ARETE – DELIVERABLE

WP4 – D4.1 ARETE Use Scenarios

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No 856533

Deliverable number:	D4.1
Due date:	31st of January 2021
Nature[1]:	R
Dissemination Level:	PU
Work Package:	4
Lead Beneficiary:	ULE
Beneficiaries:	UCD, CLB, WWL, SVU, EUN, CNR, WUE, VIC, TOU

Document History

Version	Date	Description
0.1	19/12/2020	Initial complete draft
0.2	06/01/2021	Release version for comment
0.3	25/01/2021	Comments addressed

[1] Nature:
R = Report, P = Prototype, D = Demonstrator, O = Other
Dissemination level
PU = Public
PP = Restricted to other programme participants (including the Commission Services)
RE = Restricted to a group specified by the consortium (including the Commission Services)
CO = Confidential, only for members of the consortium (including the Commission Services)
Restraint UE = Classified with the classification level "Restraint UE" according to Commission Decision 2001/844 and amendments
Confidential UE = Classified with the mention of the classification level "Confidential UE" according to Commission Decision 2001/844 and amendments
Secret UE = Classified with the mention of the classification level "Secret UE" according to Commission Decision 2001/844 and amendments

Disclaimer

This deliverable is a draft and is provided for information only.
The information contained herein is subject to change and it does not represent the opinion of the European Commission (EC), and the EC is not responsible for any use that might be made of information contained.
The final version of the deliverable will be published as soon as approved by EC.

Disclaimer

Table of Contents

Executive Summary

D4.1 is the first deliverable of WP4 – User-centred Interactive Design. It consists of an introduction (Section 1) and a conclusion (Section 4), which sandwich two major sections reporting on the work that WP4 has conducted in the period of Month 1 to 14.

Section 2 presents the empirical tasks of WP4, utilizing a range of established Human-centred Design (HcD) concepts and methods, which are briefly described in Section 2.1. In Section 2.2, 2.3 and 2.4, we describe the processes and outcomes of co-designing digital artefacts of Pilot 1, 2, and 3, respectively. Specifically, given the lack of access to end-users, the HCI team have assumed the role of proxy users and applied usability heuristics to evaluate the interaction design of mock-ups and prototypes, thereby providing feedback and improvement suggestions.

Section 3 presents the analytical tasks undertaken in WP4, namely a systematic literature review (SLR) on research studies pertaining to AR educational tools (ARETs) with a focus set on usability and user experience (UX) (Section 3.1). The range of publication year is 2000-2020. With the search string comprising the key terms - augmented reality, education, learning, design, evaluation and school – 714 records have been returned by the four databases (Section 3.2). After a series of screening/filtering steps, the final batch of 48 articles have been analysed and synthesized (Section 3.3 – 3.5). Examples of intriguing findings include (Section 3.5 - 3.6): (i) the application domains of the ARETs reviewed were largely STEM; (ii) only one study involved parents at home, who were untrained to provide support, resulting in frustration in all parties involved; (iii) the majority of the studies deployed markers despite the increasing sophistication of the markerless technology, which remains costly; (iv) the collaborative learning mode for the ARETs was predominant, but the empirical evidence of its effects as compared to the individual mode was lacking; (v) the number of attempts on applying AR to support children with special needs to learn remained disappointingly low. Implications of these and other insights for ARETE use scenarios have been drawn (Section 3.7).

Overall, despite the negative impacts of the pandemic, thanks to the ongoing close collaborations among the partners, WP4 achieved its main goal in the first project year, albeit regrettably not to its full extent.

1 Introduction

The main goal of WP4 - User-centred Interactive Design – is to identify, update and integrate, user-based insights into designing and developing the ARETE digital artefacts, rendering them to be highly useful, usable, desirable and pleasurable. Methodologically, we draw on the User/Human-centred Design (UcD/HcD) approaches, including Participatory Design (PD), Usability and User Experience (UX) methods and tools. They are applied for formative and summative evaluations of ARETE's digital artefacts with different fidelity levels.

WP4 is led by the University of Leicester (ULEIC) where a team of Human-Computer Interaction (HCI) specialists reside. Clearly, no HCI work can be done in a vacuum, the team needs to work closely with the other WP4 partners to ensure the attainment of its goal. Nonetheless, to insure transparency and clear responsibility, in the following text, the term 'the HCI team' is used to indicate our role in the related tasks.

The original deadline of this deliverable, D4.1, was M9. Unfortunately, the HCI work with heavy reliance on access to end-users – teachers and students – has severely been disrupted by the pandemic. The proposal to defer the release of D4.1 to M15 was made with the hope that the grave situation would be eased over the summer. To our dismay, the situation has not improved. The problem of inaccessible end-users has persisted since March 2020. Consequently, the HCI team need to rely on previous collaborative experience with teachers as well as school children and on analytic approaches (e.g. Heuristic Evaluation) without involving end-users. Specifically, one ongoing key task of WP4 is to co-create the use scenarios envisioned by the partners of the three Pilots from the interaction design perspective. The HCI team assumed the role of students interacting with the prototypes under evaluation when providing feedback (Section 2). Admittedly, results based on proxy users could be less ideal although the use of such alternatives is not uncommon, which could also help mitigate the delay of WP4's input to WP3, WP5 and WP6.

Originally, *in situ* observations at schools were planned to be carried out in order to understand the usage of educational technologies, including AR, in real-life contexts. Data so collected were meant to provide contextualised inputs to substantiate the Pilots' use scenarios. According to the project plan, ethics application documents needed to be prepared and approved (WP1) prior to any user-based studies being performed. In parallel, WP4 launched the process of a systematic literature review (SLR)[2] on research studies investigating AR educational tools (ARETs) with a focus on usability and UX design and evaluation (Section 3). Insights from the SLR could inform the planned in situ observations. Upon receiving the ethics approval, we were about to contact schools, asking for permission to carry out observational studies. But this was exactly the time when national lockdowns were executed in many of the European countries. This meant that in situ observations were becoming not so meaningful or even impossible. Given that classroom activities, if allowed at

[2] Some of the SLR processes and results presented in this deliverable have been submitted as a manuscript to *International Journal of Human Computer Interaction* (IJHCI), Elsevier, and is currently under review. While this manuscript focuses on the HCI aspect of ARETs, we aim to prepare another publication on the pedagogical aspect of ARETs.

all, were adapted to the COVID-secure arrangement instead of regular settings, we planned to conduct online interviews and web-based questionnaires to collect teachers' and students' experiences and visions of using AR. We have explored different dissemination channels, but the response rates were low. Details will be documented in D4.2 (M18).

In this deliverable one of the main foci is on the SLR, for which we followed the established PRISMA method and implemented the three stages: identification, screening/filtering, and synthesis. From the four bibliometric databases, we identified 714 records, which were narrowed down to the final batch of 48 quality articles. Results of the SLR not only allow us to position as well as enhance the Pilots' use scenarios but also enable us to draw implications for other use scenarios. Some of the implications such as identifying innovative methods to evaluate user experience with markerless AR may be realised within the lifetime of ARETE whereas broadening the range of stakeholders of ARETs, such as parents, museum curators, children with special needs, could entail further effort beyond the project's duration.

2 Applications of Human-Centred Design (HcD) Frameworks to the Pilots

In this section, we first present some basic HcD concepts and methods to be deployed in WP4. Then we describe the processes and outcomes of the WP4 work applied to the digital artefacts of the three Pilots.

2.1 Basic Human-centred Design (HcD) Concepts and Methods

As the focus of this deliverable is on the applications of HcD, we do not delve into the vast body of the related theoretical frameworks. Instead a concise summary of each key approach is presented as background.

2.1.1 Human-centred Design (HcD)

HcD is variously known as User-centred Design (UCD) (9241-11:1998/9241-210:2010) of which the tenet is to involve end-users in the entire process of system development to ensure that their needs and preferences as user requirements for the system will be taken into account. The ongoing involvement of end-users is to enable their feedback on prototypes of different fidelity to be collected and addressed by the development team. This process can maximise user acceptance and adoption of the system delivered. In ARETE, we aim to involve teachers and students in the process of HcD.

2.1.2 Usability

Usability is defined as "extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use" (ISO 9241-11:2018). For the ARETE project, usability of the developed apps is especially important to ensure that no problems with the software impede the learning process and that the students and teachers are satisfied with the way they reach their (learning) goals.

2.1.3 User Experience (UX)

User Experience is defined as "user's perceptions and responses that result from the use and/or anticipated use of a system, product or service" (ISO 9241-210:2019). This is a very broad definition, but compared to usability it is a shift towards the experiential quality of the

interaction. For ARETE a positive User Experience with the developed apps is especially important, to make sure students and teachers enjoy using the software (and will continue to do so) to reach their learning goals.

2.1.4 Formative Evaluation

Formative evaluation is known as diagnostic evaluation, which is aimed to collect and analyse end-users' feedback to identify usability problems and improvement strategies, thereby improving the interaction quality of the artefact under scrutiny. To support the development efforts of other work packages (WP3 and WP5), WP4 performs formative evaluation of the prototypes (Section 2.2, 2.3 and 2.4), empirically with end-users (e.g. focus groups, surveys) and analytically without them (e.g. Heuristic Evaluation).

2.1.5 Summative Evaluation

Summative evaluation is conducted to collect input and feedback with the aim to check and rate the artefact under scrutiny. In the HcD process, it is usually performed at a key stage of the development process (e.g. final version, release version) to determine if the software reaches specific goals, for example, high levels of usability and user experience. For ARETE the summative evaluation of the HCI aspects will be incorporated into the Pilot evaluations through close collaborations between WP4 and WP6.

2.1.6 Focus Group

Focus Groups are semi-structured group interviews facilitating the discussion of topics that are of interest for the moderating researcher(s). Participants are encouraged to share their feelings and thoughts through prompts (e.g. presenting ideas) or asking them questions directly. For WP4, Focus Groups are performed with teachers to collect their input and feedback on scenarios and functionality options or interface design alternatives.

2.1.7 Heuristic Evaluation (HE)

In Heuristic Evaluations feedback on software artefacts is generated without end-user involvement by checking for compliance with or violation of usability heuristics (cf. the ten widely used ones proposed by Jakob Nielsen, 1994[3]). The main result of HE is a list of usability problems and their impact (e.g. severity). To support developers, this list typically includes recommended modifications that can be implemented to address and resolve the usability problems identified (Section 2.2 and 2.3).

2.2 Pilot 1: Interaction Design of WordsWorthLearning (WWL)

Two major evaluation tasks on the digital artefacts designed for Pilot 1 have been conducted, namely Heuristic Evaluation of the website (Section 2.2.1) and Feedback on the app's scripts (Section 2.2.2). The process and sample outcomes are reported in the following.

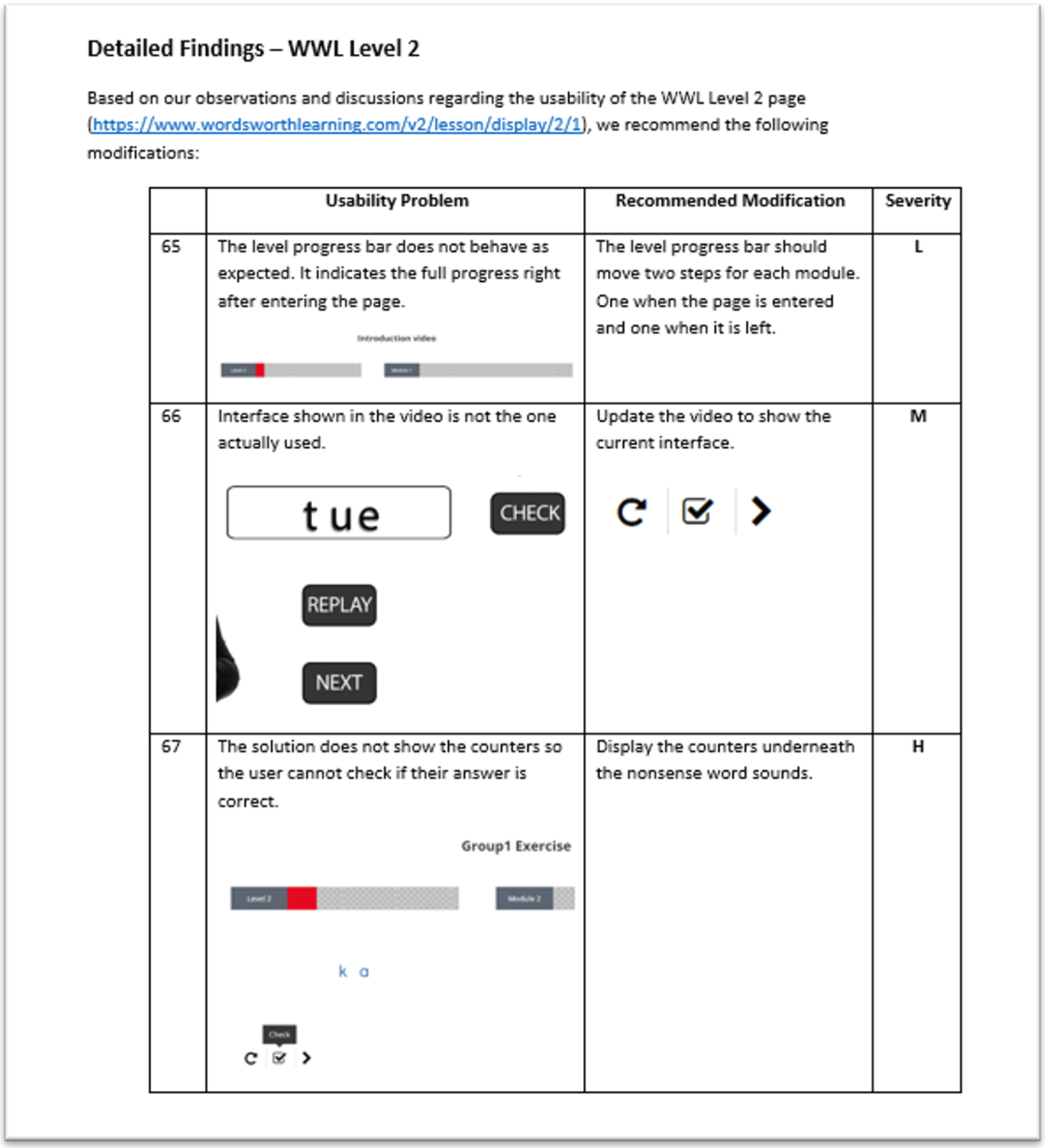
2.2.1 Heuristic Evaluation of the Website

In the early phase of the project (February 2020), Heuristic Evaluation of the current WWL application was performed by a team of HCI specialists in WP4. This activity served two

[3] https://www.nngroup.com/articles/ten-usability-heuristics/

purposes. It allowed the HCI specialists to get familiar with the content and functionality of the program to be transformed into an AR app in the course of the ARETE project. This was useful as background when providing input and feedback during the AR app design and development process. In addition, and more importantly, the HE resulted in feedback from the user-perspective, which could be used by WWL to inform the design decisions when working on the AR app. To perform the Heuristic Evaluation the team went through the WWL program level by level and step by step, assuming the role of a student (and acting as a proxy providing feedback) as well as applying usability heuristics when evaluating the interface and interaction design.

As described in the methods section above (see 2.1), the Heuristic Evaluation resulted in tables of Usability Problems, together with Recommended Modifications to address and resolve the problems identified. An example from the generated HE report can be seen in Figure 2.1, more details are presented in *Appendix A: Excerpts from the Heuristic Evaluation report for the WWL program*. After the sessions in which usability problems were identified and recommended modifications were generated, the list of usability problems was circulated around the team. Each team member then assigned severity ratings from high over medium to low. These ratings indicate how important and urgent it is for the designers and developers to address the problem. Discrepancies in severity ratings were finally discussed to reach a consensus for each usability problem. In total the Heuristic Evaluation resulted in 48 low-, 44 medium-, and 14 high-severity usability problems being discovered and recorded. Besides problems the report also presented positive observations, for example “The colourful course landing / overview page is very visually appealing, especially for younger children.” to let the designers and developers know about good aspects of the website.



Detailed Findings – WWL Level 2

Based on our observations and discussions regarding the usability of the WWL Level 2 page (https://www.wordsworthlearning.com/v2/lesson/display/2/1), we recommend the following modifications:

	Usability Problem	Recommended Modification	Severity
65	The level progress bar does not behave as expected. It indicates the full progress right after entering the page.	The level progress bar should move two steps for each module. One when the page is entered and one when it is left.	L
66	Interface shown in the video is not the one actually used.	Update the video to show the current interface.	M
67	The solution does not show the counters so the user cannot check if their answer is correct.	Display the counters underneath the nonsense word sounds.	H

Figure 2.1: Example page from the WWL program HE report.

2.2.2 Feedback on the app script

To further support the WWL app development the team of HCI specialists provided feedback on the script. This includes input and change recommendations on phrasing as well as interface and interaction design suggestions of aspects of the app reflected in the script. The text that is said by *TipTop*, the little robot guiding the learners through the program, was evaluated from the perspective of the target group based on the experience the team had, for example, from previous projects with school children. Feedback about the text here was, for example, to break down long explanations in shorter fragments, to make the content easier to understand for the target group, or suggesting more child-friendly terms, expressions, and phrases, to keep the communication playful and engaging. App and interaction design was also commented on as proxy for students as well as based on HCI

guidelines. Examples for recommendations to improve the app were having more sounds, to make the virtual environment more realistic and engaging, or having animations of TipTop explaining how to play the games or perform the exercises instead of only giving the instructions verbally. The added visuals could make the explanations more interesting and easier to follow. An overarching comment on content and design was to ensure consistency (e.g. in expressions and interactive elements), to make sure that the users could apply learned patterns throughout the whole app. Feedback was provided in several iterations for different levels of the app and thus parts of the script. Example sections of the annotated script from different iterations and levels can be seen in Figure 2.2 and Figure 2.3, more examples can be found in *Appendix B: Excerpts from the feedback on the WWL app scripts*.

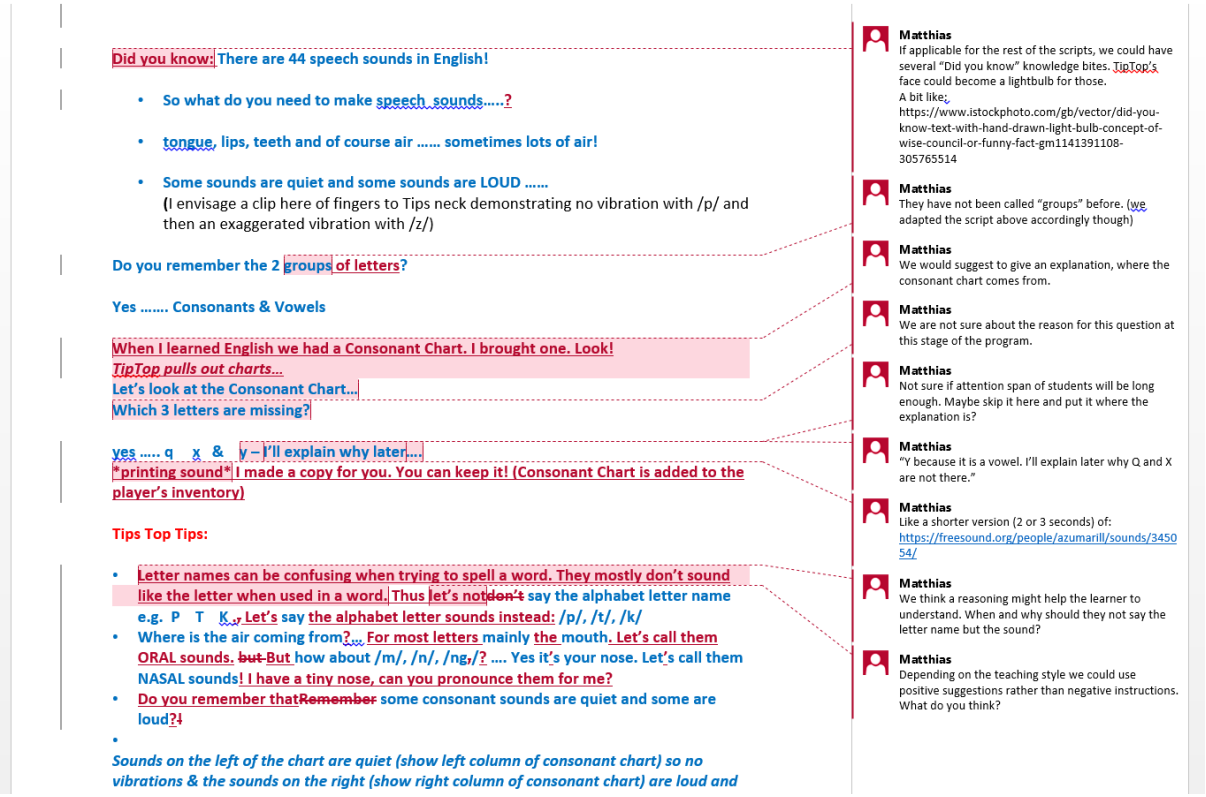


Figure 2.2: Example of the feedback and change suggestions on the first iteration of the WWL app script.

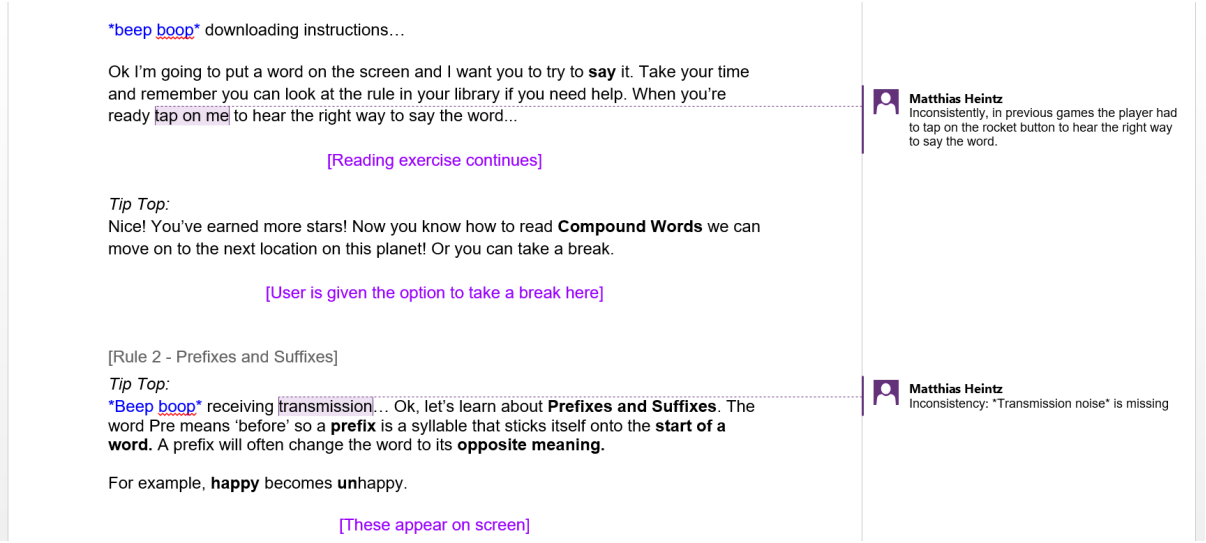


Figure 2.3: Example of feedback and change suggestions on a later iteration and level of the WWL app script.

2.3 Pilot 2: Interaction Design of CleverBooks (CLB)

Two major evaluation tasks on the digital artefacts designed for Pilot 2 have been conducted, namely Heuristic Evaluation of the app (Section 2.3.1) and Feedback on the screen and interaction design alternatives (Section 2.3.2). The process and sample outcomes are reported in the following.

2.3.1 Heuristic Evaluation of the app

The team of WP4 HCI specialists conducted Heuristic Evaluation of the CleverBooks Geometry app in the early phase of the project (February 2020). Assuming the role of students working with the app, the team tested the usability and user experience by interacting with each element and functionality of the app as a user would. Similar to the Heuristic Evaluation conducted for WWL (Section 2.2.1), usability problems were identified, recommended modifications provided, and their severity rated (first individually and then discussed in the team until consensus was reached). An example for the feedback given can be seen in Figure 2.4, more details are presented in *Appendix C: Excerpts from the Heuristic Evaluation report for the CLB app*. Besides the usability problems, the team also made note of positive observations again (e.g. "It is nice that the (verbal) feedback to the test answers changes and is not always the same."), which were likewise communicated in the HE report. The results of the Heuristic Evaluation can help to improve the current version of the software, as well as inform the development of further functionality for Pilot 2.

Detailed Findings – CleverBooks Geometry app

Based on our observations and discussions regarding the usability of the CleverBooks Geometry app, we recommend the following modifications:

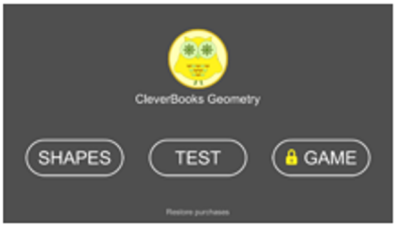
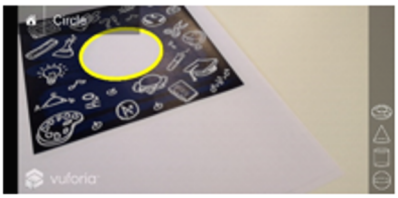
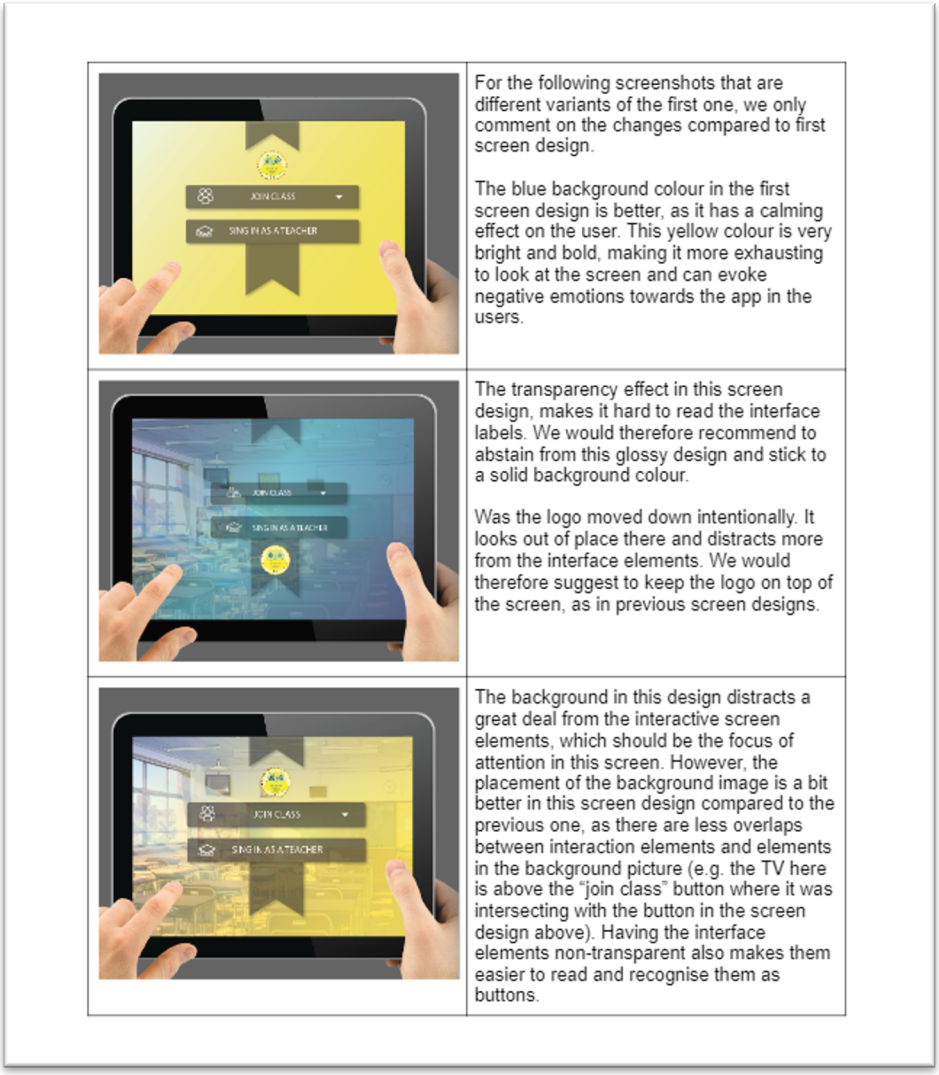
	Usability Problem	Recommended Modification	Severity
1	The "Restore purchase" button reacts to touching, but does not have any functionality. 	Remove button or add functionality.	H
2	On the iPhone there is no "SHOP" button on the home screen. 	Also remove the shop button from the Android screen, as for the user it has not the same priority as the other buttons (but the same design). The shop should be presented in a "more information about the app" area.	L
3	There is no help or instructions on how to use the app and how to get the necessary markers.	Add a "Help" section and the relevant information to the app.	H
4	The volume button overlays the vuforia logo, which makes it hard to spot. 	Move the button or the logo so that they do not overlap.	M

Figure 2.4: Example page from the CLB geometry app HE report.

2.3.2 Feedback on screen and interaction design alternatives

To further support the CLB development process, WP4 has provided HCI guidelines in regards to interface and interaction design for the mobile app. Once several alternative design ideas had been created we also reviewed, evaluated, and provided feedback on the interface design options from an HCI and user perspective. Specifically, graphical feedback in regards to redesign suggestions was provided. This kind of feedback consisting of drawings supported by textual comments was provided using the Participatory Design online tool PDot Capturer (Heintz & Law, 2018). An example for this kind of feedback can be seen in Figure 2.5.

Figure 2.5: PDotCapturer feedback on a CleverBooks app screen design.



	For the following screenshots that are different variants of the first one, we only comment on the changes compared to first screen design. The blue background colour in the first screen design is better, as it has a calming effect on the user. This yellow colour is very bright and bold, making it more exhausting to look at the screen and can evoke negative emotions towards the app in the users.
	The transparency effect in this screen design, makes it hard to read the interface labels. We would therefore recommend to abstain from this glossy design and stick to a solid background colour. Was the logo moved down intentionally. It looks out of place there and distracts more from the interface elements. We would therefore suggest to keep the logo on top of the screen, as in previous screen designs.
	The background in this design distracts a great deal from the interactive screen elements, which should be the focus of attention in this screen. However, the placement of the background image is a bit better in this screen design compared to the previous one, as there are less overlaps between interaction elements and elements in the background picture (e.g. the TV here is above the "join class" button where it was intersecting with the button in the screen design above). Having the interface elements non-transparent also makes them easier to read and recognise them as buttons.

Figure 2.6: Example page from the textual feedback on CLB app designs.

As a number of different interface design alternatives were created by CLB, for example with different colour schemes or interface element arrangements, in addition to feedback in regards to the layout of interface elements, textual feedback regarding the advantages and disadvantages of different alternatives or explaining redesign suggestions was provided. An example page from the report to CleverBooks with this type of feedback can be seen in Figure

2.6, more details are presented in *Appendix D: Excerpt of feedback on app design and alternatives by CLB*.

2.4 Pilot 3: Interaction Design of PBIS Prototypes

Pilot 3 will be launched towards the end of Year 2 of the project. Nonetheless, following the basic principles of the Human-centred Design approaches (Section 2.1), it is of paramount importance for the HCI specialists in WP4 to provide input from the user perspective in the early phase of development, which, at the time of writing this deliverable, has primarily focused on the main character and interaction design for the bespoke AR-based PBIS app.

2.4.1 Character Design

Character design refers to the creation of the alien character, who will be guiding the learners by presenting expected and unexpected behaviour in the animations to be developed. When it was designed, feedback from the HCI perspective was provided. Additionally, WP4 presented a strong stance on the involvement of end-users, particularly students, in the decision making process. This was introduced in the form of questionnaires, asking children for their preferences in regards to different design choices (see Figure 2.7), which were developed and analysed by the partners responsible for the character design.

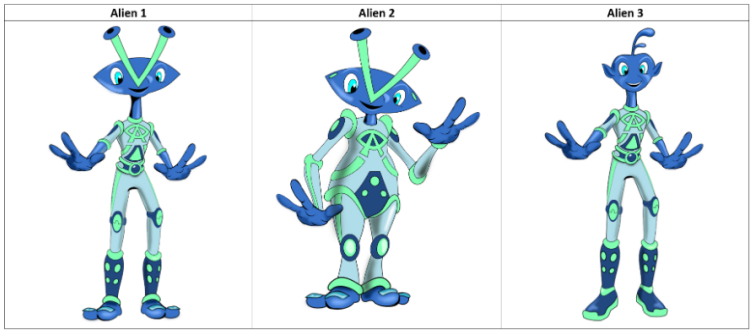


Figure 2.7: Character design choices (Picture taken from the online questionnaire for children).

According to the results of the poll submitted to children, as the target group for selecting the character, the majority vote went for Alien character 1: the skinny alien with the inverted ARETE logo head. The alien has been characterized following some additional suggestions collected from children:

1. The feet with the two toes should be replaced with the boots as worn by alien 3.
2. The face was viewed as to pointy on the sides and participants asked if the face on the sides could be rounded off a little bit more to make it look more natural.
3. The antennas on the head were liked, but suggestions were made to put them a little higher (away from the nose, to differentiate between nose and antennas).
4. Some children found the character too skinny, but as the pudgier character was liked least, so the skinny character has been considered.

2.4.2 Interaction Design

For the user interface design of the PBIS app, feedback from the user perspective was provided mostly based on usability heuristics (e.g., “Recognition rather than Recall” “Consistency and standards”, Nielsen 1994) as well as general User Experience (UX) guidelines (e.g., minimising user mental load through minimalistic design; enhancing user engagement with attractive UI design and meaningful as well as timely feedback). This was due to the issue

of not having access to end-users, especially students, because of the COVID19 pandemic and its impact on schools. When providing feedback the role of students was assumed, and experience from previous projects with students was taken into consideration.

WP4 partners have actively been involved in the user interface and interaction design process. They have an account for the Balsamiq Cloud instance that is used to create low-fidelity prototypes of the app. This way the interface was actively shaped by providing suggestions for individual screens or sequences. In addition to these screen and interaction design suggestions, verbal feedback and participation in the discussions in regards to interface options, choices, and design alternatives was provided. One important alteration from the HCI perspective was to break up overloaded and therefore complex screens into a series of screens. This reduces the visual complexity, increasing usability and UX. It also supports the user in interacting with the software, because the different screens guide the user through the process, whereas when all functionality is on one screen, the user has to figure out what to do in which order by themselves.

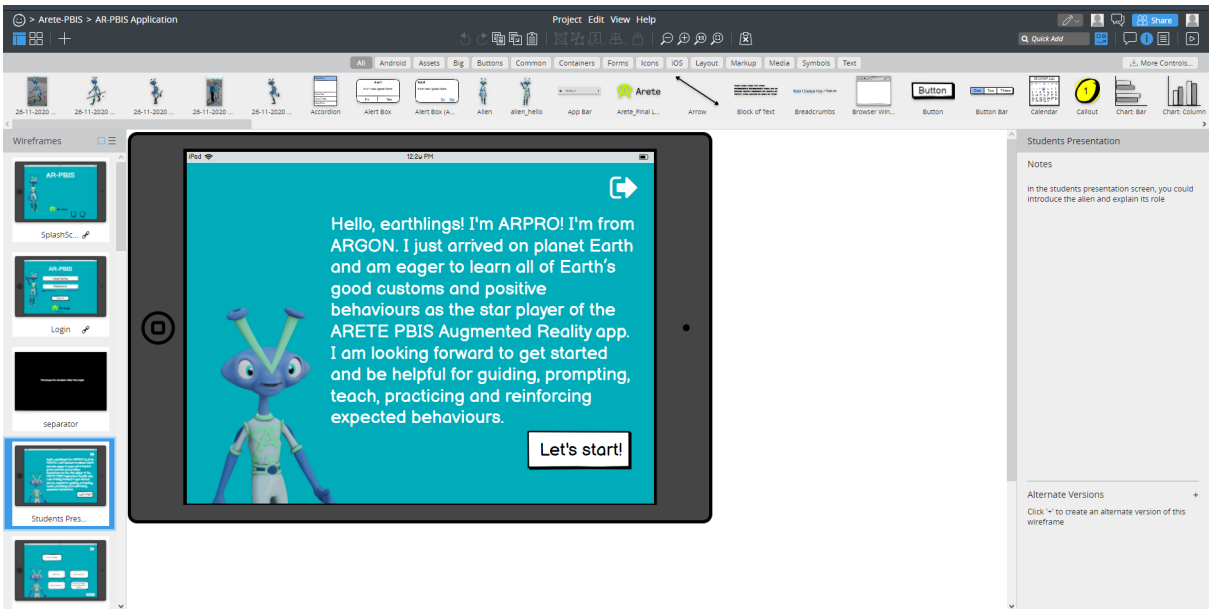


Figure 2.8: Welcome screen (designed in Balsamiq) before WP4 intervention: One big paragraph of text.

The first screen that was simplified in that matter was the “welcome screen”. In the initial proposal, the alien would say a long text, introducing themselves, providing background information, and a description of their role in the app and program (see Figure 2.8). This was realised in the mock-ups as a speech bubble with a lot of text. To ease the consumption of this information for the students, our first suggestion was to replace reading this long paragraph of text with listening to the alien saying it. However, it is not feasible to have audio in the envisioned use scenario. On the large scale, this is due to internationalisation of the app for the Pilot and for distribution all over Europe. Having audio recordings in many different languages is not possible. While this issue could be overcome, there are also pedagogical considerations that prohibit audio output: As the students will use the app in groups they cannot use headphones. Using the speaker of the mobile device will then lead to a cacophony of sound in the classroom, if several different groups work with the app in the

same room and at the same time. Therefore, a text-only solution was sought after. However, reading that much text on the screen can be tiring and exhausting, and might cause the students to skip this screen, missing out on important information. The proposed solution was to break down the long monologue of the alien into several shorter dialogues between alien and students. This way the learners will be more engaged, as they have to select answer options, and the information the alien provides will be easier to read and process, as it will be broken down into smaller portions. An example screen of the proposed solution can be seen in Figure 2.9, more details of the suggestion are presented in *Appendix E: Alien introduction as dialogues instead of monologue.*

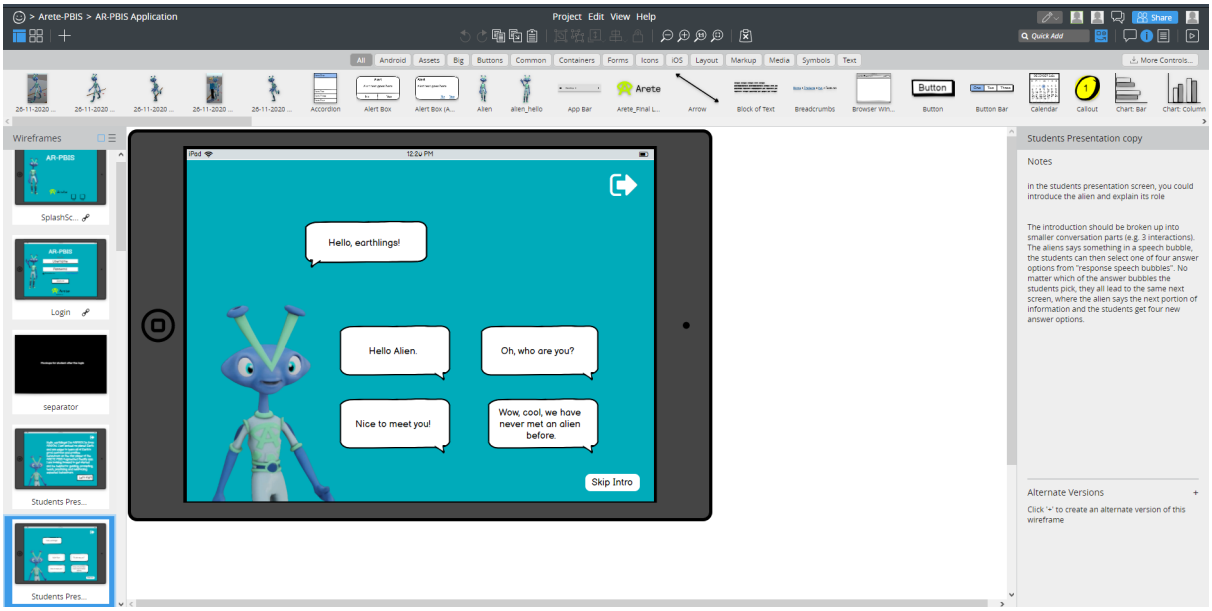


Figure 2.9: Welcome screen design in Balsamiq after WP4 intervention: The monologue in form of a big paragraph of text is broken up into a dialogue between alien and students.

Another screen that was quite overloaded in the initial design was the one to record and play back the videos of students performing the expected behaviours themselves (see Figure 2.10), which happens after the students have seen the alien performing the expected and unexpected behaviour. The proposed solution here is, to break the screen down into the different steps the students have to perform in the sequence they are presented to them, to create their videos. An example step can be seen in Figure 2.11, the entire sequence in *Appendix F: Recording sequence*. It can be seen that most of the option selection is presented in the form of a dialogue with the alien, to be consistent with the interaction design of the app and to create an active and engaging experience for the students.

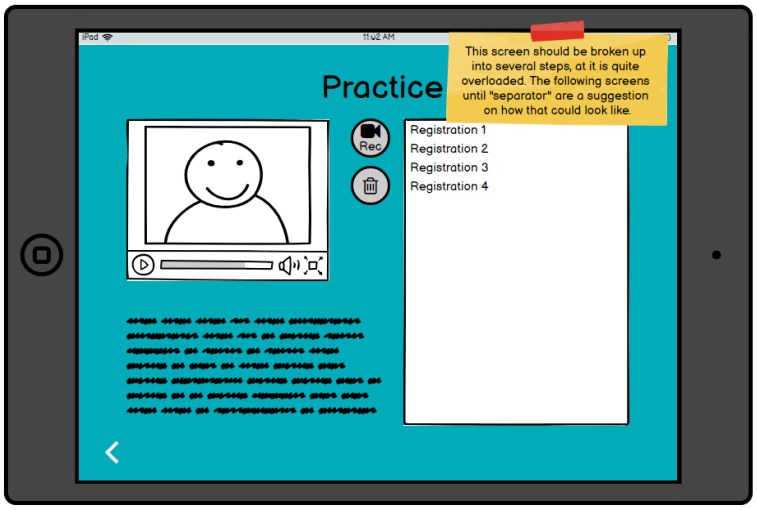


Figure 2.10: Initial design of the recording screen before WP4 intervention: High complexity.

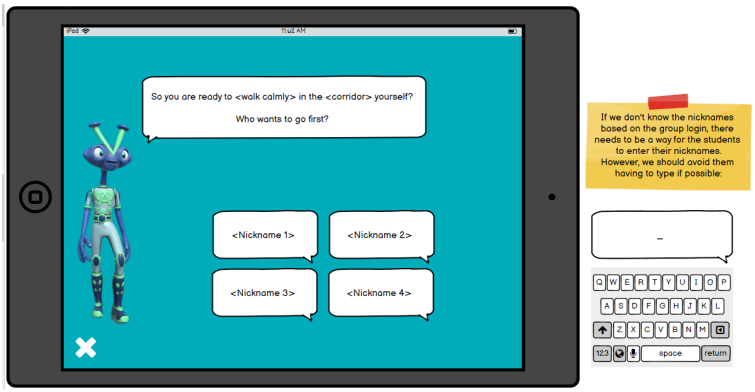


Figure 2.11: First screen and step of recording sequence after WP4 intervention: Selection of group member to be recorded.

Another example where the interaction with the app is made more engaging for the students and where the consistent interaction paradigm of "talking" to the alien is applied is the suggestion to replace standard software dialogs (see Figure 2.12) with dialogue options in a conversation with the alien (see Figure 2.13).

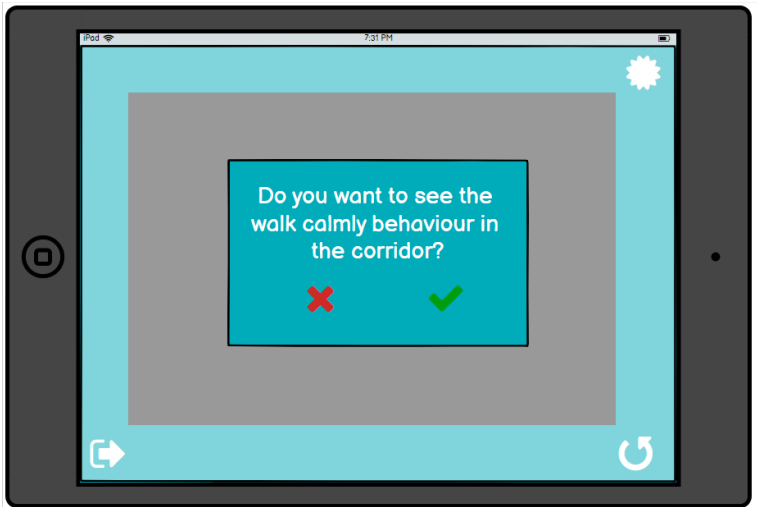


Figure 2.12: Dialog before WP4 intervention: Traditional pop-up.

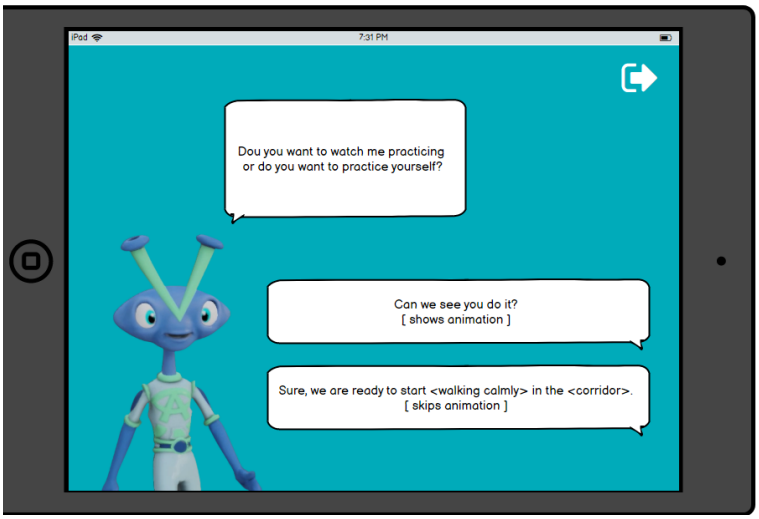


Figure 2.13: Dialog after WP4 intervention: Conversation with the alien.

To ease and streamline the selection process for the behaviour to be learned, we suggested to replace the proposed menus (see Figure 2.14) with scanning of a QR code (see Figure 2.15). The content of the QR code would then automatically tell the app, which behaviour in which setting to display. This avoids possible disturbance of the lesson by students (accidentally or to explore the other options) choosing the wrong menu entries.

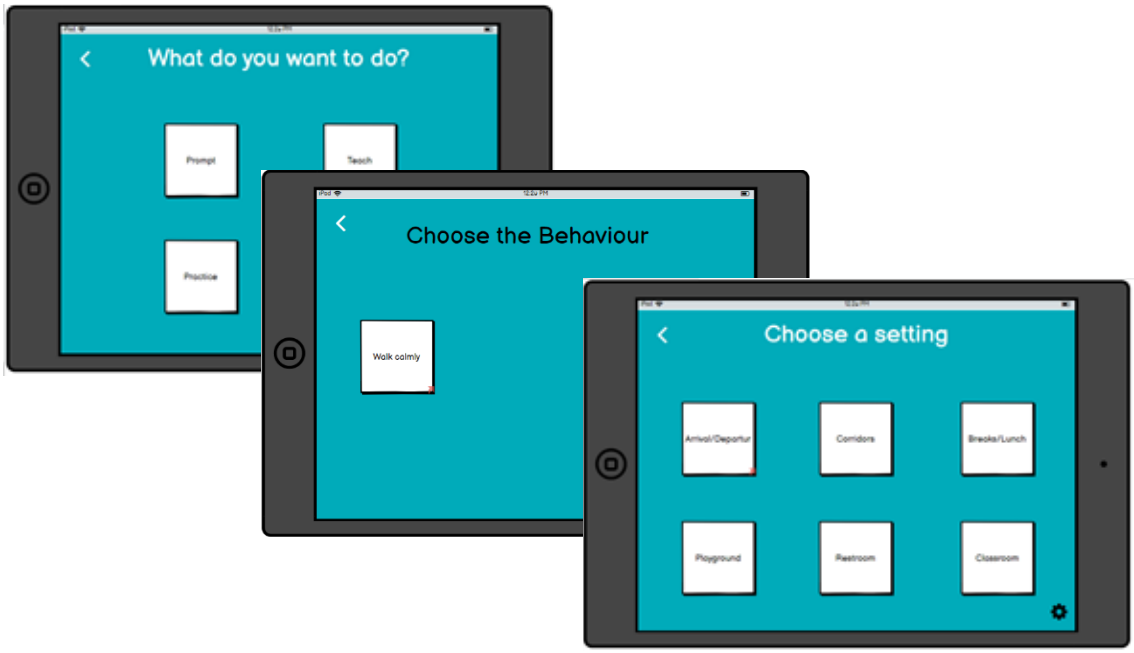


Figure 2.14: Behaviour selection before WP4 intervention: Different menus.

Figure 2.15: Behaviour selection after WP4 intervention: One screen to scan QR code.

3 Systematic Literature Review on AR Educational Tools (ARETs)

In this section, we first present an overview of the SLR carried out by the HCI team. Then we elaborate on each of the three main stages of the SLR. Next we report the major findings and then discuss the implications thus drawn.

3.1 Overview

The AR technology started to gain root in education about twenty years ago. Since then more than ten *systematic literature reviews* (SLRs) (e.g., Santos et al., 2013; Ibáñez et al., 2018; Garzon et al., 2019) on AR educational tools (ARETs) have already been published, albeit with varied quality, scope and scale. Put briefly, an SLR aims to identify relevant research studies on a specific topic, analyze and synthesize constructs of interest systematically, thereby producing a broad as well as deep understanding of that topic and drawing implications for future research and practice (Siddaway et al. 2019).

The existing SLRs on ARETs address primarily their educational impacts rather than their usability and user experience (UX), which are critical qualities for determining the acceptance and adoption of AR as teaching and learning tools. To gain deeper insights into the interaction quality of AR tools used by schools in order to inform the future research on AR in general and the work on ARETE in particular, the HCI team conducted an SLR on ARETs designed for learners from kindergartens up to secondary schools. Our SLR followed the well-recognized *Preferred Reporting Items for Systematic Reviews and Meta-analysis* (PRISMA) guidelines (Moher et al., 2009) and involved searches in four databases and existing SLRs. The process of **identification** and **screening/filtering** has resulted in a batch of 48 included papers, which have then been subjected to the process of **synthesis** to draw key insights and implications.

3.2 Stage 1: Identification

The SLR was scoped to studies of which target groups were pre-school up to secondary school learners. This scoping was based on the consideration that the design and evaluation issues related to ARETs were particularly relevant to school-aged children. The search string includes the terms “Design” and “Evaluation”, broader than “usability” and “user experience”, which are necessarily used in *title, abstract* and *keywords* to which searches are confined. The four databases used are large-scale bibliometric ones that are commonly used for SLR.

Search string:
"Augmented Reality" AND ("Education" OR "Learning") AND "School" AND ("Design" OR "Evaluation")
Databases:

- Scopus
- Web of Science (WoS) Core Collection
- ACM Digital Library Full-text Collection
- IEEE Xplore

The last set of searches was conducted on 1 July 2020, aiming to cover papers published in the first half of year 2020 while leaving sufficient time for analysis. The initial searches resulted in altogether 714 records. Each record was assigned an identifier.

3.3 Stage 2: Screening and Filtering

The course of screening and filtering results in four progressively refined scopes. As depicted in Figure 3.1, the two outer circles contain the papers fitting the scope of design and evaluation to different extents, whereas the two inner circles narrow the scope of papers to usability and UX, with the innermost one meeting the strictest eligibility criteria for synthesis.

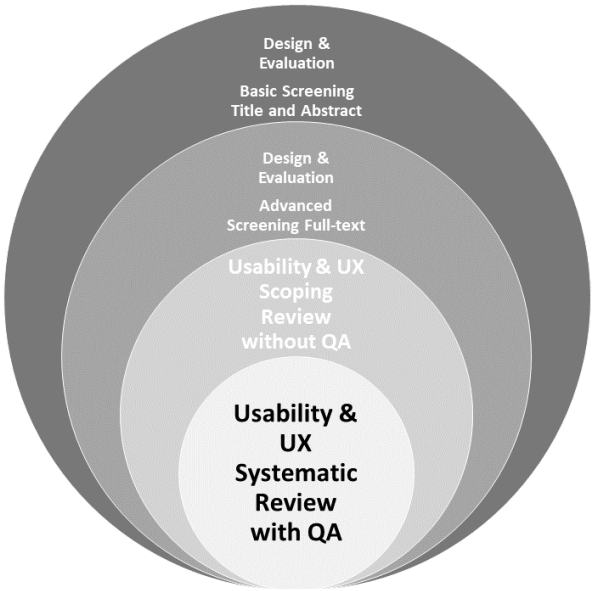


Figure 3.1. Four refined scopes of papers resulting from the progress of screening. and filtering.

As depicted in Figure 3.2, Scopus returned the highest number of records, followed by WoS. The total number of records having duplicate(s) in one or more than one of the other databases is 144. Only one instance of a duplicate record is placed in the source labelled "Overlap", but such a record can have two or more identifiers. After consolidating the duplicates, 536 unique records remain. Interestingly, ACM returned only 18 records and 11 are duplicates of Scopus and 6 of WoS.

3.3.1 Basic screening

The relevance of each of the 536 unique records was screened. Specifically, the title and abstract of each record was inspected to check for relevance by applying some of the inclusion/exclusion criteria (Table 3.1). This first screening filtered out 213 records for reasons such as the target groups were university students.

Table 3.1: Inclusion/Exclusion criteria for screening papers

Inclusion Criteria	Exclusion Criteria
in1) The design and/or evaluation of the AR application is aimed to serve an educational goal(s); in2) Target group is from pre-school up to secondary schools (pre-university); in3) Access to full-text; in4) Essential information about the AR application and methodological approaches is provided; in5) Peer reviewed	ex1) Target group is from post-secondary institutions; ex2) Theoretical or review-focused; ex3) The term 'augmented reality' mentioned while actually virtual reality is used; ex4) Written in non-English; ex5) Insufficient information is provided about the AR application or methodological approaches

3.3.2 Advanced screening

The 323 papers retained after the basic screening were further screened in full text with the use of the inclusion/exclusion criteria (Table 3.1). As a result, 98 papers were eliminated with the major reasons being inaccessible full text (in3), literature review only (ex1), VR instead of AR (ex3), and non-English (ex4). This left 225 papers for subsequent analysis. The distribution of the papers filtered in and out over the four databases and overlap is shown in Figure 3.2 (Yes vs. No). A *data extraction scheme* (Table 3.2) was developed to pull out relevant information from individual papers.

Table 3.2: The data extraction scheme

High-level Attribute	Low-level Attribute	
Paper information	identifier, author, title, publication year, source	
Basic	domain, research goals/questions, theoretical framework	
Methodological approaches	*Context*	activity, setting, hardware, software
	Participant	target group, special condition of participants, participant age range, sample size
	Data	method, data collection instrument and data type, data analysis instrument and data type
Results	challenges, perceived quality by learner, perceived quality by educator, effectiveness for learner, effectiveness for educator	
Miscellaneous	Comments	

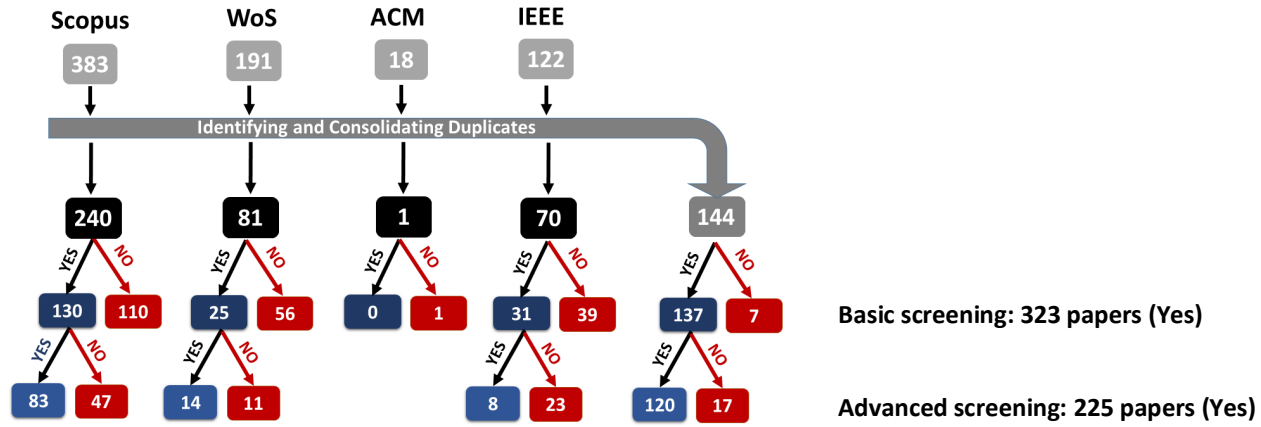


Figure 3.2. The results of basic and advanced screening stage.

3.3.3 Usability and UX without Quality Assessment – Scoping Review

Based on the attributes 'research goals/questions' and 'method' (Table 3.2) extracted from each of the 225 papers, we identified the papers that addressed usability and UX in their design and evaluation of ARETs and found that 43% (n = 97) did so. The majority (n = 128, 57%) of the research goals are pedagogical in nature: design, develop and/or evaluate ARETs for enhancing specific knowledge and ability. In accordance with the defining characteristics of scoping review that does not entail quality assessment (Arksey & O'Malley, 2005), the batch of 97 usability and UX papers fits this category (i.e. the third inner circle of Figure 3.1).

3.3.4 Usability and UX with Quality Assessment – SLR

To allow a synthesis to base on papers of a higher standard, the process of quality assessment is recommended. We employed two measures - Google Citation Index (GCI) and h-index provided by Scimago[4] Journal Rankings (SJR) - to support us to make informed decisions on including papers in the final batch for synthesis. This last filtering step led to the final batch of 47 papers. The overall workflow and results of each screening/filtering step are depicted with a PRISMA template (Figure 3.3).

[4] https://www.scimagojr.com/index.php

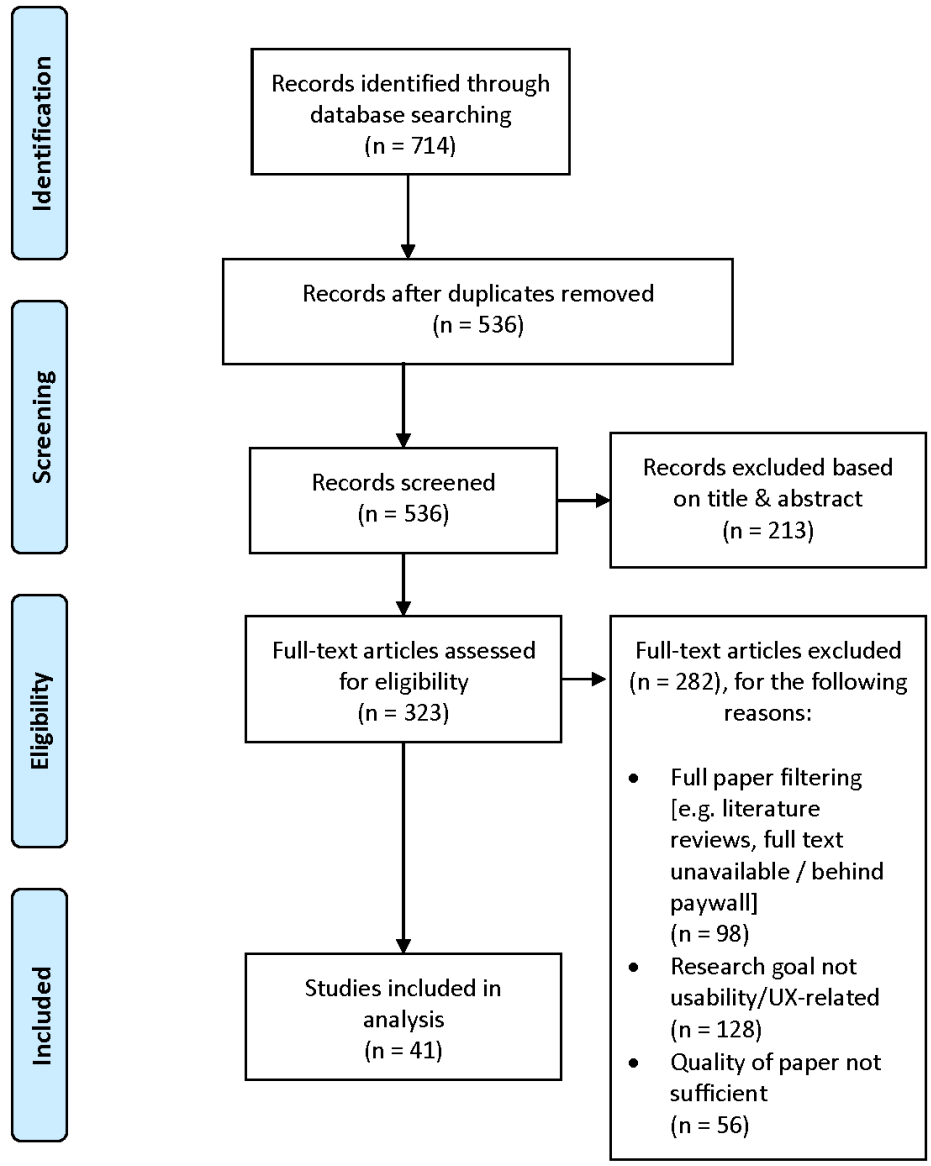


Figure 3.3. The PRISMA flowchart.

3.3.5 Extra papers from existing SLRs

Three of the existing ten SLRs on ARETs studied the usability/UX aspect as a part of the review. From the list of papers included in (Akçayır & Akçayır, 2017; Dey et al., 2018; Santos et al. 2014), we identified papers that were duplicates of our batch and also unique ones of which 7 meet our criteria. They were not captured by our searches because the search word "school" was not used in their title, abstract or keyword, although their target groups were school-age learners. With these extra papers, we have 48 (= 41 + 7) eligible for the SLR.

3.4 Stage 3: Qualitative Synthesis

In addition to the data extraction process (Table 3.2), the final batch of 48 papers were further analyzed with the following coding scheme (Table 3.3), which consists of two major

dimensions – Methods and Data, Results and Follow-up – and attributes. The information coded was synthesized to identify patterns and insights (Section 3.5).

Table 3.3: The coding scheme for usability and UX articles included in the SLR

Usability/UX: Methods and Data ▪ *Usability/UX Frameworks* ▪ *Scope* ▪ *Design Goals* ▪ *Evaluation purpose* ▪ *Research protocol:* ▪ *Informant* ▪ *Data type* ▪ *Data collection instrument* ▪ *Data analysis techniques*
Usability/UX: Results and Follow Up ▪ *Overall results* ▪ *Detailed descriptions* ▪ *Relation with Learning Effect* ▪ *Mediating variables* ▪ *Responses*

3.5 Results on General Patterns

3.5.1 Patterns of Basic Attributes

Papers by year: In searching the four databases (Section 3.1), the earliest publication year of the records returned is 2000. Figure 3.4 illustrates the changes over time in the last twenty years. While the increase was gradual in the first decade (2000-2009), it was more rapid in the second decade (2010-2019) with a visible jump from 2017 to 2018.

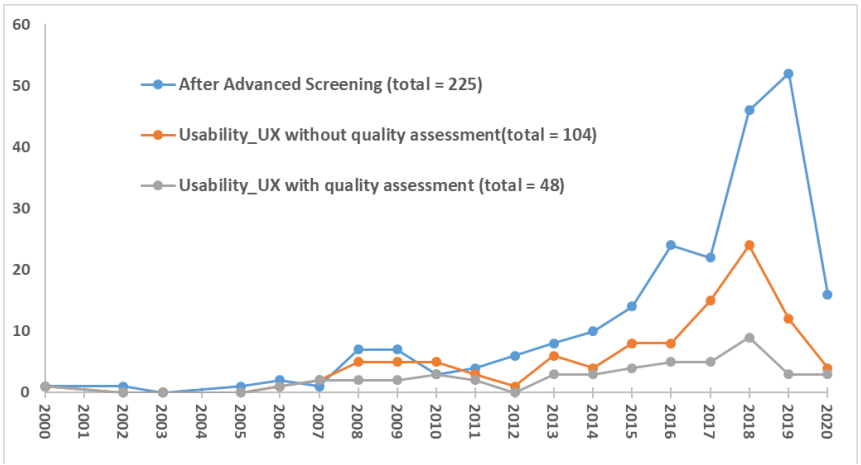


Figure 3.4: The number of papers per year in the three filtering phases.

Papers by sources: The papers were published in three types of sources: journals, conferences and books. We categorized them by seven disciplines, which inevitably overlap to some extent (Table 3.4). Out of the 48 papers, 31 are sourced from journals. Given our

focus on usability and UX of ARETs, it is not surprising that Education Tech is the most frequent category, followed by HCI. In contrast, surprisingly the number of papers from sources addressing explicitly MR/VR is limited; it can be attributed to their focus on technical development.

Table 3.4: Distribution of the papers by sources

		Design	Education Tech	Engineering & Comp. Sci.	Entertainment & Games Tech	HCI	Science & Tech	VR/MR	Subtotal
SLR (with QA)	Journal	1	*22*	1	1	5	1	0	*31*
	Conference	0	4	3	2	5	2	1	*17*
	Subtotal	*1*	*26*	*4*	*3*	*10*	*3*	*1*	**48**

Papers by application domain: The range of application domain of ARETs as described in the papers is broad (Table 3.5). We categorized them at the subject level and then clustered them to three major domains of which STEM, subsuming seven subjects, is the largest with 58% (28 out of 48 papers). The subject "Integrated Science" is referred to general science education for primary school level when the division of biology, chemistry and physics is not yet in place. Maths, mostly geometry, proved a popular subject, given the power of AR for 3D visualization. Language learning is another popular subject where AR is typically used to visualize learning scenarios, enhancing the motivation. The subject "Common Knowledge" is referred to the integrated study at the primary/lower secondary level, exploring basic scientific, social and civic topics. The subject "Cognitive and social skills" covers topics like creativity, computational thinking, memory management, emotional intelligence and symbolic play.

Table 3.5: Distribution of papers by application domain

	STEM							Humanities		General Knowledge & Skills			
	Biology	Chemistry	Physics	Integrated Science	ICT	Environ-mental Science	Maths	Lanaguage	Cultural Studies & History	Common knowledge	Cognitive & Social Skills	P.E.	Art & Design
SLR (with QA)	*5*	3	5	6	1	2	6	6	3	6	2	2	1

3.5.2 Patterns of Contextual Attributes

Pattern in Hardware: Different types of hardware were deployed in the ARETs as described in the papers reviewed (Table 3.6). By 'Mobile devices', we refer to phones and tablets. For the category of 'Custom made', it refers to the technical setup where the researchers integrated different hardware components, such as displays, cameras, projectors, headsets and scanners, in specific ways to address their research questions. Salient examples of the categories 'HMD' (head-mounted display), 'Tracker', and 'Large screen' are Hololens, Kinect and smart TV, respectively. An intriguing observation is that the number of marker-based AR applications has been consistently higher than their marker-less counterparts (Figure 3.5). One plausible reason is the reliance on GPS to support outdoor marker-less AR experience, but it is hard to ensure the stability and precision (high resolution) of GPS. Another marker-less setup is mid-air gesture-based interaction such as Kinect, but the need of equipment might hamper its adoption. In contrast, markers are easy and economical to produce, for example, with the support of a tool such as Vuforia, and everyday objects can be used as markers (e.g. P179), thereby fostering natural interaction and immersive experience.

Pattern in Software: We identified six categories (Table 3.7), however, almost half of the papers did not provide any information on the software used to create their applications. Many of the studies deployed multiple software tools; among them, Vuforia and Unity are common and often used together. Examples of '3D modelling software' are Blender, Google SketchUp, and 3DS MAX. The category 'Frameworks/Toolkits/Libraries' includes tools for low-level programming support, such as Android SDK, ARCore SDK, Open Inventor toolkit, OpenGL, NyArToolkit, etc. Examples of 'Existing AR software (customized)' mentioned are Studierstube, ARIS editor and app, Aurasma, etc. For 'Asset editing software', examples are Windows Movie Maker, Adobe Photoshop, Audacity, etc. Overall, there seems no discernible trend in the software tools deployed.

Table 3.6: Distribution of hardware used in the papers included in the SLR.

	Mobile Device	Computer /Laptop	Webcam	HMD	Tracker	Large screen	Custom made	Not specified
SLR with QA	27	14	14	8	4	3	2	1

Table 3.7. Distribution of software tools used in the papers included in the SLR.

	Vuforia	Unity	3D Modelling Software	Frameworks/Toolkits /Libraries	Existing AR Software (customized)	Asset editing software	Not specified
SLR with QA	8	7	4	14	8	1	21

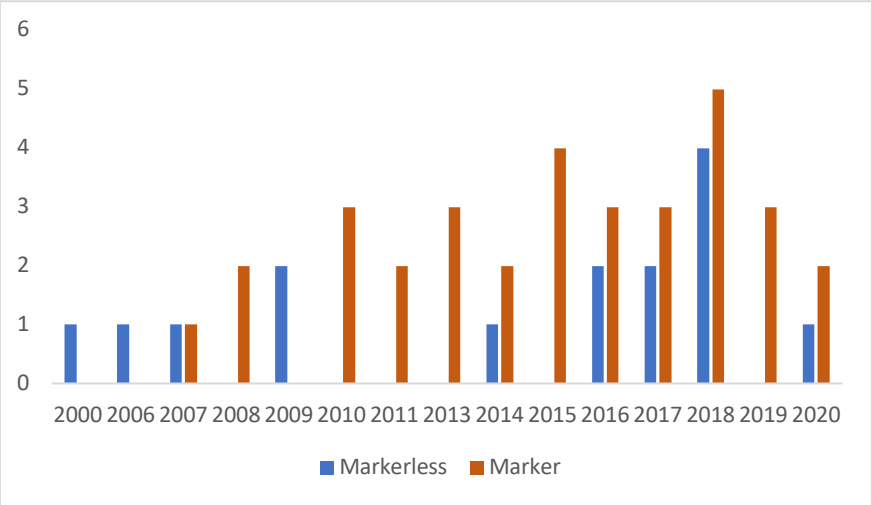


Figure 3.5: The number of ARETs with(out) markers per year.

Pattern in Participant Age and Sample Size: We applied the International Standard Classification for Education (ISCED) 2011 scheme, which defines different levels without specifying associated age ranges. With reference to different educational systems, we identified the respective ranges of the ISCED levels, as shown in Table 3.8, which clearly indicates that the majority of ARETs reviewed were for Level 1 (Primary Education). Furthermore, the sample size of the empirical work tended to be moderate with 16 studies in the SLR having 30 to 49 participants. There were a handful of studies involving more than 70 participants.

Table 3.8: Distribution of age groups included in the SLR and Scoping Review

ISCED:	Level 0	Level 1	Level 2	Level 3		
Education:	Early Childhood	Primary	Lower Secondary	Upper Secondary	Mixed	Unspecified
Age range:	4-5	6-12	13-16	17-19		
SLR with QA	0	28	10	4	6	0

Pattern in Participant Condition: The existing SLRs reported that very few research studies of ARETs focused on target groups with special needs. We corroborate this observation with the batch of papers we reviewed. Among the 48 papers for the SLR, only one targeted students with physical disabilities to learn science (P019) and one on autism (P256). In other words, only 4% of these ARET research studies addressed students with special needs. This is the issue worthy to investigate which factors contribute to the low rate of application.

Pattern in Settings: We categorized the settings where the ARETs were deployed into four major groups: in classroom (n = 33), outdoors (n = 13), museum (n = 5), and at home (n = 1). A handful of studies involved more than one setting (e.g. in classroom & museum, [P019]; in classroom, museum, and at home, [P128]). Most of the studies took place in the classroom where the control of the learning activities and infrastructure (e.g. mobile devices, the internet connectivity) tended to be more manageable than outside the classroom. The studies taken place outdoors, including playgrounds within a school premise and field trips, faced different challenges such as low GPS accuracy [P018], poor visibility [P167], and bad detection of nature objects used as markers [P179].

Pattern in Types: While there are different ways to typify the ARETs reviewed, we focused on two dimensions that we deem more relevant: Game-based vs. Non-Game-based and Individual vs. Collaborative. Results are shown in Table 3.9. Many of the studies reviewed were of collaborative and non-game-based type (n = 16). These observations could be related to some factors: social learning theories embraced by the related projects, limited availability of the devices where the ARET was run to all students at one time, and higher costs for developing game-based contents. Note, however, none of the 48 studies included both individual and collaborative modes, making it difficult to draw solid conclusions on the relative strengths (or drawbacks) of either of the learning modes.

Table 3.9: Distribution of types defined by game-basedness and collaboration

	Individual	Collaborative	Unspecified	*Total*
Game-based	4	10	2	*16*
Non-game-based	6	16	6	*28*
Both	2	1	1	*4*
Total	*12*	*27*	*9*	

Patterns of Learning Outcomes: While our SLR focused on the usability and UX of the ARETs, it is relevant to get an overview of the learning outcomes. As the target groups of these research studies were students, we synthesized the learning effects of the related ARETs. In 13 studies students were reported to benefit from improved learning experience (e.g.

enhanced interestingness of the topics), 13 studies from supporting understanding of the related subject matters, and 14 from increased motivation.

Concerning the learning effect, 13 out of 48 papers reported that the experimental group performed as good as or better than the control group on specific topics, including writing [P182], animal classification [P011], electromagnetism [P074] and pedestrian navigation [P175] whereas eight papers reported that the ARETs helped improve specific knowledge or skill, including collaborative skills [P330], recognizing emotions in avatar (Lorusso et al., 2016) and symbolic play (Kotzageorgiou et al., 2018).

3.6 Results on Usability and UX

This section reports the synthesis results focusing on the usability and UX aspects of the ARETs deployed in the 48 papers reviewed.

3.6.1 Scope, Goals and Methods

For 37 of the 48 studies, the usability/UX work was for evaluation only whereas the other 11 studies aimed to address both design and evaluation goals. The core concepts such as ease of use, satisfaction, efficiency, fun, flow, and engagement were reported to underline the design and evaluation of the ARETs. Nonetheless, some of the concepts such as satisfaction were not explicitly defined or operationalized. Furthermore, 36 of the usability/UX evaluation studies were summative, 7 were formative, and 5 were both.

The variety of usability/UX methods employed in the 48 studies was small with questionnaire being the predominant one used in 34 studies, followed by interviews (n = 18), observation (n = 12), and focus group (n = 3). This pattern corroborates the findings of the previous SLRs (e.g. Santos et al., 2013). Furthermore, three studies reported analysing interaction behaviours by using video recordings of participants when they were implementing the task scenarios with the ARET. Slightly more than half of the studies (n = 26) employed more than one method (e.g. combining questionnaire, interview and observation) whereas fifteen, three, and one studies relied only on questionnaire, observation or interview, respectively. Only a few studies attempted non-typical methods: two studies [P231, P232] asked children participants to draw their interaction experiences with the ARETs and be interviewed to explain the drawings; in one study (P019) researchers deployed objective physiological measures (heart rate, eye strain) and subjective questionnaires (i.e. Comfort Rating Scale) to yield quantitative data for their formative as well as summative evaluation. 13 studies collected only quantitative data whereas 14 studies collected only qualitative data and 21 studies mixed data.

The range of data collection instruments was small. Out of 48 studies, 23 used homegrown questionnaires, which were either created from scratch by the authors (e.g., [P128, P134]) or taken from a combined set of existing questionnaires. Only eight of these homegrown questionnaires were reliability tested with Cronbach's alpha. Furthermore, 13 studies employed standardized usability/UX questionnaires, such as SUS, NASA-TLX and User Engagement Scale (UES), and only four reported Cronbach's alpha. Other methods like interview, observation and focus group were conducted in a loose manner without using standardized questions or templates.

3.6.2 *Perceived Usability/UX and Usability Problems (UPs)*

Participants, including learners and educators, in 31 out of the 48 studies, were reported to have positive perceptions of the usability and UX of the ARETs concerned. Comments on high usability, such as easy to use, easy to scale the AR model, easy to navigate, low cognitive load, and high level of satisfaction, were documented (e.g. [P167, P001, Ex007]). Positive emotional responses, such as fun, engaging, and playful, were often reported (e.g. [P107, P109]). On the contrary, two studies had negative usability/UX results and 15 studies had mixed responses within individual studies; the negative perceptions were related to different usability problems (UPs), which are summarized in Table 3.10. Note, however, half of the 48 studies did *not* report the findings of UPs, which could explain the poor usability and UX found. In contrast, many of the remaining studies reported more than one UP. The most frequently identified UPs were AR-specific, namely, the design and usage of markers. Other common AR-specific UPs were related to the perception, manipulation, control, and positioning of 3D virtual objects. Some of the issues such as slow rendering, which were reported in the papers published in the early 2000s when the work on ARETs started to take off, have been resolved due to improved algorithms and more powerful computer processors. The UP of the video sound quality could arguably be relevant, because the feature might contribute to the holistic user experience with the ARET concerned.

Table 3.10: Category of usability problems identified in the papers included in the SLR.

Category of Usability Problems	Count	Papers
Marker-related: usage, detection, control, occlusion, transfer across contexts, objects in nature	6	P074, P108, P223, P179, P230, P376
Perceptual quality of 3D virtual object: realism, visibility (outdoor), aesthetic design	5	P134, P175, P167, P220, Ex001
Precision: misplacement of virtual objects (avatar), GPS	5	P131, P018, P009, P001, P007
Small screen size	5	P013, P001, P007, P220, P230
Software stability: crashes and rebooting	5	P072, P131, P018, P175, Ex001
Virtual object manipulation and control (gestural and hand recognition)	5	P108, P109, P256, P335, P376
Dual handling of physical device and virtual object	4	P074, P167, P220, Ex005
Understandability: AR mechanism, User interface element	4	P134, P109, P256, P128
HMD: weight, motion sickness	3	P001, P007, P019
Infrastructure setup: camera position and image projection on real-life objects	2	P256, Ex005
Slow rendering	2	P001, P007
Sound quality of video	1	P220

In analyzing whether and how these UPs were addressed within the respective studies, only in three studies [P109, P256, P376] did the authors report that the related UPs were handled with success. Specifically, in [P109], the changes included installing UI buttons on both sides of a tablet to facilitate controlling the ARET; providing a tutorial on 3D depth, amplifying perceptual cues (e.g., adding shadowing), and rendering visuals simpler. In case of [P256], simplifying the AR game mechanics and adding meaningful images to indicate the start

position of the game resulted in an improved understanding of interaction design. For [P376], one marker and a menu supporting switches between solids to be visually augmented were deployed to replace multiple markers, and a pinch-to-zoom feature was also added.

Furthermore, in eleven studies, the authors presented some planned improvement actions as future work, albeit with different degrees of concreteness. Among them, four suggested adding a tutorial could resolve some UPs; one was more specific: “a short tutorial for introducing the device by a virtual friendly pet” [P033] whereas one simply wrote “a short tutorial”. Some proposed generic actions such as “robuster tracking” [Ex005] and “focus on simplicity” (P128) whereas some had UP-focused actions. For instance, in [P134], the authors proposed using a road map instead of a satellite map to address the issue of poor map tile quality. In [P179], to address the problem of marker recognition, the authors suggested using computer vision and machine learning to identify nature objects rather than transforming an object in nature into a marker. Nevertheless, the remaining studies acknowledged the presence of UPs without specifying any remedial actions.

3.6.3 Relations between Usability/UX and Learning Effect

By learning effect, we refer to the measure showing the change in specific knowledge, skill or ability as a result of learning with the ARET concerned. Nineteen out of the 48 studies did not attempt to take such a measure; the relative high percentage of such studies can be attributed to our paper selection criteria (Section 3.1). A number of these studies focused on developing the application right from the interaction design perspective. Nonetheless, 15 of the studies did not relate the learning effect to the usability/UX findings, either quantitatively or qualitatively. In other words, whether learners gained knowledge, skill or ability from an ARET seems independent of their perceptions and responses from interacting with it. However, it could be that the authors just did not discuss the relation explicitly. Nine and five studies showed the positive and negative relations between the usability/UX findings and learning effect, respectively. The mediating variables mentioned for the **positive relations** were **novelty of the tool, motivation, flow, presence, and instant feedback**, whereas those for the **negative ones** were **task difficulty, lack of engagement, and difficulty in marker manipulation**. These issues will be followed up in D4.4 where findings of each Pilot are reported.

3.7 SLR Insights and Implications for ARETE Use Scenarios

In this section, we present the main results of the SLR along three aspects. For each aspect, we analyse the insights gained and their implications for ARETE use scenarios. For individual implication, we discuss how it can be (or has already been) applied within the project and beyond it.

3.7.1 Target groups, Subjects, Settings and Types

SLR Insights: The trends of target groups, learning subjects and settings suggest that there are significant gaps to be bridged. First, it is necessary to provide parents with enough training to support their children to deploy ARETs at home, given the proven benefits of such educational technologies. This is particularly salient in the wake of the current pandemic when home-schooling has become essential. Irrespective of the recurrence of such a crisis, which hopefully will never happen again, children’s self-directed learning in formal as well as

informal settings should be fostered with scaffolding to be provided by informed teachers, parents and carers. Arguing along this line, the range of age groups and the scope of learning subjects should also be expanded to ensure a broad coverage. Clearly, developing AR-based content entails knowledge and skillsets different from those for traditional learning materials. Usable and useful authoring tools that can facilitate teachers and parents to co-create contents with children can be viable options to address the observed gaps. Second, the use of ARETs for learners with special needs should further be explored. Although the number of studies (only 2) was small, they all suggested the potential of ARET in this regard, especially the game-based approach (cf. [P256]).

Implications for ARETE use scenarios:

Implications	**Practices**
Involve end-users, including students, teachers and parents, in co-designing the ARETE prototypes to ensure their acceptance and adoption.	This has been our planned approach all along. We have explored a range of options to gain access to end-users, and addressed the access issue with user proxy and analytic methods.
Offer the training on the use of AR educational tools to parents, who can then be encouraged to support their children to learn with such tools at home with confidence.	Parents are not originally included as end-users. It will incur extra effort to address this emerging need. Nonetheless, the consortium can discuss its possible realization within or beyond ARETE. For instance, the ARETE toolkit being developed can take this target group into consideration. The training material to be created for teachers could also be made available to parents.
Forge stronger collaboration between schools and informal learning settings such as museums to develop AR-based learning contents for blended learning.	Museums are not originally included as a target setting. But the UCD Science Festival in every June with thousands of student participants can have similar impact. Nonetheless, If resources permit, a small scale exploratory study in museums (science, history) using the ARETE toolkit can be conducted to identify what adaptations are required.
Widen the range of domain areas to which ARETs can be applied.	Pilot 1 and Pilot 2 address English literacy and STEM – the basic areas whereas Pilot 3 targets a new domain – PBIS, which can build reference frameworks for other novel and critical domains.
Explore the opportunity of applying the AR technology to support children with special needs, especially using the multisensory approach.	A satellite research project can be conducted to address strengths and limitations of applying the ARETE toolkit to children with special needs (e.g. Down Syndrome).
Study the effect of integrating game-based learning (GBL) techniques into ARETs to further enhance their motivational potential and impact	In all three ARETE Pilots, GBL approaches have been or can be applied to a varied extent. The WWL app includes literacy games that can be played to practice learning to read and spell. The CLB app already includes a quiz which allows students to playfully test their knowledge. A game with multi-user interaction is planned to be developed for ARETE. The PBIS app is still in

	its early planning and development stage, thus it can be discussed how game-based learning could be included.
Substantiate the effect of the collaborative learning mode for ARETs with more empirical evidence, especially here is a lack of comparison study on individual vs. collaborative mode given organizational and methodological challenges.	The WP3 development toolkit for collaborative learning will enable more group-based use scenarios, such as PBIS in WP5. WP4, together with WP6, will explore the possibility of conducting comparison scenarios and viable means to collect and analyse data systematically to provide empirical evidence.

3.7.2 Hardware and software

SLR Insights: Despite the advantages of mobile devices for ARETs, the potential of HMDs can be explored. Nonetheless, their affordability is a significant barrier. Clearly, high-quality tablets and phones can probably lead to good usability and positive user experience, but they are also more expensive. This can especially become an issue in school settings where several devices, not just one, need to be acquired to allow individuals or small groups of students to experience an ARET. This budgetary concern may be eased by some joint private-public partnership. Furthermore, the design and development of markers entail further research efforts to address the usability problems identified, especially the issue of lighting and dealing with low-quality camera.

Implications for ARETE use scenarios:

Implication	**Practice**
Study systematically the usability and UX of marker-based ARETs to identify recommendations for resolving usability problems with markers.	In the three ARETE Pilots, both marker-based and markerless technology are used. Pilot 1 is developing markerless AR, as they already have experience with a marker based approach for their app from a previous project (AHA). The basic idea of markerless augmented reality, is to superimpose graphics, audio and other sensory enhancements on a real-world environment, in real time, on a mobile device. This means less restriction, compared to 'marker' based approach, as schoolchildren can use their app anytime, anyplace, without the need for physical markers. In addition it is anticipated that a number of students in the Pilot 1 cohort will also present with co-morbid diagnoses associated with reading / spelling disorders and dyslexia. Conditions such as ADHD, Developmental Co-ordination Disorder (DCD) or visual processing difficulties can adversely affect their ability to access the AR using a physical marker-based approach, so an easier, less cumbersome path is preferred.

	Pilot 2 is going to use the well-established markers developed and already deployed successfully by CLB. For Pilot 3 the idea is to apply a mixed solution that uses marker-based and marker-less solutions. Pilot 3 is also planning to expand the existing research in the field of marker-less AR to include new and innovative interaction with the AR objects in the PBIS app. Formative and summative evaluation activities will pay particular attention to issues that can arise from the use or non-use of markers.
Explore cost-effective use scenarios with the marker-less AR technology (e.g. Hololens) to compare with strengths and limitations of marker-based AR.	Empirical studies, albeit small-scale, will be coordinated by the HCI team to compare the learning effect of the marker-based and markerless AR educational apps and to examine the issue of evaluating users' real-time emotional responses when interacting with the apps.

3.7.3 Usability/UX and Learning Effect

SLR Insights: It is surprising to note the relatively low number of studies attempting to measure the learning effect in empirical research on educational technology. It can be a methodological artefact of the SLR process as we included papers focusing on usability/UX. But it can also be attributed to the fact that the research on ARETs is emerging; many studies were still at the exploratory phase. Nevertheless, we deem it recommendable to encourage authors/researchers to assess systematically the learning effect, usability/UX qualities, and their relations.

Furthermore, it would be beneficial to identify usability problems of the ARET prototypes and learn from design issues to inform better design and development. The prevailing sole reliance on the use of questionnaire may not serve this purpose. Hence, the ARET researchers should be enabled to conduct comprehensive usability evaluations of prototypes to gain insights.

Implications for ARETE use scenarios

Implication	**Practice**
Evaluate the ARETE prototypes from both the pedagogical and usability/UX aspects to identify their relations, gaining better insights into factors influencing the impact of the intervention.	With the close collaboration between the HCI team and the pedagogical experts (Univ. Würzburg), the evaluation approaches of the ARETE Pilots will address this issue.
Employ multiple methods to evaluate the design of the ARETE prototypes, including both lab-based usability testing and field studies (classroom-based) to collect quantitative and qualitative feedback from teachers and students.	While questionnaire remains an important evaluation method to be used in the ARETE Pilots, other established methods including focus groups, individual interviews and observations are also applied. Alternative methods are being used in conjunction with

	established ones, including software-based evaluation tools such as PDot (Section 2.3) and psycho-physiological tools measuring emotional responses to interaction with the AR applications.

4 Conclusion

The first year of WP4, like some other ARETE WPs, has been severely affected by the pandemic, because it has rendered access to end-users extremely difficult, if not impossible. Nonetheless, the HCI team, in close collaboration with the partners, have mitigated the adversity with some alternative approaches, including prototype evaluations with proxy users and usability heuristics. While the feedback and improvement suggestions could have been more rigorous if end-users had been involved, they were constructive in terms of enhancing the quality of interaction design, as acknowledged by the beneficiary partners – WWL and CLB. Examples include the increased understandability and attractiveness of UI objects (like the CLB app screens, for which improvement suggestions and design-decision support between alternatives were provided) and dialogues (like the increased child-friendliness of the script for TipTop, the robot in the WWL app).

WP4 conducted the empirical evaluations of the ARETE prototypes to support the realization of the planned ARETE use scenarios, which can be substantiated by the insights gained from the SLR. Furthermore, the SLR can also broaden the scope of use scenarios, which may be realizable within the project's lifetime. All in all, as the viability of WP4's work hinges crucially on the access to end-users, we hope that a wide use of vaccines will overcome the harsh challenges we faced last year.

References

Part I: All 48 papers included in the SLR sorted by ID number:

Ex001 Di Serio, Á., Ibáñez, M. B., & Kloos, C. D. (2013). Impact of an augmented reality system on students' motivation for a visual art course. *Computers & Education, 68*, 586-596.

Ex002 Sin, A. K., & Zaman, H. B. (2010). Live Solar System (LSS): Evaluation of an Augmented Reality book-based educational tool. *2010 International Symposium on Information Technology (Vol. 1, pp.1-6).*

Ex003 Juan, C. M., Llop, E., Abad, F., & Lluch, J. (2010). Learning words using augmented reality. *10th IEEE International Conference on Advanced Learning Technologies,* pp. 422-426.

Ex004 Juan, C. M., Toffetti, G., Abad, F., & Cano, J. (2010). Tangible cubes used as the user interface in an augmented reality game for edutainment. *10th IEEE International Conference on Advanced Learning Technologies,* pp. 599-603.

Ex005 Liu, W., Cheok, A. D., Mei-Ling, C. L., & Theng, Y. L. (2007). Mixed reality classroom: learning from entertainment. *Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts*, pp.65-72.

Ex006 Sayed, N. E., Zayed, H. H., & Sharawy, M. I. (2011). ARSC: Augmented reality student card an augmented reality solution for the education field. *Computers & Education, 56*(4), 1045-1061.

Ex007 Chiang, T. H., Yang, S. J., & Hwang, G. J. (2014). An augmented reality-based mobile learning system to improve students' learning achievements and motivations in natural science inquiry activities. *Journal of Educational Technology & Society, 17*(4), 352-365.

P001 Kaufmann H., Schmalstieg D., Wagner M. (2000). Construct3D: A virtual reality application for mathematics and geometry education. *Education and Information Technologies*, *5*, 263–276.

P007 Kaufmann H., Schmalstieg D. (2006). Designing immersive virtual reality for geometry education. *IEEE Virtual Reality Conference* (VR 2006) (pp. 51-58).

P009 Squire K.D., Jan M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology, 16*(1), 5-29.

P011 Freitas R., Campos P. (2008). SMART: A system of augmented reality for teaching 2nd grade students. *People and Computers XXII Culture, Creativity, Interaction 22*, 27-30.

P013 Liu T.-Y., Chu Y.-L. (2008). Handheld augmented reality supported immersive ubiquitous learning system. *2008 IEEE International Conference on Systems, Man and Cybernetics* (pp. 2454-2458).

P018 Dunleavy M., Dede C., Mitchell R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education and Technology*, *18*(1), 7-22.

P019 Arvanitis T.N., Petrou A., Knight J.F., Savas S., Sotiriou S., Gargalakos M., Gialouri E. (2009). Human factors and qualitative pedagogical evaluation of a mobile augmented reality system for science education used by learners with physical disabilities. *Personal and Ubiquitous Computing, 13*(3), 243-250.

P033 Juan M.C., Furió D., Seguí I., Rando N., Cano J. (2011). Lessons learnt from an experience with an augmented reality iPhone learning game. *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology* (pp. 1-8).

P048 Wojciechowski R., Cellary W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers & Education, 68*, 570-585.

P055 Salvador-Herranz G., Pérez-López D., Ortega M., Soto E., Alcañiz M., Contero M.(2013). Manipulating virtual objects with your hands: A case study on applying desktop Augmented Reality at the primary school. *46th Hawaii International Conference on System Sciences* (pp. 31-39).

P072 Cai S., Wang X., Chiang F.-K. (2014). A case study of Augmented Reality simulation system application in a chemistry course. *Computers in Human Behavior, 37*, 31-40.

P074 Ibáñez M.B., Di Serio Á., Villarán D., Delgado Kloos C. (2014). Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers & Education, 71*, 1-13.

P098 Lin H.-C.K., Chen M.-C., Chang C.-K. (2015). Assessing the effectiveness of learning solid geometry by using an augmented reality-assisted learning system. *Interactive Learning Environments, 23*(6), 799-810.

P107 Lu S.-J., Liu Y.-C.(2015). Integrating augmented reality technology to enhance children's learning in marine education. *Environmental Education Research, 21*(4), 525-541.

P108 Fleck S., Hachet M., Christian Bastien J.M. (2015). Marker-based Augmented Reality: Instructional-design to improve children interactions with astronomical concepts. *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 21-28).

P109 Radu I., Doherty E., DiQuollo K., McCarthy B., Tiu M. (2015). Cyberchase Shape Quest: Pushing geometry education boundaries with augmented reality. *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 430-433).

P128 Hsiao H.-S., Chang C.-S., Lin C.-Y., Wang Y.-Z. (2016). Weather observers: a manipulative augmented reality system for weather simulations at home, in the classroom, and at a museum. *Interactive Learning Environments, 24*(1), 205-223.

P131 Laine T.H., Nygren E., Dirin A., Suk H.-J. (2016). Science Spots AR: a platform for science learning games with augmented reality. *Educational Technology Research and Development, 64*(3), 507-531.

P134 Laine T.H., Suk H.J. (2016). Designing mobile augmented reality exergames. *Games and Culture, 11(*5), 548-580.

P167 Ahmed S., Nasir B., Khan J.A., Ali S., Umer M. (2017). MAPILS: Mobile augmented reality plant inquiry learning system. *IEEE Global Engineering Education Conference (EDUCON)* (pp. 1443-1449).

P174 Joo-Nagata J., Martinez Abad F., García-Bermejo Giner J., García-Peñalvo F.J. (2017). Augmented reality and pedestrian navigation through its implementation in m-learning and e-learning: Evaluation of an educational program in Chile. *Computers & Education, 111*, 1-17.

P175 Cai S., Chiang F.-K., Sun Y., Lin C., Lee J.J. (2017). Applications of augmented reality-based natural interactive learning in magnetic field instruction. *Interactive Learning Environments, 25*(6), 778-791.

P179 Alakärppä I., Jaakkola E., Väyrynen J., Häkkilä J. (2017). Using nature elements in mobile AR for education with children. *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 1-13).

P182 Wang Y.-H. (2017). Exploring the effectiveness of integrating augmented reality-based materials to support writing activities. *Computers & Education, 113*, 162-176.

P201 Georgiou Y., Kyza E.A. (2018). Investigating the coupling of narrative and locality in augmented reality educational activities: Effects on students' immersion and learning gains. *13th International Conference of the Learning Sciences* (ICLS) 2018, Volume 1. London, UK: International Society of the Learning Sciences.

P205 Layona R., Yulianto B., Tunardi Y. (2018). Web based Augmented Reality for Human Body Anatomy Learning. *Procedia Computer Science, 135*, 457-464.

P214 Leonard S.N., Fitzgerald R.N. (2018). Holographic learning: A mixed reality trial of Microsoft Hololens in an Australian secondary school. *Research in Learning Technology, 26*.

P220 Efstathiou I., Kyza E.A., Georgiou Y. (2018). An inquiry-based augmented reality mobile learning approach to fostering primary school students' historical reasoning in non-formal settings. *Interactive Learning Environments, 26*(1), 22-41.

P222 Wu P.-H., Hwang G.-J., Yang M.-L., Chen C.-H. (2018). Impacts of integrating the repertory grid into an augmented reality-based learning design on students' learning achievements, cognitive load and degree of satisfaction. *Interactive Learning Environments, 26*(2), 221-234.

P223 Alhumaidan H., Lo K.P.Y., Selby A. (2018). Co-designing with children a collaborative augmented reality book based on a primary school textbook. *International Journal of Child-Computer Interaction, 15*, 24-36.

P231 Schaper M.-M., Santos M., Malinverni L., Zerbini Berro J., Pares N. (2018). Learning about the past through situatedness, embodied exploration and digital augmentation of cultural heritage sites. *International Journal of Human-Computer Studies, 114*, 36-50.

P232 Malinverni L., Valero C., Schaper M.-M., Pares N. (2018). A conceptual framework to compare two paradigms of augmented and mixed reality experiences. *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 7-18).

P256 Takahashi I., Oki M., Bourreau B., Kitahara I., Suzuki K. (2018). An empathic design approach to an augmented gymnasium in a special needs school setting. *International Journal of Design, 12*(3).

P303 Hsu T.-C. (2019). Effects of gender and different augmented reality learning systems on English vocabulary learning of elementary school students. *Universal Access in the Information Society, 18*(2), 315-325.

P314 Hsu H.-P., Wenting Z., Hughes J.E. (2019). Developing Elementary Students' Digital Literacy Through Augmented Reality Creation: Insights From a Longitudinal Analysis of Questionnaires, Interviews, and Projects. *Journal of Educational Computing Research, 57*(6), 1400-1435.

P320 Ewais A., Troyer O.D. (2019). A Usability and Acceptance Evaluation of the Use of Augmented Reality for Learning Atoms and Molecules Reaction by Primary School Female Students in Palestine. *Journal of Educational Computing Research, 57*(7), 1643-1670.

P330 López-Faican, L., Jaen, J. (2020). EmoFindAR: Evaluation of a mobile multiplayer augmented reality game for primary school children. *Computers & Education, 149*, 103814.

P335 del Río Guerra, M.S., Martínez, A.E.G., Martin-Gutierrez, J., López-Chao, V. (2020). The limited effect of graphic elements in video and augmented reality on children's listening comprehension. *Applied Sciences*, *10*(2), 527.

P376 Rossano, V., Lanzilotti, R., Cazzolla, A., & Roselli, T. (2020). Augmented Reality to Support Geometry Learning. *IEEE Access, 8,* 107772-107780.

W104 Gopalan, V., Zulkifli, A. N., & Bakar, J. A. A. (2016, August). A study of students' motivation using the augmented reality science textbook. In *AIP Conference Proceedings* (Vol. 1761, No. 1, p. 020040).

W106 Huang, T. C., Chen, C. C., & Chou, Y. W. (2016). Animating eco-education: To see, feel, and discover in an augmented reality-based experiential learning environment. *Computers & Education, 96*, 72-82.

Part 2: In-text bibliography

Akçayır, M., & Akçayır, G. (2017). Advantages and challenges associated with augmented reality for education: A systematic review of the literature. Educational Research Review, 20, 1-11.
Arksey H, and O'Malley L. (2005). Scoping studies: towards a methodological framework. International Journal Social Research Methodology, 8, 19-31.
Dey, A., Billinghurst, M., Lindeman, R. W., & Swan, J. (2018). A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. Frontiers in Robotics and AI, 5, 37.
Garzon, J., & Acevedo, J. (2019a). Meta-analysis of the impact of Augmented Reality on students' learning gains. Educational Research Review, 27, 244-260.
Heintz, M., and Law, E. L. C. (2018, September). PDotcapturer and PDotanalyser: participatory design online tools for asynchronous idea capturing and analysis. In *Proceedings of the 10th Nordic Conference on Human-Computer Interaction* (pp. 622-633). ACM.
Ibáñez, M. B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, 123, 109-123.
ISO 9241-11:2018. Accessed online: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en 10/12/2020.
ISO 9241-210:2019. Accessed online: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en 16/12/2020.
Kotzageorgiou M., Kellidou P.-M., Voulgari I., Nteropoulou-Nterou E. (2018). Augmented reality and the symbolic play of pre-school children with autism. *Proceedings of the 17th European Conference on e-Learning* (pp. 273-280).
Lorusso M.L., Giorgetti M., Travellini S., Greci L., Zangiacomi A., Mondellini M., Sacco M., Reni G. (2016). Giok: An alien stimulates pragmatic and social skills in pre-school children. *Proceedings of the 4th Workshop on ICTs for Improving Patients Rehabilitation Research Techniques* (pp. 89-92).
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS med, 6*(7), e1000097.
Nielsen, J. (1994). *Usability engineering*. Elsevier.
Santos, M. E. C., Chen, A., Taketomi, T., Yamamoto, G., Miyazaki, J., & Kato, H. (2013). Augmented reality learning experiences: Survey of prototype design and evaluation. *IEEE Transactions on Learning Technologies, 7*(1), 38-56.
Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology, 70,* 747-770.

Appendices

Appendix A: Excerpts from the Heuristic Evaluation report for the WWL program

As the entire HE report (see section 2.2 for details) would be too long to add in this appendix, we present some representative excerpts from it, covering different webpages and levels of the program.

ARETE
Heuristic Evaluation
WordsWorthLearning Website
10 - 18 February 2020

Aim of session

To get an overview over the current functionality of the WordWorthLearning (WWL) website to be transformed into an Augmented Reality application in the course of the ARETE project and to provide feedback from an HCI perspective, the University of Leicester ARETE team performed a Heuristic Evaluation.

Method

Three HCI specialists from the University of Leicester ARETE team went through the process of exploring the WWL website to provide feedback from an HCI perspective and to check for any usability problems or bugs. Several tasks and sequences were tried out to ensure the correct performance of WWL under different circumstances.

The main evaluation took place between 10/02/2020 and 18/02/2020 over 4 sessions (10/02: 10:30 – 14:45; 11/02: 09:30 – 12:45; 12/02: 09:45 – 11:00; 13/02: 12:00 – 13:00, 18/02: 09:30 – 11:00), during which the reviewers tried to imagine themselves in the role of children (from the age of 6) and adults with spelling difficulties (common with dyslexia), aware of a wide range of computer expertise among these population. Some concepts closely related to usability and user experience (such as the aesthetic and affective factors) were also taken into account for this evaluation.

Observations made during the Heuristic Evaluation were documented for this report. After the session, the usability problems were circulated around the review team so that all could independently assign severity ratings (H - high / M - medium / L - low) for fixing each problem:

- Low importance (L) rating is given for problems, which would be noticed by end-users, and might affect their overall sense of the quality of the interface, but would not hinder them significantly in achieving their objectives.
- Medium importance (M) rating is given for problems, which would be noticed by end-users and may confuse, delay, or distract them briefly and temporarily.
- High importance (H) rating is given for problems, which would be an obstacle for end-users, either preventing them from achieving their goals, or causing significant delay, disruption, confusion, or annoyance.

Finally discrepancies in severity ratings were discussed and consensus was achieved for each usability problem.

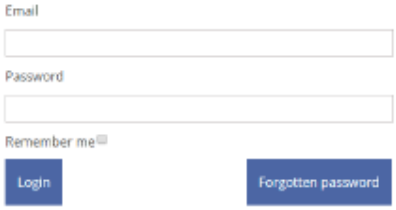
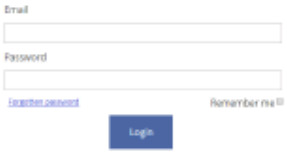
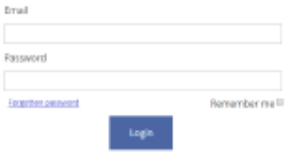
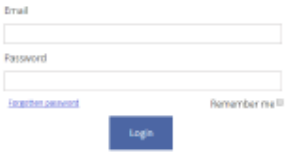
Equipment

- One laptop running on Windows 10
- Screen resolution: 1366 x 768
- Browser: Google Chrome, Version 79.0.3945.130 (32-bit)

1

Detailed Findings – WWL login page

Based on our observations and discussions regarding the usability of the WWL login page (https://www.wordsworthlearning.com/v2/auth/login), we recommend the following modifications:

	Usability Problem	Recommended Modification	Severity
1	The "Forgotten password" button is in the place where users would expect the "OK" or "Login" button, based on common dialogue designs. 	Swap the position of "Login" and "Forgotten password" button. Suggested layout: 	M
2	"Login" and "Forgotten password" button have the same appearance.	As the main task on this screen / in this dialogue is to log in, the emphasis should be on "Login" button. Suggested layout: 	L
3	The button-design of "Forgotten password" makes it look like a way to submit the form. As this is not the case, this design is misleading.	"Forgot password" should be a link instead of a button. Suggested layout: 	L

	Usability Problem	Recommended Modification	Severity
4	Help text leads to an additional step for the user to reach their goal (they have to search for the contact button before being able to send a message). If you need help, send us a message using the contact button.	Help text should be a link to the contact form instead of or in addition of instructions on how to contact WWL.	M
5	The text is misleading, as it is a "contact entry in the menu" not a "button". If you need help, send us a message using the contact button.	Rephrase the text: "If you need help, send us a message using the contact entry in the main menu above.".	L
6	Having the help text in bold makes it look like the most important task on the page.	Have the text being not bold instead.	L
7	Footer does not look nice, as it does not align with the rest of the page and is asymmetrical. 	Redesign footer layout. E.g., so that it is consistent with the footer in the learning area of the page: 	L
8	Footer is not visually separated from the content of the page. That makes it harder for users to distinguish between content and footer information.	Clearly visually separate page footer and content. E.g., consistent with the footer in the learning area of the page: 	L
9	Linked and non-linked text in the footer has the same design. The user can thus not visually distinguish between links and text. 	Design of linked and non-linked text should be visually different. E.g., consistent with the footer in the learning area of the page: 	M

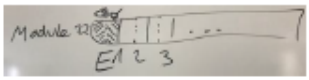
	Usability Problem	Recommended Modification	Severity
15	Changing your password should not be done in the same dialogue as resetting a forgotten password, as these are two different concepts.	Create separate dialogues for resetting and changing passwords. Changing your password should be provided as an option in your profile or settings once you are logged in.	M

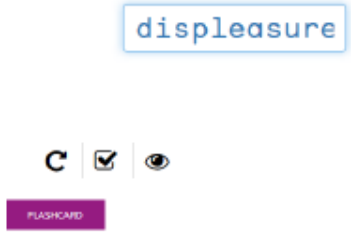
Detailed Findings – WWL course landing page

Based on our observations and discussions regarding the usability of the WWL course landing page (https://www.wordsworthlearning.com/v2/), we recommend the following modifications:

	Usability Problem	Recommended Modification	Severity
16	White font on bright coloured backgrounds make it hard to read the text. 	Use a black outline around the white text to improve readability.	M
17	The layout lacks uniformity in the level description boxes.	Make the level description boxes all the same length.	L
18	The colons at the start of the level descriptions look out of place.	As the design already separates level names and description, the colons should be omitted.	L
19	Using words to indicate the Module status adds a lot of reading for the user. 	Instead of words the module status should be indicated by icons (e.g. a green tick mark for "finished").	L
20	Missing visual separation between the different modules makes it harder to identify information that belongs together (visual grouping). 	Leave more space between the different module columns in each row.	L

	Usability Problem	Recommended Modification	Severity
36	The background pattern of the progress bars is not uniform (you can see where it repeats itself). 	Either do not use a pattern for the background (plain colour) or at least make sure it is tiled in a way that the repetition is not recognisable.	L
37	It is unclear, why the module bar is longer than the level bar and if the length of the progress bar correlates to length of time. 	The relation between module and level progress bar should be more apparent. For example: 	L
38	The Vowel chart does not have a title, like the Consonants chart. 	Add a title to the vowel chart.	M
39	A lot of effort is required to scroll back to the top of the page after looking at all the charts.	A link "back to the top of the page" would be useful.	M
40	After the video is finished, there is only the option to continue to the next module. In case the user missed something, they are not able to repeat the video directly. 	There should be an option to replay the video and the icons to navigate to the charts and to open the info popup should also not disappear.	M
41	Users expect high resolution videos even on the Internet.	A video with a higher resolution should be made available for users with high bandwidth.	L
42	The title of Module 3 does not tell the user what Module 3 is about. 	Provide a meaningful title for Module 3.	L

	Usability Problem	Recommended Modification	Severity
59	The module progress bar during the "Consonant Chart Exercise" gives some indication of how far into the exercise the user is. But because users have to map the length of the progress bar to the percentage of exercise completed accurate progression cannot easily be ascertained. 	Either indicate the different exercises in the module progress bar or indicate the last exercise, e.g. "Exercise 5 / 17".	L
60	The module progress bar in the "Consonant Chart Exercise" does not behave as expected. It moves only forward after finishing an exercise, thus not indicating the progress in doing an exercise.	The progress bar should move two steps for each exercise. One when the sound is played back and one when the next button is pressed. 	L
61	For the "Vowel Chart Exercise" the solution to the "oo" exercise is presented inconsistently to the visualisation on the chart. 	The "oo" in the solution should be underlined so that the unique identifier for this sound is correct: "oo in book".	M

	Usability Problem	Recommended Modification	Severity
88	The input box in the spelling exercises is sometimes too small for the solution to fit, 	Make sure the input box is long enough for the intended input.	L
89	There is a "Next Module" button at the end of the last module, which leads to a warning/error message. 	Remove the "Next Module" button when there is no next module.	H

	Usability Problem	Recommended Modification	Severity
106	When playing back a sound in the charts this seems to be added to the history of your browser, leading to the back and forward button in the browser not working as expected.	Playing back a sound in the chart should not alter the browsing history.	H

Positive observations:

- The colourful course landing / overview page is very visually appealing, especially for younger children. Matching the WWL logo increases brand recognition.
- It is nice that the module progress bar is in sync with the video progress bar and can also recognize and is updated accordingly if you go back or jump ahead in the video.
- It is very good that the charts are interactive, making good use of being digital content, so that the user can hear the sounds in place.
- It is good that the user cannot move on to the next exercise before checking the current one. The system prevents it and gives a good message, explaining what is expected from the user to solve it.
- It is a nice touch that the colour of the avatar in the exercises changes with the level the user is in.

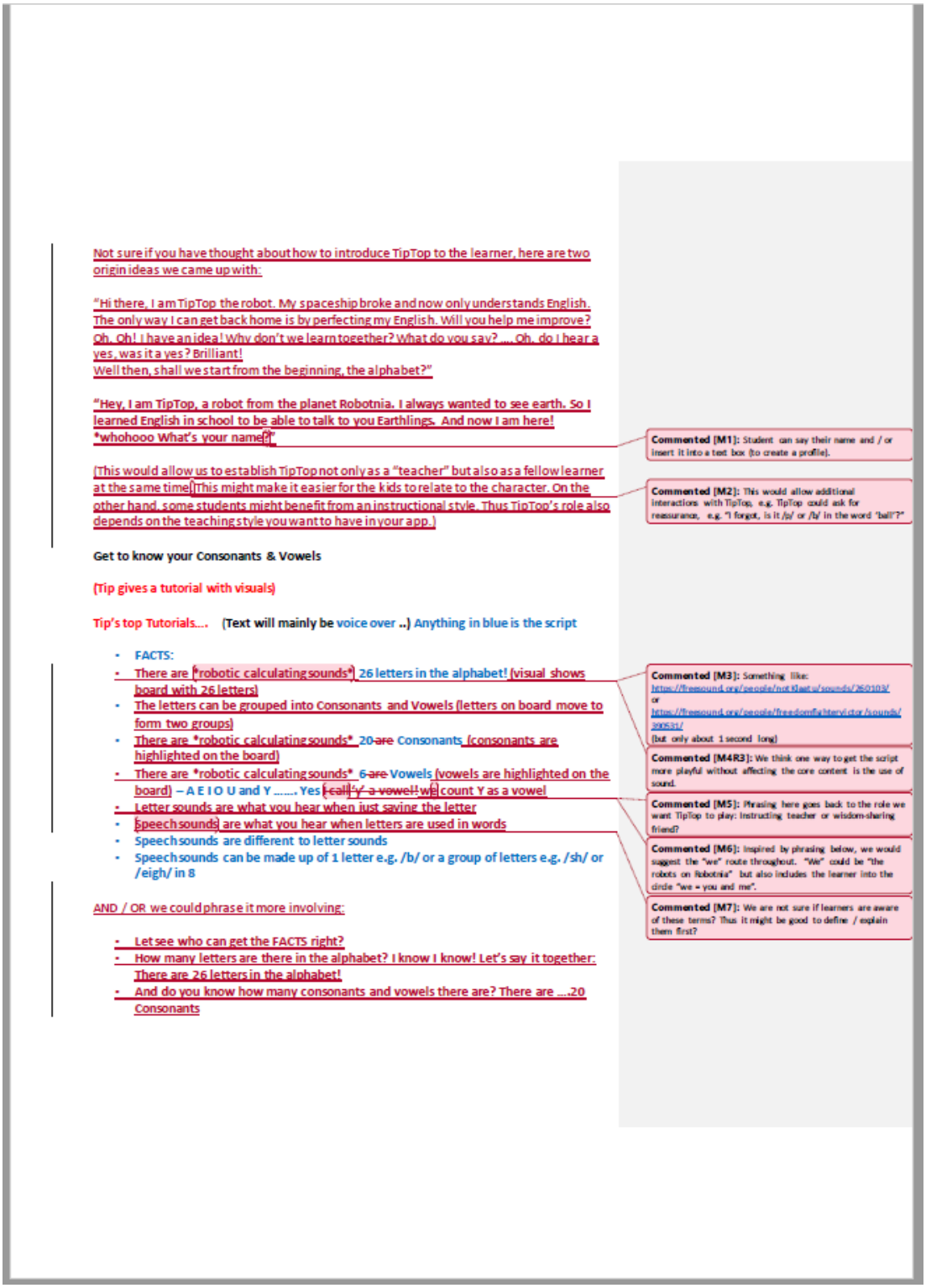
Summary

Some problems are highlighted on specific examples based on where they were encountered by the reviewers, however, they should be generalised (e.g. "heading of a specific section should clearly express the content" => "all headings in WWL should clearly express the content of the page"). Most of the 106 problems identified are related to outdated videos and inconsistency in layout and about half of the usability problems are only of low severity.

- 48 low-severity problems were encountered while using WWL. These do not represent a big impediment for the user to accomplish tasks, but would be beneficial to address to avoid confusion or dissatisfaction.
- 44 medium-severity problems could be obstructing the correct interactions with WWL. These issues should be fixed relatively soon as they represent a problem for the user.
- 14 high-severity problems should be immediately addressed to guarantee the proper performance of WWL.

Appendix B: Excerpts from the feedback on the WWL app scripts

As the entire WWL app scripts feedback (see section 2.2 for details) would be too long to add in this appendix, we present some representative excerpts from it, covering different aspects and levels of the app.



Not sure if you have thought about how to introduce TipTop to the learner, here are two origin ideas we came up with:

"Hi there, I am TipTop the robot. My spaceship broke and now only understands English. The only way I can get back home is by perfecting my English. Will you help me improve? Oh, Oh! I have an idea! Why don't we learn together? What do you say? .... Oh, do I hear a yes, was it a yes? Brilliant!
Well then, shall we start from the beginning, the alphabet?"

"Hey, I am TipTop, a robot from the planet Robotnia. I always wanted to see earth. So I learned English in school to be able to talk to you Earthlings. And now I am here!
whohooo What's your name?"

(This would allow us to establish TipTop not only as a "teacher" but also as a fellow learner at the same time. This might make it easier for the kids to relate to the character. On the other hand, some students might benefit from an instructional style. Thus TipTop's role also depends on the teaching style you want to have in your app.)

Get to know your Consonants & Vowels

(Tip gives a tutorial with visuals)

Tip's top Tutorials.... **(Text will mainly be voice over ..)** Anything in blue is the script

- FACTS:
- There are *robotic calculating sounds* 26 letters in the alphabet! (visual shows board with 26 letters)
- The letters can be grouped into Consonants and Vowels (letters on board move to form two groups)
- There are *robotic calculating sounds* 20 ~~are~~ Consonants (consonants are highlighted on the board)
- There are *robotic calculating sounds* 6 ~~are~~ Vowels (vowels are highlighted on the board) – A E I O U and Y ..... Yes ~~I call 'y' a vowel!~~ we count Y as a vowel
- Letter sounds are what you hear when just saying the letter
- Speech sounds are what you hear when letters are used in words
- Speech sounds are different to letter sounds
- Speech sounds can be made up of 1 letter e.g. /b/ or a group of letters e.g. /sh/ or /eigh/ in 8

AND / OR we could phrase it more involving:

- Let see who can get the FACTS right?
- How many letters are there in the alphabet? I know I know! Let's say it together: There are 26 letters in the alphabet!
- And do you know how many consonants and vowels there are? There are ....20 Consonants

Commented [M1]: Student can say their name and / or insert it into a text box (to create a profile).

Commented [M2]: This would allow additional interactions with TipTop, e.g. TipTop could ask for reassurance, e.g. "I forget, is it /p/ or /b/ in the word 'ball'?"

Commented [M3]: Something like: https://freesound.org/people/notklaat/sounds/260103/ or https://freesound.org/people/freedomfaltheryl.tor/sounds/390531/ (but only about 1 second long)

Commented [M4R3]: We think one way to get the script more playful without affecting the core content is the use of sound.

Commented [M5]: Phrasing here goes back to the role we want TipTop to play: Instructing teacher or wisdom-sharing friend?

Commented [M6]: Inspired by phrasing below, we would suggest the "we" route throughout. "We" could be "the robots on Robotnia" but also includes the learner into the circle "we = you and me".

Commented [M7]: We are not sure if learners are aware of these terms? Thus it might be good to define / explain them first?

- And 5 vowels …!(error noise is produced)….TipTop says: there are 6 Vowels – A E I O U and Y …….. Yes I call 'y' a vowel too!
- This makes two groups of letters: Consonants and Vowels
- I learned that writing is not same as speech and
- Speech sounds are different to letter sounds which are used in writing
- Speech sounds can be made up of 1 letter e.g. /b/ or a group of letters e.g. /sh/ or /eigh/ in 8

Did you know? There are 44 speech sounds in English!

- So what do you need to make speech sounds…..?
- tongue, lips, teeth and of course air …… sometimes lots of air!
- Some sounds are quiet and some sounds are LOUD ……
  (I envisage a clip here of fingers to Tips neck demonstrating no vibration with /p/ and then an exaggerated vibration with /z/)

Do you remember the 2 groups of letters?

Yes …….Consonants & Vowels

When I learned English we had a Consonant Chart. I brought one. Look!
TipTop pulls out charts…
Let's look at the Consonant Chart…)
(Which 3 letters are missing?)

yes ….. q x & (y – (I'll explain why later)….)
(printing sound¶) I made a copy for you. You can keep it! (Consonant Chart is added to the player's inventory)

Tips Top Tips:

- Letter names can be confusing when trying to spell a word. They mostly don't sound like the letter when used in a word.) Thus let's not~~don't~~ say the alphabet letter name e.g. P T K , Let's say the alphabet letter sounds instead: /p/, /t/, /k/
- Where is the air coming from?... For most letters mainly the mouth. Let's call them ORAL sounds. ~~but~~ But how about /m/, /n/, /ng/? …. Yes it's your nose. Let's call them NASAL sounds! I have a tiny nose, can you pronounce them for me?
- Do you remember that~~Remember~~ some consonant sounds are quiet and some are loud?!

Sounds on the left of the chart are quiet (show left column of consonant chart) so no vibrations & the sounds on the right (show right column of consonant chart) are loud and

Commented [M8]: It would be helpful for the user to have a brief explanation for the reason why the letter "y" counts as a vowel (vowel sound), before moving on.

Commented [M9]: If applicable for the rest of the scripts, we could have several "Did you know" knowledge bites. TipTop's face could become a lightbulb for those. A bit like: https://www.istockphoto.com/gb/vector/did-you-know-text-with-hand-drawn-light-bulb-concept-of-wise-council-or-funny-fact-gm1141391108-305765514

Commented [M10]: They have not been called "groups" before. (we adapted the script above accordingly though)

Commented [M11]: We would suggest to give an explanation, where the consonant chart comes from.

Commented [M12]: We are not sure about the reason for this question at this stage of the program.

Commented [M13]: Not sure if attention span of students will be long enough. Maybe skip it here and put it where the explanation is?

Commented [M14]: "Y because it is a vowel. I'll explain later why Q and X are not there."

Commented [M15]: (like a shorter version (2 or 3 seconds) of: https://freesound.org/people/aramill/sounds/345054/

Commented [M16]: We think a reasoning might help the learner to understand. When and why should they not say the letter name but the sound?

Commented [M17]: Depending on the teaching style we could use positive suggestions rather than negative instructions. What do you think?

really vibrate. (I envisage using the same demo as above for /p/ & /d/ but more exaggerated)
(Will there be charts or is this all in the 'bank / library' – need to discuss this with Kest)

- So what do we use to Make the consonant "SOUNDS"? Let's check it out!

- Have a game here – Student must click on which icon(s) displayed on the screen are involved in making each sound – aim is to build up oral awareness. They will hear a sound then they have to click all relevant corresponding icons
  - Tongue icon
  - Teeth icon
  - Lips icon
  - Air from mouth icons (puff, hiss, gently flow, forced push)
  - Air from nose icon
  - Loud / vibrating sound
  - Quiet / no vibrating sound

Keeping the animations – play the video and then student must click what are in use ? Don't need to use the web online videos explaining movement in detail for the app.

- Starting at the top of the chart....?????? **(yet to determine how these sounds will be presented)**
- See the video, hear the sound then click on the relevant icons to demonstrate what is being used.....

(no dialogue needed here, these sounds will be played and the student points to relevant icons)

- **p**
  - Lips icon
  - Air from mouth icon
  - Quiet / no vibrating sound
- **b**
  - Lips icon
  - Air from mouth icon
  - Loud / Vibrating sound
- **t**
  - Tongue icon
  - Teeth icon
  - Air from mouth icon
  - Quiet / no vibrating sound
- **d**
  - Tongue icon
  - Teeth icon
  - Air from mouth icon
  - Loud / vibrating sound

Commented [M18]: Maybe we could have a built in tutorial here, by TipTop answering the first question, by selecting some icons, and then asking the player, if that was correct: "Did I get that right?" followed by the prompt to the player to try it themselves: "Thanks! Now you have a go!"?

Commented [M19]: Will the app analyse if the correct icons have been clicked? Then TipTop could say "Correct!", "Well done!" and the like, or "Instead from the nose, air comes from the mouth when making this sound." and "Remember, you also use your lips for this sound!"

Air from mouth icon
Loud / vibrating sound

- m

Lips icon
Air from nose icon
Loud / vibrating sound

- n

Tongue icon
Teeth icon
Air from nose icon
Loud / vibrating sound

- ng

Tongue icon
Air from nose icon
Loud / vibrating sound

Need to bring in a game here to teach the consonant sounds with a reference to the charts / sound bank.
E.g. a sound is called out and they pick it out from an array of letters? To make it more difficult or for bonus points if they can identify what articulators are being used (tongue, teeth, lips) and if its voiced or voiceless they get more points.

Tips tutorial again ….

- Now let's have a look at the VOWEL CHART
- When I was a kid I was taught that there are only 5 vowels in English and these are……
  A E I O U
- It's just not true!
- There are 19 vowel sounds in English and you got to know them!

Tips tutorials…

Can you work out what shape your lips are when you say /ee/? /oo/, /ar/ & /ow/?
Tick which one is relevant SPREAD, ROUND, RESTING & MOVING. (Maybe these should be 4 closed boxes?)

- Let's have a closer look at the Vowel Chart:
- 1st let's find all the vowels in Green:
- In the SPREAD Group there is ee and ae
- In the ROUND Group there is oe
- In the MOVING Group there is ue and ie
- What do they all have in Common?
- They all have an 'e' at the end

Commented [M20]: See comment on Consonant Chart above.

Commented [M21]: We found that this sounds rather harsh. Is "vowel sounds" the same as "vowel"?
We would suggest something like: "On Robotnia we were taught that there are only 5 vowels in English. And that they are A E I O and U. That's true. But those 5 vowel letters result in 19 different vowel sounds in English! That's impressive, huh? *proudly: And I have learnt them all!"

Commented [M22]: The transition from one section to the other should follow a pattern and be smoother, currently it feels unconnected and a bit jumpy. (Also depends on the game design, if there is a change of location in between, then having a break in the script makes more sense.)

- Do you know what it's called? :
- Silent 'e' / Magic 'e' / Don't Know…… We call it a silent 'e' simply because it is silent – you NEVER hear it or say it.
- But it has a really important job to do – it tells you to say the ALPHABET name of the 1st vowel e.g. A, E, I, O, U
- These are the only 5 times that you say the name of an alphabet letter!

Ok, now let's go back up to the SPREAD Group (*add in a start button and an arrow to show the direction the chart should be read – like a clockface – this is redundant if a bank / library idea is being used*)
~~At~~ For the sounds at the beginning your lips are very spread, ~~and~~ They become less and less spread as you move around the circle or along the bank / library shelf ????? Perhaps have boxes which open to show the sound options?

E.G.

ee	in *bee*
i	in *ink*
e	*in elephant*
ae	*in snake*
a	*in apple*

Ok, now let's go to ~~for~~ the ROUND Group – ~~at~~ for the sounds at the beginning your lips start off very round … like a kiss, ~~and then~~ They become less ~~rounds~~ as you move around.
E.G.

oo	in tooth
oo	in *book*
oe	*in toe*
or	*in orange*
au/aw	*in sauce or saw*
al / all	*in bald or ball (not a pure vowel but we snuck them in!)*
o	*in box*

OK. Now let's move on to the RESTING Group…
E.G.

er/ur/ir	in hammer, bird and turf
u	in cup
ar	in car

Finally the MOVING Group…… …where your lips move from one shape to another to make the sound.
E.G

ue	in statue
ie	in eye
oi/oy	in oil & boy
ou/ow	in house & cow

TIPS TOP TIPS:

General comments:

- The first few levels have been nicely improved, but we noticed a decline in friendliness, playfulness, and interactions of Tip Top with the player from Level 5 onwards. Phrases of encouragement aimed at the user (other than "you have earned more stars") seem to be missing from the later levels. We would therefore suggest to adapt the Levels 5 – 7 based on Levels 1 – 4.
- There also seem to be some inconsistencies in later levels with previous ones, e.g. with the sequence of landing on a planet, scanning the planet, and receiving transmissions about rules/data/facts/etc. followed by the "transmission noises" sound or which button (Tip Top, Audio button, …) to press to check the pronunciation in pronunciation games or in the introduction of games (with "Let's play a game to …").
- There seem to be some Tip Top recordings needed, that are not yet reflected in the script, which are currently only mentioned in the notes in square brackets.

Introduction/Begin Level 1

[User opens the app for the first time]

Tip Top:
Hi!

I didn't expect to see anyone else out here in space! I'm Tip Top, and I'm from WoWo World! I can tell by looking at you that you must be from Planet Earth, is that right? Wow you've travelled a very long way. Have you come to visit the Learning Solar System too?

The Learning Solar System is a biiiig system of planets where people come to learn all about English reading and spelling. The rulers of WoWo World have sent me here to gather as much information as I can. I'm not supposed to come home until I've visited every planet in the Learning Solar System, but it's so large, and I've never been on a mission like this before...

[Tip top looks worried, then brightens up]

Hey I have an idea! Let's explore the Learning Solar System together! That would be so great! Once we enter the Learning Solar System we'll need to visit every planet to learn something new about English reading and spelling. On each planet we'll be able to get your Galactic Passport stamped, and once it's full we can get our tickets home. So let's get going on this adventure together!

planet. BUT, if you like, you can always travel back to planets and areas we have visited before, to remind yourself what they were about and play the games again!

[Visuals on screen to illustrate this]

Got it? Ok, let's visit the first location on the planet, tap here to get started.

[The first zone on the planet pulses and the user taps it]
[The screen changes]

Tip Top:
"Beep boop" Hang on, I'm getting a transmission about this area…"transmission noises" I'm receiving some important facts – did you know that:

There are "robotic calculating sounds" 26 letters in the alphabet?

The letters can be grouped into Consonants

[The word consonant is displayed on the left side of the screen]

and Vowels

[The word vowel is displayed on the right side of the screen]

There are "robotic calculating sounds" 20 Consonants

[Consonant side populates]

And there are "robotic calculating sounds" 6 Vowels
A E I O U and Y
Yep! We count Y as a vowel!

[Vowel section is populating while Tip Top announces each one]

Tip Top:
This is how consonants and vowels look when they're in the alphabet, but that doesn't really help us when we're trying to speak! When we're talking – except for A E I O and U – we never say the name of letters, just their sounds!

We call these **speech sounds.**

["Speech sounds" appears on screen]

Speech sounds can be made up of 1 letter, for example
"b"

Commented [MH2]: To make it easier for the user to follow the explanation and create a mental model of the world we would suggest to use one term consistently throughout (i.e. "area", instead of area, location, zone).

Woah check it out! You've earned a star for completing this part of the mission! Remember if we collect enough stars we can exchange them for rewards.

[Star going into the rewards button]

We can go to our next stop and play a game, or you can take a break now.

[User has the option to continue or save and quit.]

TipTop:
Now that we've learned all the consonant speech sounds, let's play a game to help us remember all the sounds better!

[First Interactive consonant exercise]

"beep boop" Ok I'm downloading the game rules... "modem sounds" Yay, this will be fun!

Here is how it goes: You will hear a sound. Then you will have to pick the correct consonant from the options that appear. Let me do the first one!

[The audio "p" plays and 4 speech sound consonant options appear on screen. Tip Top flies to the option P and presses it. The P option is highlighted and after a short pause the "correct" indication (sound and visual) are played.]

This is the right consonant for this sound. Ok now it's your turn. Remember you can look at your consonant chart at any time by tapping this button. And I'm right here if you need me, just tap me if you want some help.

Commented [MH3]: What happens if the user taps on Tip Top?

[Consonant chart button highlights]

[The next sound is played and a new set of options appear]

Tip Top:
If you want to hear the sound again, just tap the audio button.

[Audio button highlights.]

[The game continues until the user has gotten every question right.]

[If the user gets a question wrong first time]

Tip Top:
Oops! Let's try again. I find that saying the sound out loud helps me think. And don't forget you can look at the consonant chart at any time to help you!

Well done! You've clearly worked hard at memorizing your flashcards! You've earned more stars! Are you ready to move to the next location on Planet PLARSH?

Commented [MH14]: We would suggest to keep the tone consistent with previous levels, for example: "We have collected plenty of stars here. I think we can move to the next area. What do you say? Shall we? Lets goooooo!"

[User taps button and screen changes]

[Story]

Tip Top:

Welcome to the Planet PLARSH storytelling centre! The people of Planet PLARSH love to tell each other stories, and reading stories is a great way to practise your reading skills. So let's have a look at the story of the day!

[Some interaction to start the story. User reads the short story]

Tip Top:

That was great! The people on Planet PLARSH are so inventive! You've earned more stars!

[Quiz]

Ok we're almost ready to leave Planet PLARSH, but first we need to take this quick quiz to get a stamp for our passport!

(Quiz is done here)

[When user passes the quiz there is some celebration animation and a visual of their passport being stamped]

Tip Top:

Great job! Now we can depart Planet PLARSH. Are you ready to visit our next planet or do you want to take a break?

[User is again given the option to take a break at this point]

END OF LEVEL 5

Level 6

[Introduction]

Tip Top:

Time to explore a new planet!

Commented [MH15]: Are you excited? I really am!

cow-boy

[This appears on screen]

Tip Top:
Let's play a quick game to make sure we understand the **compound words** rule when spelling!

"beep boop" downloading instructions…

I'm going to **say** a word and I want you to try to **spell** it using the **speech sounds** below. Take your time and remember you can look at the rule in your library if you need help. If you need to hear the word again just tap [this] button.

Commented [MH20]: Should be consistent: We would suggest to tap Tip Top over "Audio button" or "Rocket button".

[Audio button pulses]

Ready? Let's go…

[Spelling exercise continues]

Tip Top:
Great job! You've earned more stars! Now you know the **Compound Words** we can move on to the next location on this planet! Or you can take a break and go back to the main menu.

[User is given the option to take a break here]

[Rule 2 - Prefixes and Suffixes]

Tip Top:
"Beep boop" downloading instructions… Ok remember **Prefixes and Suffixes**? Yep, a prefix is a syllable that is ~~fixed~~ attached ~~on~~ to the start of the word and a suffix is a syllable that then goes to the end of a word. Great! Well the good news is we're going to skip this rule for now because we'll learn allllll about spelling Prefixes and Suffixes in the ZOIM Asteroid Field! We will travel there after leaving this planet. So let's move on to the next rule!

[Rule 3 - Y at the end]

Tip Top:
"Beep boop" receiving transmission… Ok, let's learn about spelling words with **Y at the end**. If you remember from earlier, when we have the letter 'y' at the end of a longer

[Quiz]

Ok we're almost ready to leave Planet BOOP, but first we need to take this quick quiz to get a stamp for our passport!

(Quiz is done here)

[When user passes the quiz there is some celebration animation and a visual of their passport being stamped]

Tip Top:
You're making so much progress! Now we can depart Planet BOOP. We now have enough stamps to get our ticket to leave the Learning Solar System, but before we do that we have one more destination to visit; the ZOIM Asteroid Field. Are you ready or would you like to take a break?

[User is again given the option to take a break at this point]

END OF LEVEL 6

Level 7

[Introduction]

Tip Top:
Wow, we've were almost finished our journey through the Learning Solar System! We've learned so much already! We now have enough stamps to get our ticket to leave the Learning Solar System!

[Celebration noises and visual of the ticket on screen]

Tip Top:
Oh but I almost forgot... see that huge asteroid field? It's called the ZOIM Asteroid Field and it's the only way out of the solar system! We will have to leap from asteroid to asteroid to make our departure, and on each one we'll learn all about prefixes and suffixes! Let's go on the final part of our journey!

[Some visual of the asteroid field on screen while Tip Top explains]

[2 Syllable Suffixes]

Tip Top:
Ok let's get going! *Beep boop* Scanning the asteroid field… this next stretch is a little more complicated, we'll be moving on to two syllable suffixes. You know by now that every syllable must have a 'working vowel' so we would expect to find **two working vowels** in a two syllable suffix.

Some examples of two syllable suffixes with two working vowels are…

-ab/le - in wash/ab/le
-ib/le - in hor/rib/le
-ous/ly - in fa/mous/ly

[These appear on screen]

Tip Top:
Of course we're talking about English here so, you guessed it, there are some exceptions to this rule. There are **two times** when a **two syllable ending** only has **one vowel**. Although you can hear a vowel in the final syllable, you do not write it!

This happens when the suffix ends in 'm'
for example...

-is/m - in bap/tis/m
-as/m - in sar/cas/m

[These appear on screen]

In these cases the 'm' is a syllable on its own.

Tip Top:
Another example of two syllable suffixes is when we have **two vowels together** in the suffix. We've already come across these on planet BOOP where we learned the **v/v rule or the two vowel rule.** For example

-uid - in flu/id
-ual - in grad/u/al
-uel - in cru/el
-eum - in mu/se/um
-eon - in ne/on

Commented [MH21]: Could some examples for two syllable suffixes be added as well?

Appendix C: Excerpts from the Heuristic Evaluation report for the CLB app

In the following we present some representative excerpts from the Heuristic Evaluation report for the CLB app (see section 2.3 for details), covering different screens of the app.

ARETE
Heuristic Evaluation
CleverBooks Geometry app
18 February 2020

Aim of session

To get an overview over the current functionality of the CleverBooks Geometry app and to provide feedback from an HCI perspective, the University of Leicester ARETE team performed a Heuristic Evaluation.

Method

Three HCI specialists from the University of Leicester ARETE team went through the process of exploring the CleverBooks Geometry app to get an overview over the current functionality and to check for any usability problems or bugs. Several tasks and sequences were tried out to ensure the correct performance of the CleverBooks Geometry app under different circumstances.

The main evaluation took place on 18/02/2020 over one session of 1.5 hours (11:30 – 13:00), during which the reviewers tried to imagine themselves in the student role. Some concepts closely related to usability and user experience (such as the aesthetic and affective factors) were also taken into account for this evaluation.

Observations made during the Heuristic Evaluation were documented for this report. After the session, the usability problems were circulated around the review team so that all could independently assign severity ratings (H - high / M - medium / L - low) for addressing each problem:

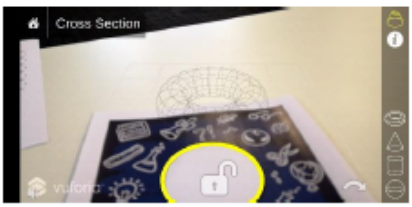
- Low importance (L) rating is given for problems, which would be noticed by end-users, and might affect their overall sense of the quality of the interface, but would not hinder them significantly in achieving their objectives.
- Medium importance (M) rating is given for problems, which would be noticed by end-users and may confuse, delay, or distract them briefly and temporarily.
- High importance (H) rating is given for problems, which would be an obstacle for end-users, either preventing them from achieving their goals, or causing significant delay, disruption, confusion, or annoyance.

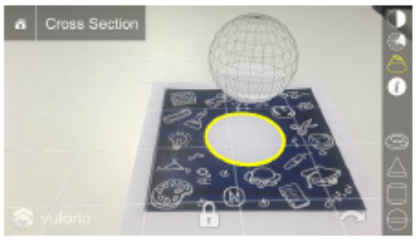
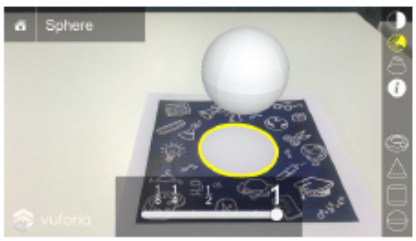
Finally discrepancies in severity ratings were discussed and consensus was achieved for each usability problem.

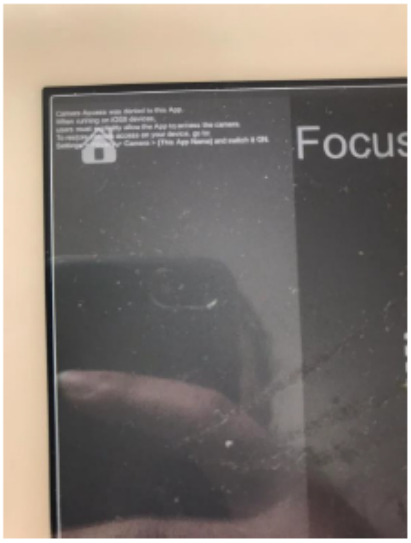
Equipment

- One Huawei P20, running Android version 9 (Pie)
- One Samsung Galaxy S9, running Android version 9 (Pie)
- One iPhone 7, running iOS 13.3.1
- One iPad, running iOS 10.3.4

1

	Usability Problem	Recommended Modification	Severity
5	When being too close to the marker attempting to create a cross section poses difficulties. It works fine from a certain distance away. 	Either make the cross section work fine from any distance or show a warning message, when the camera is too close to the marker.	H
6	If the user moves the camera too far away from the marker, the visual representation of the cross section plane ends before it intersects with the 3D object.  	Keep the cross section plane representation always the same distance away from the camera as the 3D object or make it unlimited in size.	M
7	The speaker and arrow button are rather small in comparison to the lock button, which makes it look like the lock button is more important. 	Have the arrow button the same size as the lock.	L

	Usability Problem	Recommended Modification	Severity
13	"Cross Section" changes the title, "Hemisphere" and "Fraction" do not.  	To support the user in knowing what is currently visualised the title should show the name of the 3D object followed by the name of the option selected, e.g. "Sphere – Hemisphere".	M
14	While being in "cross section" mode, when switching the displayed 3D object, the cross section mode is also left. This means two changes at the same time, when only one was actively triggered by the end user.	Stay in "cross section mode" when changing 3D objects.	M
15	Due to AR requirement of having the marker visible, the user cannot look on the bottom of a 3D object and the object can also not be rotated to do so.	It might be beneficial if the user could rotate the 3D objects to be able to inspect them from all angles.	H

	Usability Problem	Recommended Modification	Severity
21	The error message when not allowing camera access on iOS is hard to read. 	Re-position the error message and make it bigger.	M

Positive observations:

- On iPhone the purchase option for the game is behind a security measure, which is a very nice idea and asks for knowledge about shapes to unlock the purchase.

- It is nice that the (verbal) feedback to the test answers changes and is not always the same.
- It is good that the voiceover is done by a child, not an adult or an automated voice.

Appendix D: Excerpt of feedback on app design and alternatives by CLB

The following presents an excerpt of the feedback provided on CLB app designs and alternatives (see section 2.3 for details).

	The background image in this screen design does not distract from the interface elements, because there is no overlap. However, the connection between background image and app is not clear. The colour scheme and especially contrast between text and visuals is good, as it makes the labels easy to read. Having the dropdown arrow on the left hand side is unexpected, due to it being unconventional. To support the user in understanding and interacting with the interface, conventions should be followed. We would also recommend to have the icons on the left hand side of the text, as this is the more common approach and thus more familiar to the users. The logo seems to be unaligned to any of the other interface elements. We would recommend to either align it or put it in the top left hand corner to improve the visual composition of the screen for the end-user.
	There is a slight typo in "theacher" (teacher). The background of the house icon should either blend in more with the background of the screen or be more distinct, to make it a clear rectangular button. We would recommend to centre all the labels and interface elements - other than the home button - for a more pleasant visual arrangement. To comply with user interface conventions, there should be a "Submit" button visible underneath the input fields. Open question: Do teachers log in to a class like students or to an admin interface where they can manage their classes?

	This screen design is mostly fine, but the input boxes do not look like input boxes. To comply with user expectations, they should follow the conventions. The text is a little hard to read, which could be improved by making it bold or slightly bigger.
	The colourful design just around the buttons looks grabs the attention but at the same time divides it from the interactive elements. Having so many different colours can be confusing and distracting. With the white background the rainbow part looks cut out, creating the impression of an incoherent design, e.g. the logo looks detached from the rest.
	It is unclear to us, if the background image is supposed to be the live view from the camera or a background image for the app. If it is the latter, then the placement of the marker looks odd. The X- and "back arrow"-button seem to be way too small to press and distinguish. Because they are very close together, it might be hard to press one but not the other (accidentally). The differences in system actions triggered by "home button", "exit class", "back arrow", and "X"-button are unclear to us. Will users know the difference? The white text on rainbow background makes it hard to read. The icons on the right hand side are nice and clear in the meaning they are supposed to convey. However, they should not be displayed, when the class has not started yet. Instead of being a pop-up window, this

	screen for consistency.
	We assume this is an alternative way to joining a class, by scanning the respective QR code. As above, the functionality of the app used in the class should only be available after joining it and thus not be displayed in the background. This screen should be modelled after the "Join class" screens in the beginning of this document.
	See and combine feedback from above relevant for this screen. Login should not be displayed as stars, only the password.
	Error message needs to be more expressive than just stating "error" to follow the HCI recommendations to "allow users to recover from errors", for example, a reason needs to be specified in a language the user can understand. Also different options on how to progress should be offered, for example a "try again" button (or "forgot password", depending on the underlying error). The X-button should not be rainbow coloured, to be consistent with user expectations and easily recognisable. The only possible interaction for the user here seems to be to close the "Login error" message, everything else seems to be deactivated by a pink overlay. The X-button is way too small for that.

	See feedback on similar screens above. This error message is slightly better, as it gives some information in regards to the error that occurred. However, it should still be more precise and meaningful and offer alternative interactions to different solutions, other than just closing the error message. Avoid having several pop-ups on top of each other, as this clutters up the screen significantly.
	See feedback on similar screens above.
	Language selector and its options look too small for touch interaction. We would recommend taking up the whole screen with all the flags to choose from. The name of the app should be bigger, as it is kind of lost in between the other elements. Why are there “Continents” in the “Geometry” app. Is this the teacher version of the app or what does the “teacher hat” signify?
	See feedback on similar screens above. Not sure, what the “SCHOOL” refers to? Compared to the interface above, icons OR text should be used consistently.

Appendix E: Alien introduction as dialogues instead of monologue

The alien monologue in form of a large paragraph of text (see Figure 16 and section 0 for details) is broken down into smaller sections presented in the form of a dialogue between alien and students, as can be seen in Figure 17 and Figure 18.

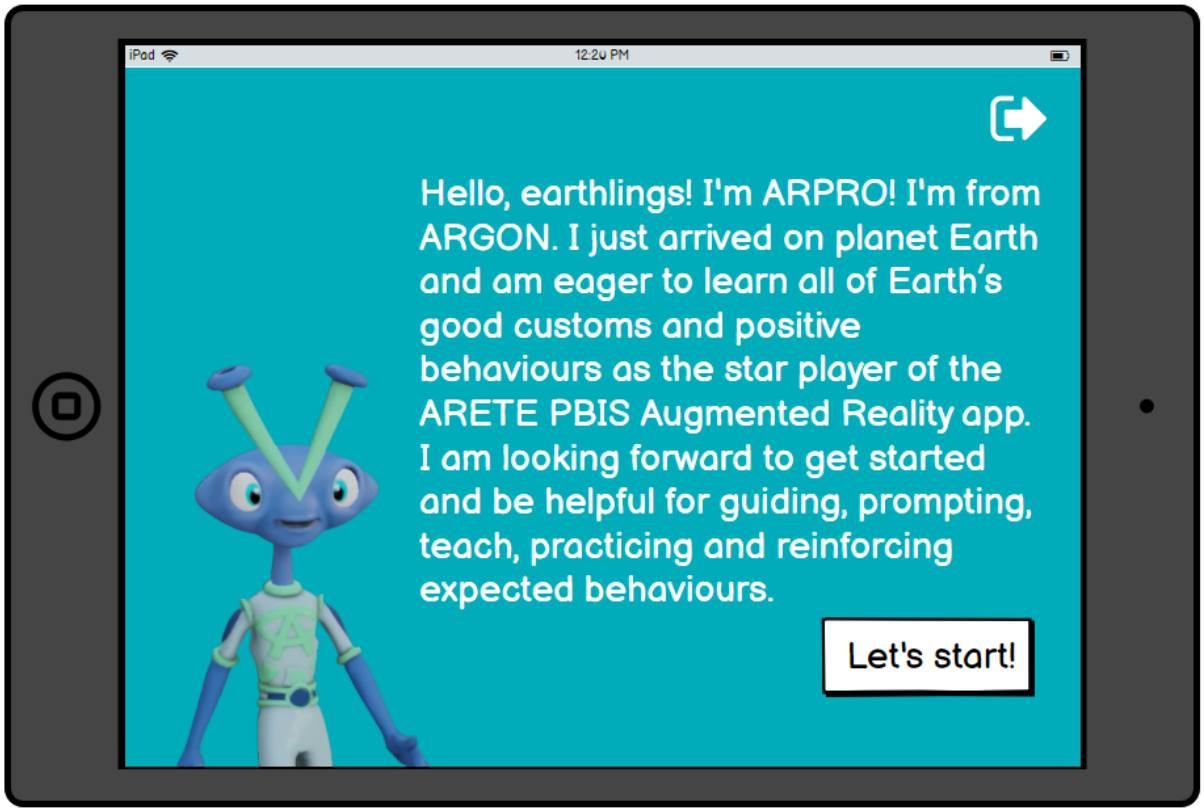


Figure 16: Before.

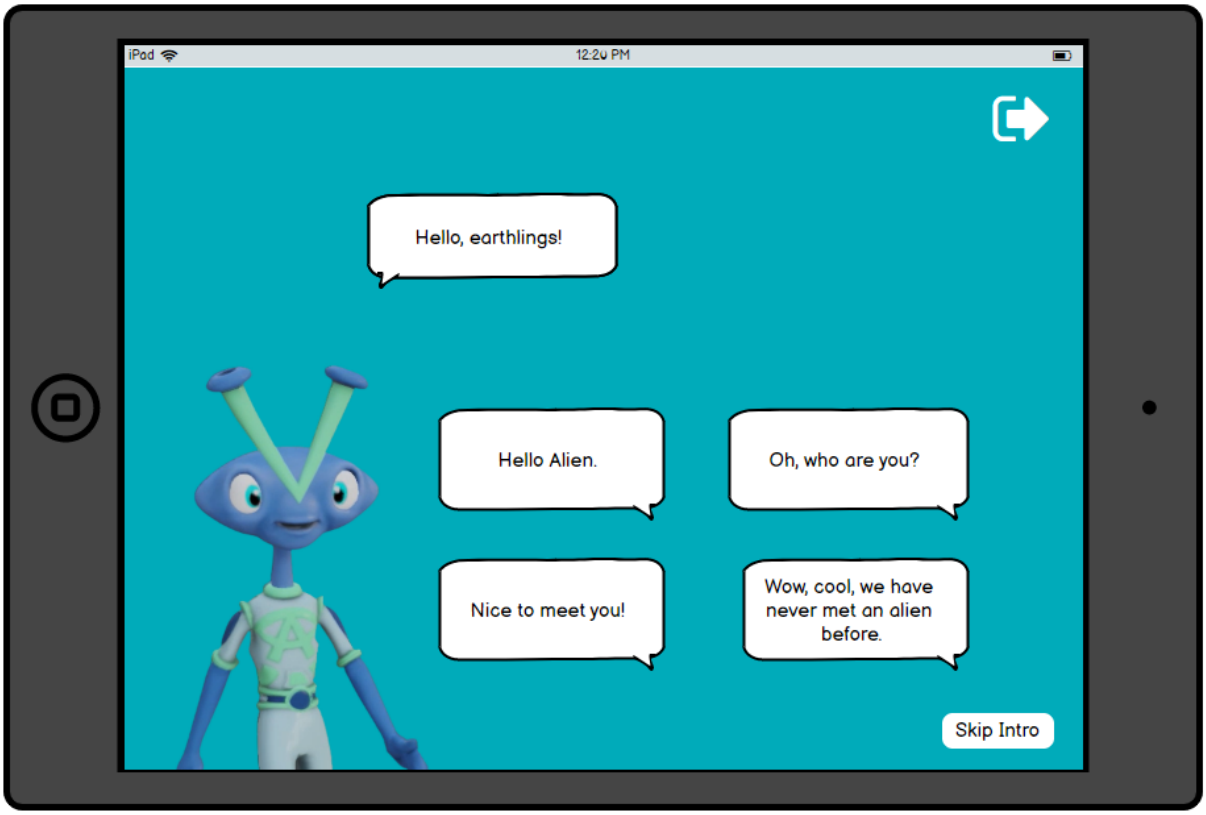


Figure 17: After 01.

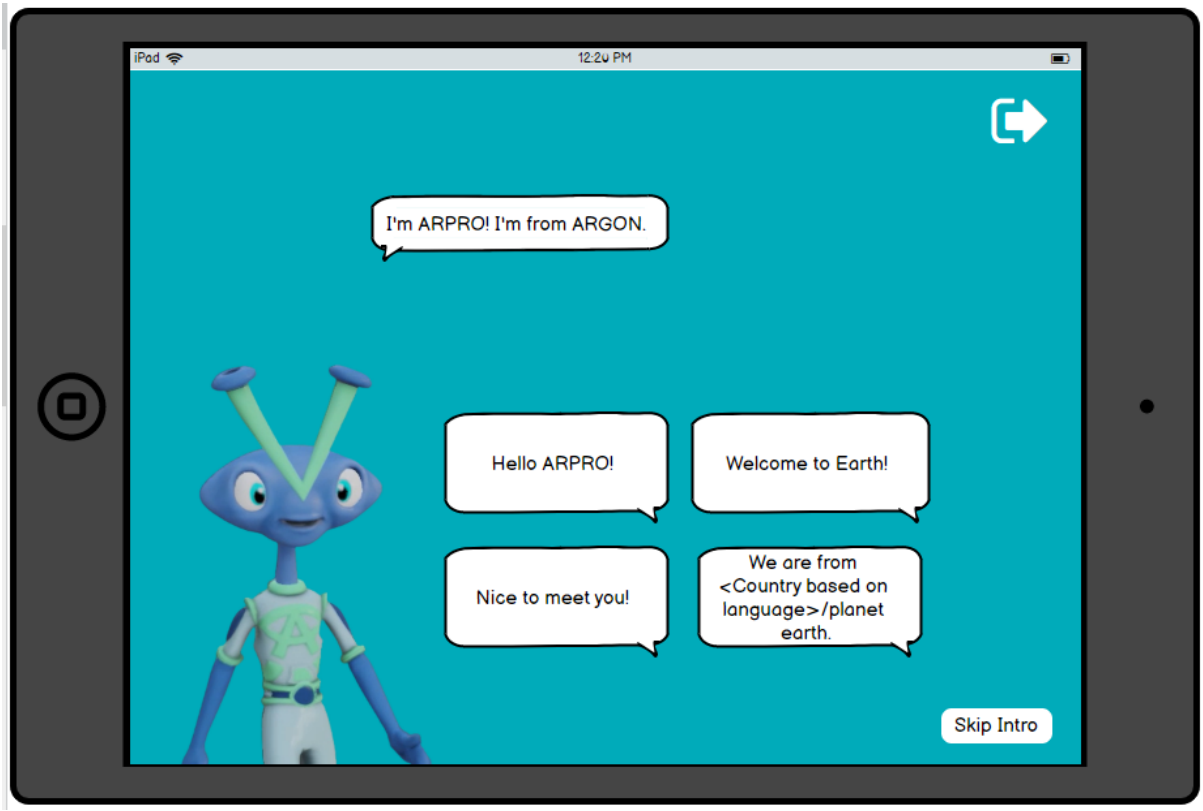


Figure 18: After 02.

Appendix F: Recording sequence

The following screenshots shows the suggestions on how to break down the complex recording screen into a sequence of screens.

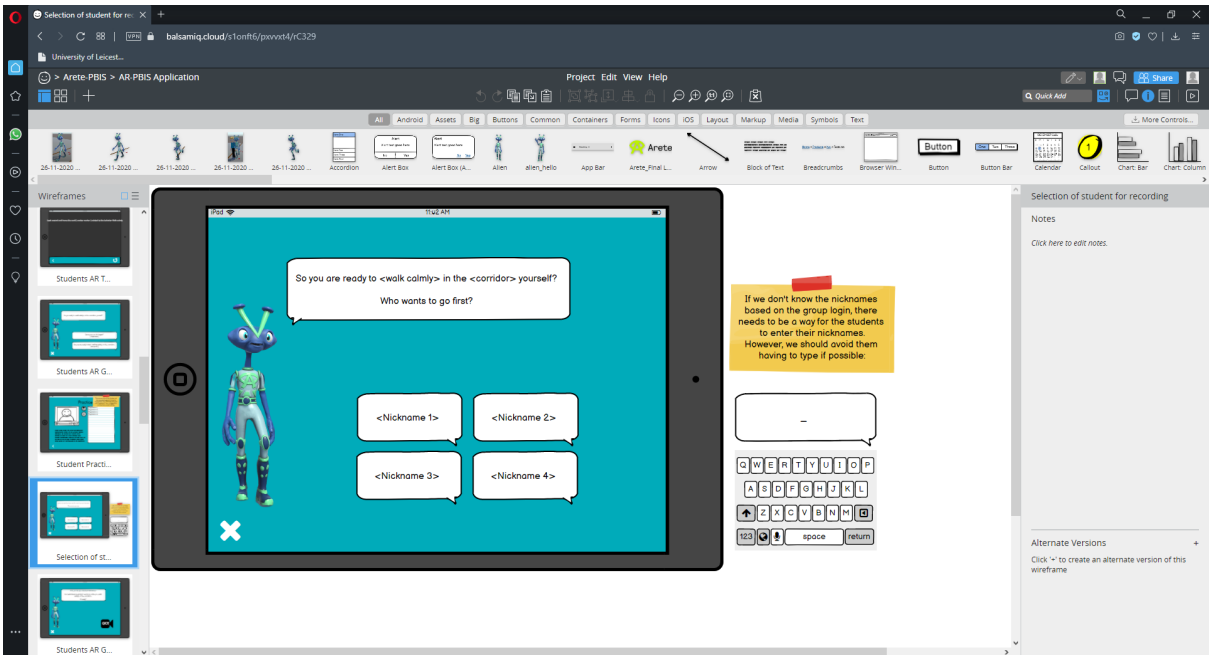


Figure 19: Select group member to record.

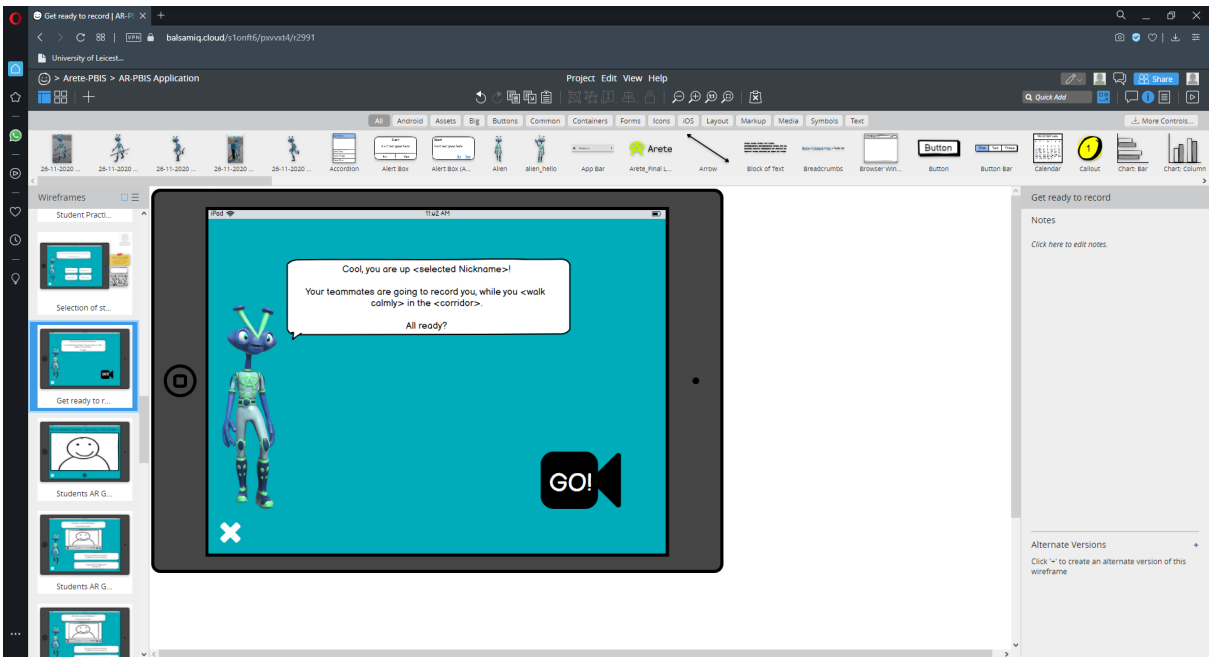


Figure 20: Get ready to record.

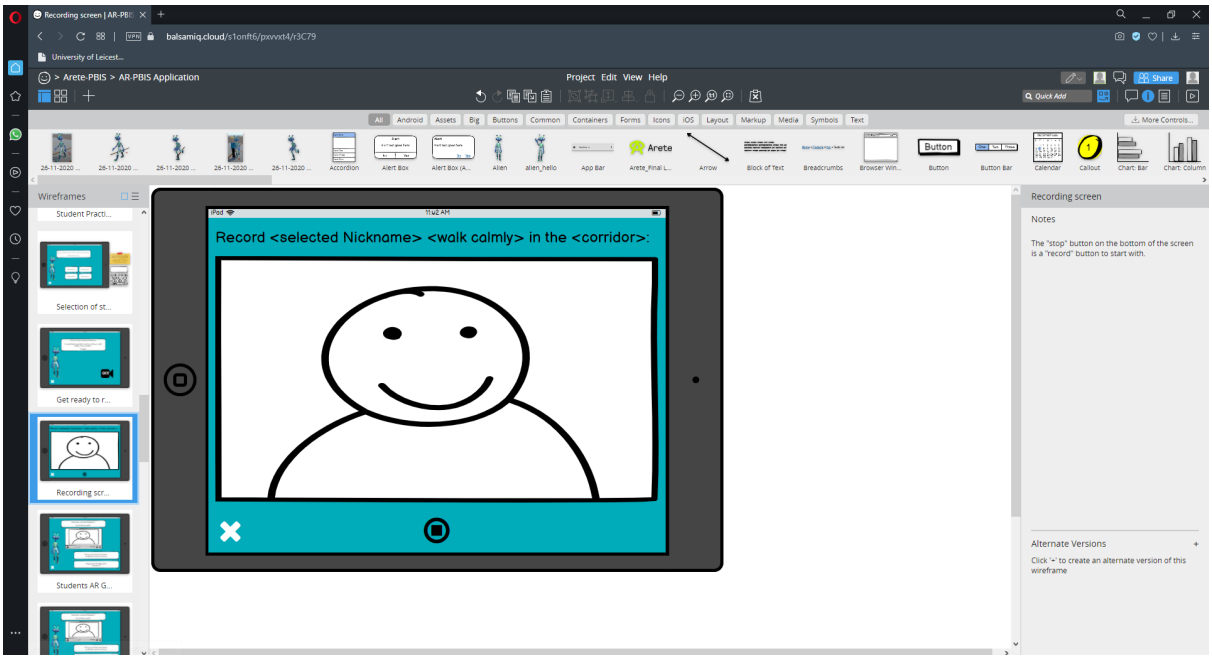


Figure 21: Behaviour recording.

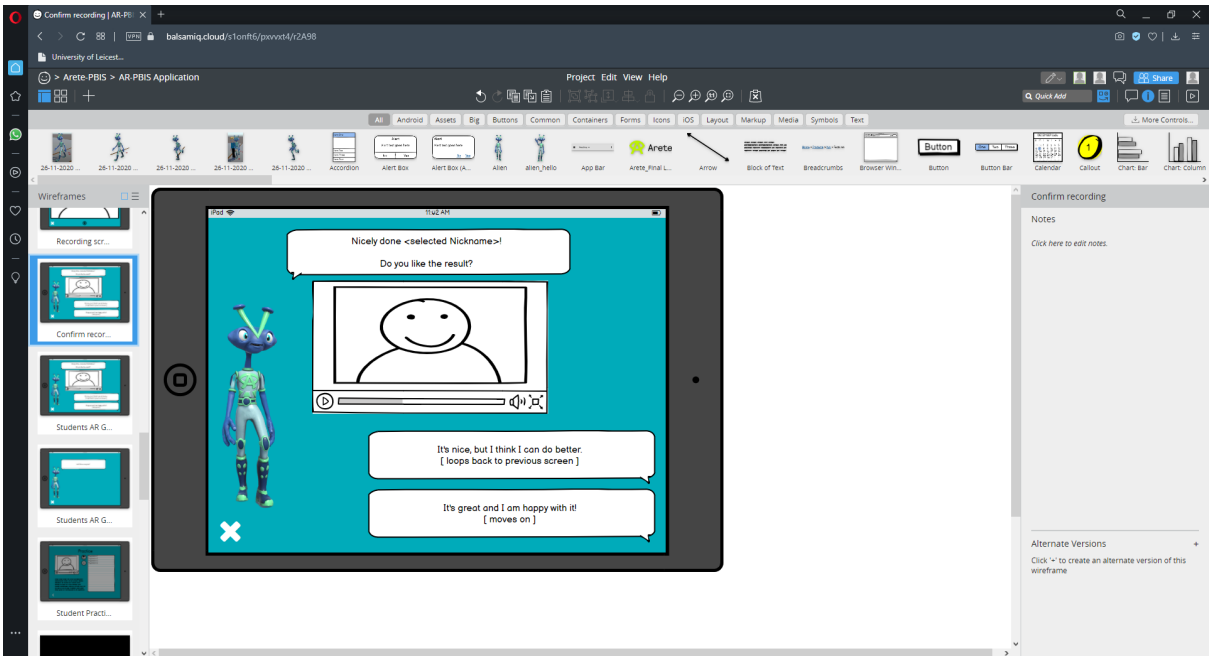


Figure 22: Confirm recording.

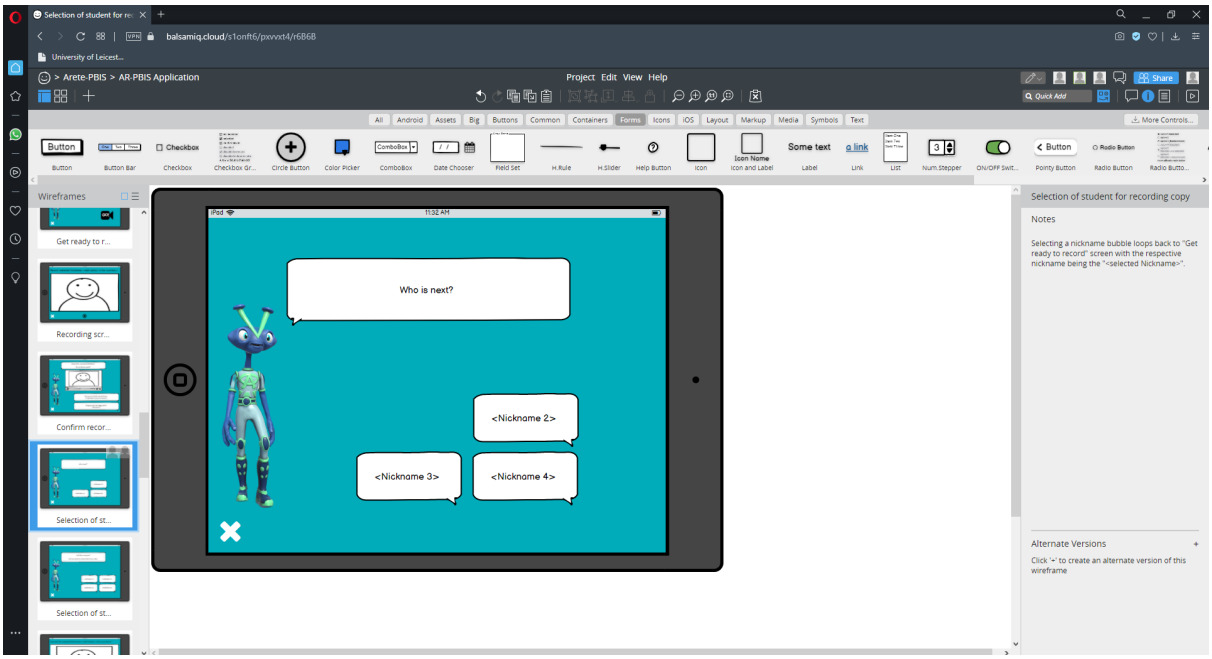


Figure 23: Select next student for behaviour recording.

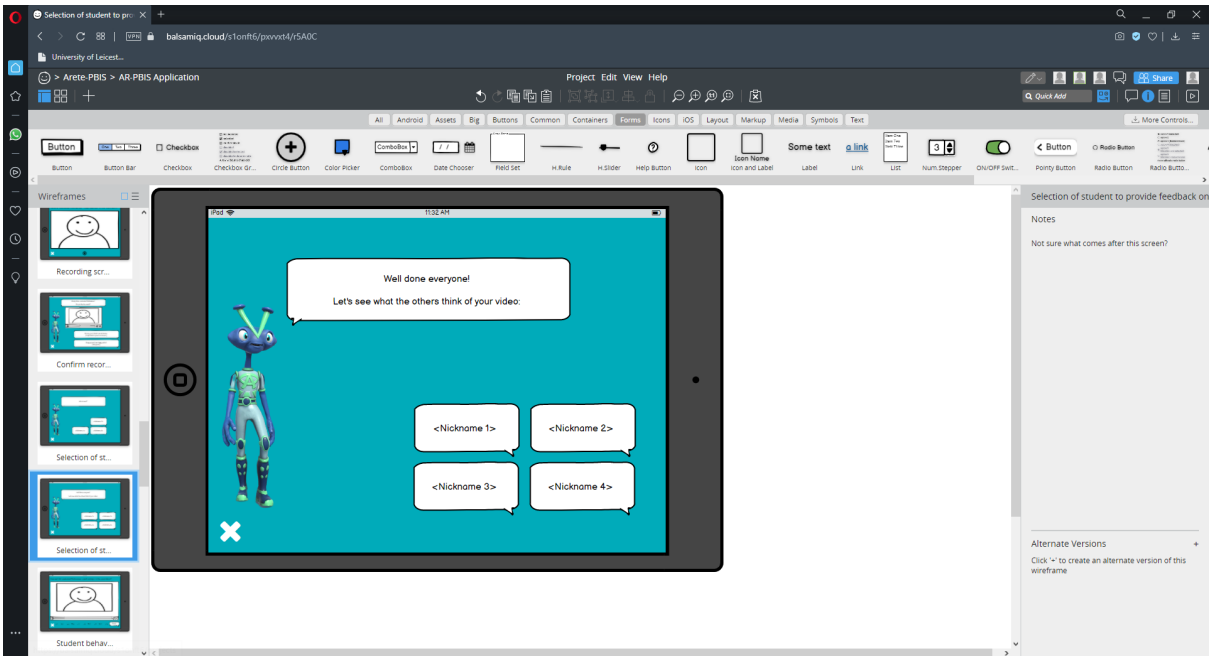


Figure 24: Select student to provide feedback for.

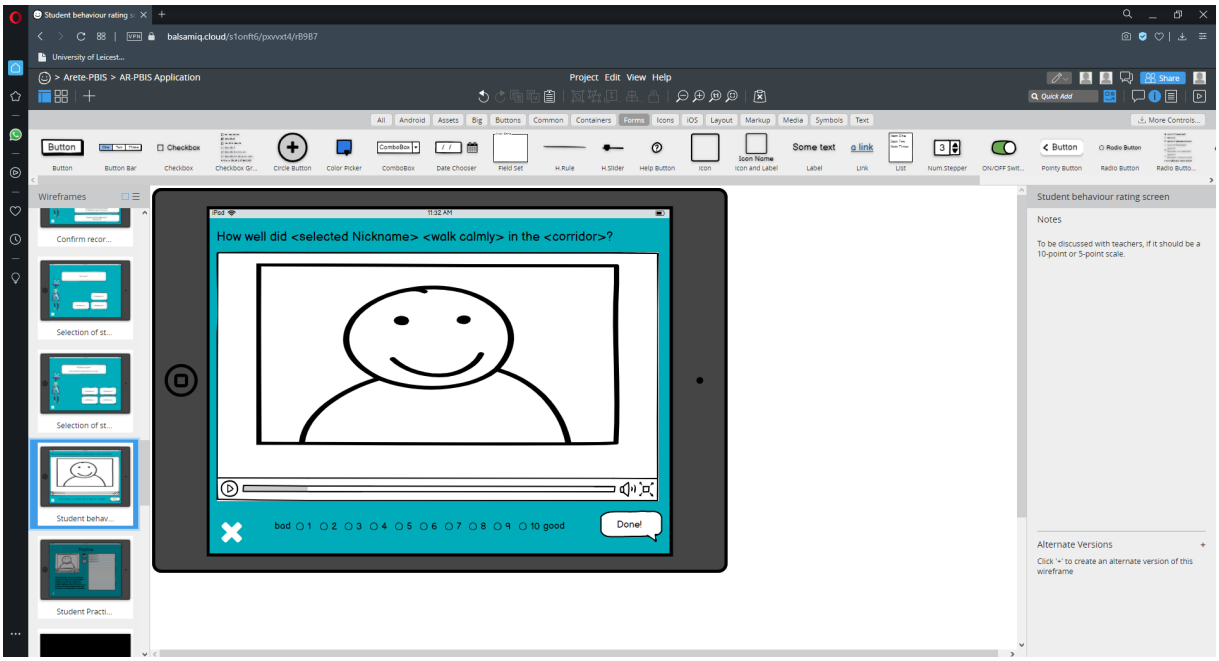


Figure 25: Student behaviour rating by group.