https://doi.org/10.37546 /JALTSIG.CALL.PCP2022-04

4

Ready, set…? End-user testing of an AR treasure hunt learning game

Samuel Taylor and Adam Stone

Abstract

Formative user testing, aimed at obtaining feedback from students, instructors, and administrators at a higher education institution in Japan, was undertaken on an AR treasure hunt learning game app. It involved the collection of usage data from data-mining tools, anecdotal evidence from surveys and interviews, and descriptive evidence from observations of tester behavior to investigate aspects of app learning, engagement, usability. Analysis indicated that testers (N=21) rated learning significantly higher than both engagement ($p<0.01$) and usability ($p<0.01$), and that engagement was rated significantly higher than usability ($p<0.05$). Observations noted indicators of engagement, frustration, and unanticipated patterns of usage. Interviews with testers raised issues concerning usability and engagement, and provided suggested changes to the app. These were used to make improvements that, while impacting learning, also facilitated the classroom implementation of the app in pedagogy. In conclusion, it is suggested that user testing is a vital part of educational technology impact evaluations, and that key aims of the design and development of educational apps should be the easy access, enabled through high usability and engagement, of learning content situated in contexts that provide opportunities for further or applied use.

本研究では、日本の高等教育機関の学生、講師、管理者からのフィードバックを得ることを目的とした形成的ユーザーテストを、AR宝探し学習ゲームアプリで実施した。データマイニングツールによる利用データの収集、アンケートやインタビューによる逸話的証拠の収集、調査対象者の行動観察による記述的証拠の収集により、アプリの学習、エンゲージメント、ユーザビリティの側面を調査した。分析によると、調査対象者（N=21）は学習をエンゲージメント（$p<0.01$）およびユーザビリティ（$p<0.01$）の両方よりも有意に高く評価し、エンゲージメントはユーザビリティ（$p<0.05$）よりも有意に高く評価されたことが分かった。また、観察では、エンゲージメント、フラストレーション、そして予期せぬ使用パターンが指摘された。調査対象者へのインタビューでは、ユーザビリティとエンゲージメントに関する問題点が指摘され、アプリの変更案が提示された。これらは、学習に影響を与える一方で、教育上、アプリを教室で容易に実施するた

めに使用された。結論としては、教育工学におけるインパクト評価には、ユーザーテストが不可欠であり、教育用アプリの設計・開発の主な目的は、高いユーザビリティとエンゲージメントによって可能になり、さらなる活用や応用の機会を提供する上で、学習コンテンツへの容易なアクセスであることを提案した。

Keywords: Mobile learning, Augmented Reality, Digital game-based learning, User testing, Educational apps

Introduction

"KSU Treasure Hunter" is a campus guide smartphone application (app) designed and developed by the authors at Kyushu Sangyo University (KSU), Fukuoka, Japan. The app is part of a research project (see Taylor, 2020; Taylor & Stone, 2018; Taylor et al., 2019) investigating how Augmented Reality (AR) technology can facilitate active participation in Japanese higher education to meet the goals of acquiring lifelong learning abilities and independent thinking skills (Japanese Ministry of Education, Culture, Sports, Science, and Technology, 2012). The project assumes that educational technology is not simply a tool for sustaining or streamlining prevailing methods of instruction (Christensen, 2008) but a means of transforming education by supporting new kinds of relationships between learners and technology (An & Oliver, 2020). In this context, AR is considered an emerging technology as it contains the potential to change the experience of education (Cukurova & Luckin, 2018). The creation and use of AR materials is also informed by a constructivist view of learning, in which learners construct their understanding through their experiences so that learning is active and exploratory, authentic and contextualized (Pegrum, 2014).

KSU Treasure Hunter presents practical information about campus services in a bilingual, pirate treasure hunt-themed AR learning game. Posters placed in campus facilities act as AR markers, requiring students to visit facilities to learn about available services. As such, the KSU Treasure Hunter app aims to transform the experience of learning about campus for KSU students. Its design is less concerned with improving the efficiency of knowledge transmission than with enabling learners to engage with the world in new ways. The magic of AR in revealing some kind of previously hidden truth (Norman, 2018) in exploratory and contextualized learning experiences increases learner interest (Godwin-Jones, 2016), and contributes to a positive learner experience (Taylor & Stone, 2018). The AR reveals campus information in the form of a learning game. Instructional goals are mapped onto the first three levels of Bloom's taxonomy (Armstrong, 2019) in a 3-level structure, detailed in Taylor (2020), that engages app users in progressively higher-level thinking skills. As such, pedagogy is embedded in a fun learning process that involves an abstraction of reality or element of fantasy in the teaching process (Boller & Kapp, 2017). The learning game players learn the names of campus facilities and services, understand what the services entail, and apply knowledge of the services in order to solve problems commonly encountered by students. Table 1 provides an example of the learning goals, learner

actions, and examples of learning content for one service available at the KSU library.

Table 1

The learning game goals, learner actions, and examples of learning content for one service available at the KSU library

		Level 1	Level 2	Level 3
Learning goal		learn the names of campus facilities and services	understand what the service entails	apply knowledge of the services in order to solve problems commonly encountered by students
Learner action		match the name of a service presented by the pirate NPC with the name of the same service presented on a jewel rotating around the AR marker	match a description of a service presented by the pirate NPC with the name of the same service presented on a jewel rotating around the AR marker	match a problem presented by a student NPC with the name of the same service presented on a jewel rotating around the AR marker
Learning content	Prompt	career guidance literature	Preparation is everything… make informed decisions on your career choices, and know what you need to do to get the job you want.	I would like some information on jobs; I want to know if I need any special qualifications to get the kind of job I'm interested in.
	Correct response	career guidance literature		

The testing of emerging educational technology is part of iterative cycles of testing, design, development, and evaluation aimed at achieving incremental improvements to learning and teaching (Cukurova & Luckin, 2018). User testing, 'a group of usability evaluation methods that involve user participation' (Riihiaho, 2018, p. 257), is recognized as a critical part of the development process (Crowther et al., 2004). For educational apps user testing should include evaluations of usability, engagement, and learning (McQuiggan et al., 2015). Testing involves the collection of anecdotal evidence gained from surveys and interviews with testers, and descriptive evidence gained from observational note-taking of user behavior (Cukurova & Luckin, 2018) and indicators of enjoyment, engagement, and frustration (Sim et al., 2006).

It is traditional to assess usability by taking measures of the users' performance, by noting where users have difficulties with the interface, and by asking the users for their opinions of the product (Sim et al., 2006). A small number of testers, between five and ten, are sufficient for identifying problems (Riihiaho, 2018). Tests should be realistic and represent actual use (Riihiaho, 2018), however testing in lab conditions as opposed to in the field has several advantages: it is less time consuming and leads to less interruptions and, most importantly, there is no difference in the number of user interface problems identified by testers (Kaikkonen et al., 2005). Testing can also involve analysis of in-game user interactions, obtained from data mining tools built into the app, to investigate whether users are showing expected sequences of behavior and whether there are any over- or underused features (McQuiggan et al., 2015).

Ferrer et al. (2013) user tested an AR game to investigate usability, learning, and motivation. Usability concerned the design of the AR, and analysis indicated that, despite the AR game having reduced usability compared to a desktop version because it took longer to play, the AR game was more motivating for learners. Learning was found to be unaffected by mode. The measurement of the motivation of the college-student testers was achieved using survey responses. Learning was measured using pre-post measures. Kucirkova et al. (2014) evaluated engagement with educational software using analysis of user interactions with the software and with each other. It was demonstrated how the engagement of nursery school-aged children with a Spanish language learning app was shown to be affected by both usability and learning content. In particular, it was suggested that ease of access to learning content that encourages open-ended accomplishments, rather than drill-and-practice activities, facilitates greater engagement.

Method

The testing of KSU Treasure Hunter was largely concerned with usability, particularly the user interface (UI), as such, it can be described as formative, as it was aimed at obtaining feedback from users to inform further development (Riihiaho, 2018). The testing detailed in this paper was the last round of testing before the app was made available for use by instructors and students. It involved the collection of usage data from data-mining tools in the app, anecdotal evidence from surveys and interviews, and descriptive evidence from observations of tester behavior.

Testers from three distinct stakeholder groups were recruited: students, instructors, and administrators. This was done to obtain feedback from the three populations involved in the use of the app upon it being made available, each of which may have different insights on the app's objectives and limitations (McQuiggan et al., 2015). A total of ten Japanese students participated voluntarily in testing, and their participation was not connected to their English classes or English proficiency in any way. The instructor group consisted of five English instructors – two Japanese, two English, and one American – and the administrator group consisted

of six Japanese administrative university employees. Demographic data, such as age and gender, were not recorded, and testers were assigned a letter of the alphabet, from A to U, to allow for anonymity.

Testing was conducted in lab conditions, in a single room on the university campus. The placing of the AR markers in the room meant that testers could experience the app content without having to visit campus facilities. Participants tested the app on their own devices and were free to choose in which language, English or Japanese, they used the app. As they concern usability, the operating system (0S) of the device used by each tester, and whether the tester used the app in English or Japanese and with or without sound, were recorded. Information on the first two items was retrieved from in-app data-mining tools, whereas the use of sound or otherwise was noted down on paper by a researcher during observations of tester behavior. Guidance was only provided when it concerned accessing the app by downloading it onto the tester's device. There was no time limit enforced for testing, and testing was considered to be complete when each tester perceived that the app experience had finished. Testers were then given the survey to complete. After that, interviews were conducted in either Japanese or English by a researcher, depending on the tester's preference, and were audio recorded. The researcher conducting the interviews also took notes of tester responses. Tester suggestions for improvements made during the interviews were listed and categorized by whether they aligned with aspects of app learning, usability, or engagement, and coded by aspects of the app design and development, such as, UI, AR design, system architecture, game design, and feature design.

Following McQuiggan et al. (2015) and Sim et al. (2006), the survey aimed at measuring tester opinions on app learning, usability, and engagement, and their overall opinion of it. Consequently, there were three, 5-point Likert items, shown in Table 2 below, for learning, usability, and engagement, and one, 5-point app rating item. The learning item asked to what degree the app is good for learning, the usability item asked to what degree the app is easy to use, and the engagement item asked to what degree the app is fun to use.

Table 2

Survey used in testing

Item	Likert scale				
This app is good for learning about campus. This app is easy to use. This app is fun to use.	1 (strongly disagree)	2 (disagree)	3 (neither)	4 (agree)	5 (strongly agree)
Please rate this app out of five.	1 (very bad)	2 (bad)	3 (neither)	4 (good)	5 (very good)

The survey data were analyzed, using independent samples *t* tests, for significant differences between the scores of the four survey items for the whole group of testers (N=21), across the three tester groups, and across three distinct usage groups of OS, language, and sound. This was done to investigate whether the app was significantly stronger or weaker in perceptions of learning, usability, and engagement, whether the tester groups experienced the app significantly differently, and whether different usage patterns significantly influenced perceptions of the tester experience.

The 19 interview questions were informed by the app checklist in McQuiggan et al. (2015), and decided upon by the researchers after consideration of the primary concerns regarding the continued design and development of the app. Table 3 below displays the questions split into categories of app purpose, learning, usability, and engagement. The questions were shared with testers without being split into categories. The Japanese translations of the questions are provided in Appendix A. Tester responses to the interview questions were reviewed to identify any specific suggestions made regarding changes to the app. These suggestions were considered in tandem with the observations of tester behavior to create a list of actionable changes.

Table 3

The interview questions used in testing

	Purpose
1	Do you think the app works well as a campus guide?
	Learning
2	Do you think that game progress is too easy or too difficult?
3	Do you think that game progress is too slow or too quick?
4	Do you think that the game provides enough feedback?
5	Do you think that the game encourages the player to think about how to use campus services?
6	Do you think that the game experience is relevant to real life?
7	Do you think that the game is an efficient way of learning about campus services?
8	Do you think that the game is an engaging way of learning about campus services?
	Usability
9	Were there any moments when the app was difficult to use?
10	Were there any moments when navigating through the app was unintuitive?
11	Were there any moments when the text was difficult to read?
12	Do you think that the app offers enough support on how to use it?

	Engagement
13	What parts of the app were the most fun?
14	What parts of the app were the least fun?
	General
15	What is a good point of this app?
16	What is a bad point of this app?
17	Is there anything about the app that you think should be changed?
18	Would you use this app again?
19	Do you have any other opinions on the app?

Results

Survey data

Survey data, shown in Table 4 below, showed that testers (N=21) rated learning the highest (5.0) and usability the lowest (4.0). Learning was rated significantly higher than both engagement ($p<0.01$) and usability ($p<0.01$), and engagement was rated significantly higher than usability ($p<0.05$). In the separate tester groups, administrators (n=6) rated the app significantly higher than students (n=10) ($p<0.01$) and instructors (n=5) ($p<0.05$), and rated engagement significantly higher than students ($p<0.05$) and instructors ($p<0.05$). There were no other significant differences.

Table 4

Ratings of app learning, engagement, and usability for total (N=21), student (n=10), instructor (n=5), and administrator (n=6) tester groups

	Total (N=21)	Student (n=10)	Instructor (n=5)	Administrator (n=6)
Learning	5.0	4.9	5.0	5.0
Engagement	4.5	4.4	4.0	5.0
Usability	4.0	3.8	4.0	4.0
App rating	4.5	4.3	4.2	5.0

Usage data

The majority of testers tested the app in Japanese (n=15, 71%), using iOS (n=14, 67%) with the sound off (n=13, 62%). Testers who initially used the app with no sound and then turned on the sound mid-testing are recorded under 'Both'. The

student tester group largely reflected the whole group (Japanese: n=9, 90%; iOS: n=9, 90%; sound off: n=8, 80%), whereas the instructor and administrator groups had a higher proportion of Android OS (n=3, 60%; n=3, 50%) and sound usage (n=3, 60%; n=3, 50%). The instructor group had the highest proportion of English usage (n=4, 80%). The data is provided in Table 5, below. Despite these variations there were no significant differences between language, OS, and sound usage groups in survey scores of learning, usability, engagement, or overall app rating.

Table 5

Tester usage data

Group	Language		OS		Sound		
	English	Japanese	iOS	Android	Yes	No	Both
Total (N=21)	6 (29%)	15 (71%)	14 (67%)	7 (33%)	6 (28%)	13 (62%)	2 (10%)
Student (n=10)	1 (10%)	9 (90%)	9 (90%)	1 (10%)	0 (0%)	8 (80%)	2 (20%)
Instructor (n=5)	4 (80%)	1 (20%)	2 (40%)	3 (60%)	3 (60%)	2 (40%)	0 (0%)
Administrator (n=6)	1 (17%)	5 (83%)	3 (50%)	3 (50%)	3 (50%)	3 (50%)	0 (0%)

Observations

Tester behavior indicated signs of both engagement and frustration with the app. Engagement, indicated by comments and laughter, was most often observed when testers used the game introduction scene, in which the player first meets the AR NPC guide character. Frustration, indicated by sighs and looking around the room, was observed when testers were in the three-level learning game. Two testers exhibited unanticipated patterns of usage by not using a button prompt on a level summary screen to continue to level two after completing level one. To continue with the game these testers quit the app and reopened it, continued their game from the main menu, accessed the next location from the map menu, and completed only the first level for each location.

Interviews

Testers were asked about indicators of frustration and it was reported that the learning game was too repetitive, and that level one, in particular, could be made more interesting by changing the game mechanics. Other issues raised concerned usability and engagement. Table 6, below, displays tester-made suggested changes, categorized by the user experience aspects of usability and engagement. There were nine suggestions relating to usability, five relating to engagement, and none relating to learning. Inferred by the researchers, the suggestions are

Table 6

Suggested changes provided by testers and the underlying problems inferred by the researchers

Suggested change	Underlying problem	Aspect of design and development
	Usability	
Use a real campus map.	User can't use map for function of locating campus services.	Feature design
Make the campus information screen more prominent.	User not using app as campus information guide.	Feature design
Make the campus information screen look different to the map screen.	The function of the campus information is unclear.	Feature design
Put menu button labels in Japanese in the Japanese version.	User can't understand what each menu button text means.	UI
Change English font to one easier to read.	Menu and learning game content is difficult to read.	UI
Add AR use guidance.	AR content can be difficult to access using AR marker.	AR design
Make some game content playable in device.	Holding up phone to play AR game is tiring. User is tied to AR marker.	AR design
Add a menu button to the language selection screen.	User unable to view menu before beginning a game.	System architecture
Add a menu button in AR game.	User unable to view menu while in AR game.	System architecture
	Engagement	
Add greater reward for completion of learning game.	User unengaged by experiential design.	Game design
Change color of AR game feedback text to indicate if response is correct or incorrect.	AR game learning feedback is unclear.	Game design
Indicate completion of a location more clearly on the map screen.	AR game progress feedback is unclear.	Game design
Add variety to AR game.	AR game mechanics become repetitive.	Game / AR design
The pirate NPC should speak	Pirate NPC not as prominent as it could be.	AR design

accompanied by description of the underlying problem that prompted the suggestion, and the aspect of the app design and development that each suggestion concerns. The suggestions relating to usability concerned feature design, UI, AR design, and system architecture. In contrast, the suggestions relating to engagement largely concerned game and AR design, including game learning and progress feedback and game mechanics.

Discussion

Learning

In general terms, analysis of survey and usage data indicated that instructors and students had similar experiences of the app, while administrators experienced the app more positively than the other two groups, seen in higher scores for app engagement and overall rating. However, learning was the most positively experienced aspect of the app, and was not experienced differently by tester group or usage pattern. These findings suggest that the learning content is sufficient for the intended purpose of the app. Alternatively, it may also indicate that formative user testing is not the best means of collecting evidence on the impact of an emerging technology on learning.

Usability

Usability issues were shown to exist by usability being the least positively experienced aspect of the app. The usability issues were perceived by all tester groups, but they weren't distinguished by device OS, language, or sound usage. These results can be seen to reflect the tendency for AR content to foster a positive learner experience (Taylor & Stone, 2018) and motivate students to learn (Ferrer et al., 2013) despite the presence of usability issues. Analysis of the suggested changes gained from the interviews with testers supports the survey and observational data in indicating that usability issues outnumber those concerning engagement or learning. More suggestions were made that relate to usability than engagement. The most frequently occurring aspect of the app design and development in suggestions related to usability was feature design. The underlying problem of each feature design issue concerned expanding the scope of the app from a learning game to a campus guide. This suggests that the usability of the learning game is affected by the extent to which it is integrated into part of a larger campus information system. The provision of open-ended accomplishments that encourage high degrees of engagement and increase the educational value of the app (Kucirkova et al., 2014) needs to explicitly concern this context. Open-ended accomplishments for a campus knowledge system can be encouraged using a mobile app's ability to connect the user to the wider world. The challenge is to integrate activities, such as booking an appointment for a service, using links to external information related to campus services, such as resume writing tips, or using links to KSU sites, such as the LERC Moodle system, into the learning game

so that learning is active and exploratory, authentic and contextualized (Pegrum, 2014).

As a result of formative user testing, changes made to the app to improve usability concerned feature design and UI. The campus information screen was redesigned to be more distinct from the campus map screen to clarify the apps intended use as a campus guide. UI buttons were redesigned to be more clearly recognizable as interactive features to avoid the incorrect patterns of usage described above. The English font was changed to improve the clarity of learning content. All app content was made bilingual to improve the accessibility of learning content. Figure 1 below displays the changes to the UI buttons and English font on the game level three summary screen.

Figure 1

The changes to the UI buttons and English font as a result of user testing

pre-user testing post-user testing

Engagement

Engagement was the most positively experienced aspect of the app after learning, and it was more positively experienced by administrators than by the other tester groups. It was not affected by sound usage, device OS, or language use. The most frequently occurring aspect of the app design and development in the suggestions relating to engagement was learning game design. Therefore,

learner actions, determined by game mechanics, and learning feedback can be seen as particularly important for user engagement in learning games, as they make learning content accessible, engaging and comprehensible. The issue of the accessibility of learning content, the importance of which is demonstrated by Kucirkova et al. (2014), also concerns usability, and the results of testing suggest that accessibility is clearly related to app feature design, UI, AR design, system architecture, and game design.

The suggested changes regarding learning game feedback were enacted, so that green and red text was used in learning feedback to indicate correct and incorrect responses. In regard to reducing the repetitiveness of the learning game, the following modification to level one was proposed but not enacted due to time constraints. The AR marker gives access to a grid of 15 jewels, which are not tied to the AR marker but are playable in-device. The pirate NPC presents the task facing the player: to find and select three of each service as quickly as they can. Each set of three jewels is collected upon being selected in a sequence. This correct response also receives a point score, but incorrect responses, such as selecting a different service to the one selected previously, lead to a points deduction. A timer is added above the grid of jewels, and the time taken for the player to complete the level is used with the number of correct and incorrect responses to calculate a final score. This proposed change is shown in Figure 2 below. Adding time pressure is a means of increasing engagement, whereas the use of game content not tied to the AR marker improves usability by not requiring the player to hold their device in front of the AR marker for the duration of the game level.

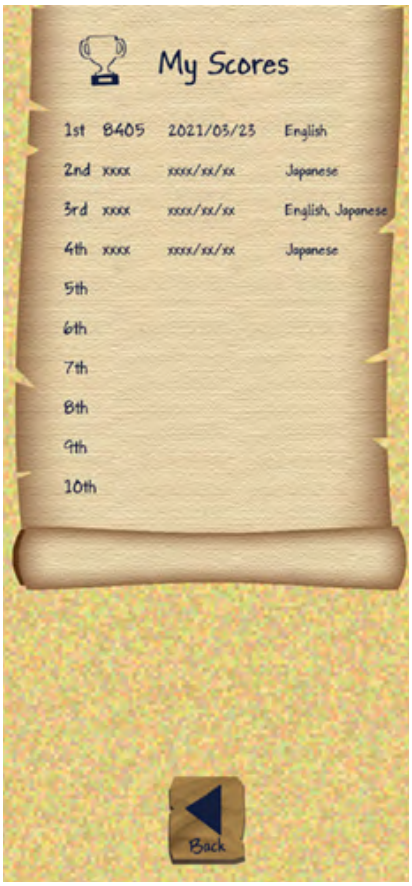
Figure 2

A possible modification to learning game level one to improve usability and engagement

Pedagogy

The changes detailed above all modify the app design, and in-app experience and behaviors, and, therefore, can be seen to concern learning (Cukurova & Luckin, 2018). There was also a change made to the app that concerned pedagogy. A 'My Scores' screen, shown in Figure 3 below, was added to the high scores that displays only the device scores. While this adds app personalization, an important feature of educational apps (McQuiggan et al., 2015), this screen also displays in which language the app was played. This information is included as a means for instructors to check that learners have used the app appropriately when implemented in classroom use.

Figure 3
The newly-added My Scores screen



Conclusion

The current study provides anecdotal and descriptive evidence of the development and testing stage of the KSU Treasure Hunter app. There are several drawbacks of the current research that temper any implications that can be made regarding the results of analysis. Firstly, it is necessarily specific to the context of KSU in learning content, and by not adopting any of the various frameworks available for the evaluation of educational apps (see McQuiggan et al., 2015) the research does not allow for the comparison of KSU Treasure Hunter with other similar apps. Secondly, a simple testing design makes contributing to usability testing (see Riihiaho, 2018), and the measurement of user engagement (see Kucirkova et al., 2014) and learning (see Ferrer et al., 2013) difficult.

Despite these limitations, the research has shown that user testing is a vital part in achieving improvements to the learning and teaching associated with emerging educational technology (Cukurova & Luckin, 2018). Following Riihiaho (2018), testing in lab conditions with a small number of testers was sufficient for

identifying problems with the app in regard to both usability and engagement. As suggested by Sim et al. (2006), asking user opinions of the app was also a productive means of informing design and development. The combination of survey, observational, and interview data provided the app designers and developers with a roadmap for future iterations, but also suggested that issues concerning learning may be best explored through other types of data collection. Future research concerned with learning could involve the collection of correlational data, for example on the relationship between user perceptions of the app and user performance as indicated by game scores. Additionally, the collection of causal evidence on the pedagogical soundness of the learning game could be gained from the classroom implementation of the app in a pedagogical sequence using pre and post-test data. In this way, future research can take advantage of the opportunity the app provides to engage in a holistic approach to an emerging technology impact evaluation that includes various types of evidence of the effect of the app on learning and pedagogy (Cukurova & Luckin, 2018).

KSU Treasure Hunter is a unique instructional and promotional tool by its integration of a campus knowledge system and an AR learning game. Formative testing suggests that key aims of the design and development of educational apps should be the easy access of learning content, that access is enabled through high usability and engagement, and that learning content should be situated in contexts that provide opportunities for further or applied use. From this perspective, the impact of KSU Treasure Hunter can be improved by enhancing campus knowledge system features and improving the learning game to include more variety in gameplay and the possibility of contextually situated open-ended accomplishments. Whether these adjustments bring the app closer to enabling a transformation of education (An & Oliver, 2020) is, as yet, unanswered.

References

An, T. & Oliver, M. (2020). What in the world is educational technology? Rethinking the field from the perspective of the philosophy of technology. *Learning, Media and Technology, 46*(1), 6–19. https://doi.org/10.1080/17439884.2020.1810066

Boller, S. & Kapp, K. (2017). *Play to learn: Everything you need to know about designing effective learning games*. ATD Press.

Christensen, C. M., Horn, M. B., & Johnson, C. W. (2008). *Disrupting class: How disruptive innovation will change the way the world learns*. McGraw-Hill.

Crowther, M. S., Keller, C. C., & Waddoups, G. L. (2004). Improving the quality and effectiveness of computer- mediated instruction through usability evaluations. *British Journal of Educational Technology, 35(*3), 289–303.

Cukurova, M. & Luckin, R. (2018). Measuring the impact of emerging technologies in education: A pragmatic approach. In J. Voogt et al (Eds.), *Second handbook of information technology in primary and secondary education* (pp. 1181–1199). Springer International Publishing. https://doi.org/10.1007/978-3-319-71054-9_81

Ferrer, V., Perdomo, A., Rashed-Ali, H., Fies, C., & Quarles, J. (2013). How does usability impact motivation in augmented reality serious games for education? *2013 5th international conference on games and virtual worlds for serious applications (VS-GAMES)* (pp. 1~8). http://dx.doi.org/10.1109/VS-GAMES.2013.6624233

Godwin-Jones, R. (2016). Augmented reality and language learning: From annotated vocabulary to place-based mobile games. *Language Learning and Technology, 20*(3), 9–19.

Japanese Ministry of Education, Culture, Sports, Science, and Technology. (2012). *Towards a qualitative transformation of university education for building a new future - Universities fostering lifelong learning and the ability to think independently and proactively.* Retrieved October 22, 2018, from: http://www.mext.go.jp/en/publication/report/title01/detail01/1380275.htm

Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability testing of mobile applications: A comparison between laboratory and field testing. *Journal of Usability Studies, 1*(1), 4–16.

Kucirkova, N., Messer, D., Sheeny, K., & Panadero, C. F. (2014). Children's engagement with educational iPad apps: Insights from a Spanish classroom. *Computers and Education, 71*, 175–184. https://doi.org/10.1016/j.compedu.2013.10.003

McQuiggan, S., Kosturko, L., McQuiggan, J., & Sabourin, J. (2015). *Mobile learning: A handbook for developers, educators, and learners*. John Wiley & Sons, Inc.

Norman, J. (2018). *L. Frank Baum's "The Master Key" imagines a kind of augmented reality.* http://www.historyofinformation.com/expanded.php?id=4698. Accessed November 14th, 2018.

Pegrum, M. (2014). *Mobile learning: Languages, literacies and cultures.* Palgrave Macmillan.

Riihiaho, S. (2018). Usability testing. In K. L. Norman & J. Kirakowski (Eds.), *The Wiley Handbook of Human Interaction* (Vol. 1). John Wiley & Sons Ltd.

Sim, G., MacFarlane, S., & Read, J. (2006). All work and no play: Measuring fun, usability, and learning in software for children. *Computers and Education, 46*, 235–248.

Taylor, S. (2020). Designing an augmented reality-enabled smartphone campus guide learning game. In P. Fotaris (Eds.) *Proceedings of the 13th international conference on game based learning ECGBL 2020* (pp. 604–612).

Taylor, S. & Stone, A. (2018). Smartphone augmented reality for EFL learners. *2018 PanSIG Journal*, 258–263.

Taylor, S., Stone, A. & Witkin, N. (2019). Developing an educational and promotional augmented reality learning game smartphone application. *EdMedia + Innovative Learning, 2019*(1), 935–938.

Appendix

The Japanese version of the interview questions used in testing.

1. アプリはキャンパスガイドとして適していると思いますか？
2. ゲームの進行について簡単すぎると思いますか、それとも難しすぎると思いますか？
3. ゲームの進行について遅すぎますか、早すぎますか？
4. ゲームは十分なフィードバックを提供していると思いますか？
5. ゲームは学内施設利用情報について考えるきっかけになると思いますか？
6. ゲームの体験は実生活に関連していると思いますか？
7. このゲームでは学内施設利用情報について効率的に学ぶことができると思いますか？
8. このゲームは学内施設利用情報について学ぶための魅力的な方法だと思いますか？
9. アプリが使いづらかった場面はありましたか？
10. アプリの操作が直感的でない場面はありましたか？
11. テキストが読みにくい場面はありましたか？
12. 使い方のサポートは十分だと思いますか？
13. アプリのどの部分が一番面白かったですか？
14. アプリのどの部分が一番面白くなかったですか？
15. アプリの良い点は何だと思いますか？
16. アプリの悪い点は何だと思いますか？
17. アプリについて、変更すべきだと思う点はありますか？
18. 再度このアプリを使いたいと思いますか？
19. アプリについて、他に何かご意見はありますか？

Authors' bios

Samuel Taylor received his MA in TESOL with distinction from the Institute of Education, University College London in 2017. His research interests include speaking, student perceptions of education, and educational technology.

Adam Stone received his Master of Applied Linguistics from Massey University, New Zealand in 2016. His research interests are English vocabulary acquisition, Computer Assisted Language Learning (CALL), Mobile Learning, educational software design and development, and Mixed, Augmented and Virtual Realities (MAVR).