 Health and welfare		9	8%
	Anatomy	9	8%
07 Engineering, Manufacturing & Construction		3	3%
	Engineering	3	3%
01 Education 02 Arts and humanities 03 Social sciences, journalism & information 04 Business, administration & law 06 Information & Communication technologies (ICTs) 08 Agriculture, forestry, fisheries & veterinary		0	0%

Standard Classification of Education (ISCED). Developers even can venture into new fields such as social sciences and business or administration, which may have new interesting affordances. Moreover, public institutions can collaborate with developers in making free, content-validated, and relevant for particular curricula educational MAR apps for a wide range of students.

5 Limitations of the Study

The design of the current research is subject to limitations. First, the search of the apps through Google Play Store is subjected to a limit of 250 apps per search term. Although the method ensures that the considered apps are the ones the user can find by using the same words, the data-set can be extended if other forms of search are used, for instance doing searches from different locations and different languages or adding results from websites. Second, the use of only six search terms can be a source of bias when creating the data-set; a more exhaustive list of terms might improve the broadness of the apps found. Even though the criteria for choosing the terms relies on the popular terms searched by users, there might be words that can also lead to AR apps and were excluded in this study. Third, the selection of mobile apps only from the Google Play Store limits the spectrum of apps. This opens the possibility for future research regarding an extension of MAR educational apps from the iOS store and the possible comparison with Google Play Store apps. Another limitation comes from using ARdoc to classify the reviews. The tool has its own "accuracy" and intrinsic biases derived from its assumptions and characteristics. In addition, even though the taxonomy used by [41] is in line with the objective of this study, the exclusion of reviews with the "Other" tag may have excluded sentences with relevant data. Finally, the extracted reviews may not be representative of the total amount of users of the MAR apps. In general, only users more inclined to share their thoughts are represented in the data-set. However, the group can still be considered diverse as the apps are used by different target audiences, and the reviews show diverse perspectives.

6 Conclusions and Future Work

The present research aimed to extract general information about educational MAR apps published in the Google Play Store and to examine the user experience through the classification and analysis of user reviews.

This study found that most users of educational MAR apps encountered various problems that hamper their experience (65% of reviews classified as problem discovery). Only a few apps appeared in the "high install" range, suggesting that the adoption of educational MAR apps is still in its infancy. More AR apps have appeared since 2017 due to the availability of Google and Apple SDKs, which may suggest that the adoption of the technology can reach maturity in the near future if the new apps are more liked by the users and developers leverage the recent advances of AR.

Although marker-based AR was the most popular technology used in the apps, its practical shortcomings may block AR adoption, as expressed by some users. Improving upon these shortcomings and focusing on future developments of AR, such as marker-less technologies, may improve the user experience and increase the adoption of the technology.

Among the problems experienced by users of the current MAR apps, the application's functionality was the most popular one. Therefore, exploring in more detail what are the causes of these problems may provide a better chance to enhance the adoption of the technology. However, since users often do not provide detailed information, these results need to be interpreted with caution. Regarding the "helpfulness" of AR apps for users, this study has found that some users, regardless of the limitations of the apps, can recognize specific scenarios where AR apps are helpful. These observations align with what several AR studies in educational settings have acknowledged regarding the advantages of education. However, unlike these controlled studies, the reviews showed some real-life examples from a subjective perspective of the user.

The most popular demands among users were content-related. Users generally request additional content as most apps only provide a few models for visualization, target images, or limited free content. Future app developments can be concentrated on creating apps with a more comprehensive experience other than merely displaying a limited amount of multimedia.

Finally, this study has found that most apps in the data-set target childhood education and the most popular field of exercise is science. These findings may offer opportunities for developers and scholars interested in expanding the use of AR in education. Still, continuous efforts are needed to make AR more accessible to a more diverse audience. This study did not assess the "educational" value of the apps, and therefore further research with a stronger focus on educational purpose is suggested. Future studies can include, for example, the assessment of multimedia learning principles inside the public MAR apps. On the other hand, developers could focus on providing detailed explanations about each application, its use, and certainly validated content to improve the user experience of educational MAR apps users.

Undoubtedly, future research towards addressing technical issues must be a priority for researchers to ensure that AR can be well-integrated with society, regardless of its use. More significant multidisciplinary efforts are needed to ensure that well-develop MAR apps with relevant educational content are available for more people no matter their level of education, country, or current device.

To aim for increased AR adoption as an educational tool, developers, educators, designers, and scholars may understand and address user concerns while boosting learners' educational value and experience. Altogether, academic research and public sources, such as user reviews explored in this study may provide the relevant information needed to achieve this goal.

Acknowledgements. This project has received funding from the European Union's EU Framework Program for Research and Innovation Horizon 2020 under Grant Agreement 812716.

References

1. Apple Developer (2021). https://developer.apple.com/augmented-reality/
2. Google Developers (2021). https://developers.google.com/ar
3. Keyword Tool (2021). https://keywordtool.io/app-store
4. Vuforia Engine (2021). https://library.vuforia.com/getting-started/overview.html
5. Ahasanuzzaman, M., Hassan, S., Bezemer, C.P., Hassan, A.E.: A longitudinal study of popular AD libraries in the Google Play Store. Empir. Softw. Eng. **25**(1), 824–858 (2020). https://link.springer.com/article/10.1007/s10664-019-09766-x
6. Akçayir, M., Akçayir, G.: Advantages and challenges associated with augmented reality for education: a systematic review of the literature. Educ. Res. Rev. **20**, 1–11 (2017). https://doi.org/10.1016/j.edurev.2016.11.002
7. Alqahtani, F., Orji, R.: Insights from user reviews to improve mental health apps. Health Inform. J. **26**(3), 2042–2066 (2020). https://doi.org/10.1177/1460458219896492
8. AppBrain: Android and Google Play statistics, development resources and intelligence (2020). https://www.appbrain.com/stats
9. Arifin, Y., Sastria, T.G., 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
10. Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Comput. Graph. Appl. **21**(6), 34–47 (2001). https://doi.org/10.1109/38.963459
11. Bacca, J., Baldiris, S., Fabregat, R., Graf, S., Kinshuk: Augmented Reality Trends in Education: a systematic review of research and applications. J. Educ. Technol. Soc. **17**(4), 133–149 (2014). http://www.jstor.org/stable/jeductechsoci.17.4.133
12. Boshell, B.: Average App File Size: Data for Android and iOS Mobile Apps (2017). https://sweetpricing.com/blog/2017/02/average-app-file-size/
13. Brown, M., et al.: 2020 educause horizon report teaching and learning edition. Tech. rep., EDUCAUSE, Louisville, CO (2020). https://www.learntechlib.org/p/215670
14. Caudell, T., Mizell, D.: Augmented reality: an application of heads-up display technology to manual manufacturing processes. In: Proceedings Twenty-Fifth Hawaii International Conference System Science, no. 2, pp. 659–669. IEEE (1992). https://doi.org/10.1109/HICSS.1992.183317
15. Chen, H., Dai, Y., Meng, H., Chen, Y., Li, T.: Understanding the characteristics of mobile augmented reality applications. In: 2018 IEEE International Symposium Performance Analysis of Systems and Software, pp. 128–138. IEEE (2018). https://doi.org/10.1109/ISPASS.2018.00026
16. Cheng, K.H., Tsai, C.C.: Affordances of augmented reality in science learning: suggestions for future research. J. Sci. Educ. Technol. **22**(4), 449–462 (2013). https://doi.org/10.1007/s10956-012-9405-9
17. Clement, J.: Statista (2020). https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/
18. D'Heureuse, N., Huici, F., Arumaithurai, M., Ahmed, M., Papagiannaki, K., Niccolini, S.: What's app?: A wide-scale measurement study of smart phone markets. ACM SIGMOBILE Mob. Comput. Commun. Rev. **16**(2), 16–27 (2012). https://doi.org/10.1145/2396756.2396759
19. Dirin, A., Laine, T.: User experience in mobile augmented reality: emotions, challenges, opportunities and best practices. Computers **7**(2), 33 (2018). https://doi.org/10.3390/computers7020033

20. Garzón, J., Acevedo, J.: Meta-analysis of the impact of augmented reality on students' learning gains. Educ. Res. Rev. **27**, 244–260 (2019). https://doi.org/10.1016/j.edurev.2019.04.001
21. Garzón, J., Pavón, J., Baldiris, S.: Systematic review and meta-analysis of augmented reality in educational settings. Virtual Real. **23**(4), 447–459 (2019). https://link.springer.com/10.1007/s10055-019-00379-9
22. Glassey, J., Magalhães, F.D.: Virtual labs - love them or hate them, they are likely to be used more in the future. Educ. Chem. Eng. **33**, 76–77 (2020). https://doi.org/10.1016/j.ece.2020.07.005
23. Harborth, D., Pape, S.: Investigating privacy concerns related to mobile augmented reality apps - a vignette based online experiment. Comput. Human Behav. **122**, 106833 (2021). https://linkinghub.elsevier.com/retrieve/pii/S0747563221001564
24. Iacob, C., Harrison, R.: Retrieving and analyzing mobile apps feature requests from online reviews. In: 2013 10th Working Conference on Mining Software Repositories, San Francisco, CA, USA, pp. 41–44. IEEE (2013). https://doi.org/10.1109/MSR.2013.6624001
25. Iacob, C., Harrison, R., Faily, S.: Online reviews as first class artifacts in mobile app development. In: Memmi, G., Blanke, U. (eds.) MobiCASE 2013. LNICST, vol. 130, pp. 47–53. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-05452-0_4
26. Ibáñez, M.B., Delgado-Kloos, C.: Augmented reality for STEM learning: a systematic review. Comput. Educ. **123**, 109–123 (2018). https://doi.org/10.1016/j.compedu.2018.05.002
27. Irshad, S., Rohaya Bt Awang Rambli, D.: User experience of mobile augmented reality: a review of studies. In: 2014 3rd International Conference on User Science and Engineering, pp. 125–130. IEEE (2014). https://doi.org/10.1109/IUSER.2014.7002689
28. ISO: ISO 9241–210: Ergonomics of human-system interaction Part 210: Human-centred design for interactive systems (2019)
29. Jeffery, B.: The Kids App Market (2019). https://www.bjornjeffery.com/2019/05/31/the-kids-app-market-a-strategic-overview/
30. Jo, M.: google-play-scraper (2020). https://github.com/JoMingyu/google-play-scraper
31. Kourouthanassis, P.E., Boletsis, C., Lekakos, G.: Demystifying the design of mobile augmented reality applications. Multimed. Tools Appl. **74**(3), 1045–1066 (2015). https://doi.org/10.1007/s11042-013-1710-7
32. Li, H., Gupta, A., Zhang, J., Flor, N.: Who will use augmented reality? An integrated approach based on text analytics and field survey. Eur. J. Oper. Res. **281**(3), 502–516 (2020). https://doi.org/10.1016/j.ejor.2018.10.019
33. Mahmood, A.: Identifying the influence of various factor of apps on google play apps ratings. J. Data Inf. Manage. **2**(1), 15–23 (2020). https://doi.org/10.1007/s42488-019-00015-w
34. Mekni, M., Lemieux, A.: Augmented reality: applications, challenges and future trends. In: 13th International Conference Applied Computer Computational Science, pp. 205–214. WSEAS Press (2014). http://www.wseas.us/e-library/conferences/2014/Malaysia/ACACOS/ACACOS-29.pdf
35. Nicholas, J., Fogarty, A.S., Boydell, K., Christensen, H.: The reviews are in a qualitative content analysis of consumer perspectives on apps for bipolar disorder. J. Med. Internet Res. **19**(4), e105 (2017)

36. Nowacki, P., Woda, M.: Capabilities of ARCore and ARKit platforms for AR/VR applications. In: Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T., Kacprzyk, J. (eds.) DepCoS-RELCOMEX 2019. AISC, vol. 987, pp. 358–370. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-19501-4_36
37. Olano, F.: Google-Play Scraper (2020). https://github.com/facundoolano/google-play-scraper#search
38. Olsson, T., Salo, M.: Online user survey on current mobile augmented reality applications. In: 2011 10th IEEE International Symposium Mixed and Augmented Reality, pp. 75–84. IEEE (2011)
39. Pagano, D., Maalej, W.: User feedback in the appstore: an empirical study. In: 2013 21st IEEE International Requirements Engineering Conference, pp. 125–134. IEEE (2013). https://doi.org/10.1109/RE.2013.6636712
40. Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C.A., Canfora, G., Gall, H.C.: How can i improve my app? Classifying user reviews for software maintenance and evolution. In: 2015 IEEE International Conference on Software Engineering, Bremen, Germany, no. 2, pp. 281–290. IEEE (2015). https://doi.org/10.1109/ICSM.2015.7332474
41. Panichella, S., Di Sorbo, A., Guzman, E., Visaggio, C.A., Canfora, G., Gall, H.C.: ARdoc: app reviews development oriented classifier. In: Proceedings 2016 24th ACM SIGSOFT International Symposium Foundations of Software Engineering - FSE 2016, vol. 13–18-Nove, pp. 1023–1027. ACM Press, New York (2016). https://doi.org/10.1145/2950290.2983938
42. Papadakis, S., Kalogiannakis, M.: Mobile educational applications for children: what educators and parents need to know. Int. J. Mob. Learn. Organ. **11**(3), 256 (2017). https://doi.org/10.1504/IJMLO.2017.085338
43. Pence, H.E.: Smartphones, smart objects, and augmented reality. Ref. Libr. **52**(1–2), 136–145 (2010). https://doi.org/10.1080/02763877.2011.528281
44. Qiao, X., Ren, P., Dustdar, S., Liu, L., Ma, H., Chen, J.: Web AR: a promising future for mobile augmented reality - state of the art, challenges, and insights. Proc. IEEE **107**(4), 651–666 (2019). https://doi.org/10.1109/JPROC.2019.2895105
45. Qiao, X., Ren, P., Nan, G., Liu, L., Dustdar, S., Chen, J.: Mobile web augmented reality in 5G and beyond: challenges, opportunities, and future directions. China Commun. **16**(9), 141–154 (2019). https://doi.org/10.23919/JCC.2019.09.010
46. Radu, I.: Augmented reality in education: a meta-review and cross-media analysis. Pers. Ubiquitous Comput. **18**(6), 1533–1543 (2014). https://doi.org/10.1007/s00779-013-0747-y
47. Roberto, P., Emanuele, F., Primo, Z., Adriano, M., Jelena, L., Marina, P.: Design, large-scale usage testing, and important metrics for augmented reality gaming applications. ACM Trans. Multimed. Comput. Commun. Appl. **15**(2), 1–18 (2019). https://doi.org/10.1145/3311748
48. Schneider, S.L.: The international standard classification of education 2011. In: Birkelund, G.E. (ed.) Cl. Stratif. Anal., Comparative Social Research, vol. 30, pp. 365–379. Emerald Group Publishing Limited (2013). https://doi.org/10.1108/S0195-6310(2013)0000030017
49. Thach, K.S.: User's perception on mental health applications: a qualitative analysis of user reviews. In: 2018 5th NAFOSTED Conference Information Computer Science, pp. 47–52. IEEE (2018). https://doi.org/10.1109/NICS.2018.8606901
50. Tian, Y., Nagappan, M., Lo, D., Hassan, A.E.: What are the characteristics of high-rated apps? A case study on free android applications. In: 2015 IEEE International Conference on Software Maintenance and Evolution, pp. 301–310. IEEE (2015). https://doi.org/10.1109/ICSM.2015.7332476

51. Wojciechowski, R., Cellary, W.: Evaluation of learners' attitude toward learning in ARIES augmented reality environments. Comput. Educ. **68**, 570–585 (2013). https://doi.org/10.1016/j.compedu.2013.02.014
52. Wu, H.K., Lee, S.W.Y., Chang, H.Y., Liang, J.C.: Current status, opportunities and challenges of augmented reality in education. Comput. Educ. **62**, 41–49 (2013). https://doi.org/10.1016/j.compedu.2012.10.024