Some Experiences from Remote User Testing with Children

Camilla B. Løkkegaard[1], Frederik Lassen[1], Frederik Kok[1], Jacob Blaabjerg[1], Jonas Asbjørn Sørensen[1], Jonas S. Walsted[1], Lars Bo Larsen[1], and Thomas V. Snitker[2]

[1] Aalborg University, Aalborg, Denmark
[clakke17,fkok14,flasse16,jblaa17,jwalst,jasa17]@student.aau.dk, lbl@es.aau.dk
[2] LEGO Group, Denmark thomas.snitker@LEGO.com

Abstract. This study investigates the potential of asynchronous user testing of a physical product with children aged 9-12 years as the target users. Firstly we lay out the challenges of remote user testing with children. Following this, a prototype was developed in collaboration with LEGO UX specialists through an iterative design process which resulted in a functional prototype. The solution was well received by the children who participated in the tests carried out in the development. The paper reports the gathered experiences and findings of the final test and lists a number of recommendations for future remote testing with children.

Keywords: Remote User Testing· Children · Physical Products.

1 Introduction

It has been a well-established practise for many years to carry out user tests during product development and before product release. Companies test their products to gain feedback and insights and to observe, if their user interfaces and products are understood in the manner in which they are intended. Testing is often done through an iterative process cycle, where the products' design is examined through end-user trials. User tests are usually done in a controlled lab environment to create optimal conditions for data collection or in the field for optimal ecological validity. This, however, can be expensive in relation to resources and time. To solve this problem one approach is to test the products remotely. This method solves a number of logistic problems such as test user recruitment and the requirement that they should be physically and temporally co-located with the test team. Therefore, remote testing is considered to be cheaper than standard testing and more convenient for the participants. However, as the environmental settings are not under tight control, this can impact the results.

Furthermore, this approach has so far only been applied for software products, such as websites and smartphone apps. Physical products are typically more complex to test due to the difficulty of observing the various interactions

the participant may have with the products. Furthermore, as software products are very easy to prototype and deploy remotely, physical products present more challenges. This paper will examine these two aspects: Remote user testing of a physical product.

2 Remote usability testing

It has become a more widely used method to do remote usability testing (RUT) in recent years. This is evident though the emergence of a number of RUT platforms, such as Preely, Maze, UserTesting and UseBerry [1] [2] [3] [4].

It is an efficient way to collect data from a products' target group in a natural environment [5]. RUT is typically divided into two different categories; Synchronous (SRUT) and Asynchronous (ARUT). The difference being that in SRUT, the user and the moderator are separated in space, but not in time. Contrarily, in ARUT, they are fully separated in both space and time.

One of the main drawbacks of the ARUT method is that the method lacks a good method to collect qualitative and observational data since the moderator cannot observe the interaction between the person and the product. That said, the same studies found that remote testing and lab testing, in general lead to similar results regarding locating general usability issues [6] [7]. The same overall results can be found in a study comparing lab-testing with SRUT [8]. They found no difference in regard to subjective ratings, stress levels or the quality of the usability results.

Another study also concluded from their experiment that asynchronous testing of a website was of comparative if not higher quality than in lab-testing regarding finding usability problems [9]. Furthermore, the study showed that the ARUT method is a lot more cost effective due to, amongst other things, moving the work from the UX researchers to the users and extend the amount of suitable test users. However, other studies found that the ARUT method requires less time from the researcher but was more time consuming for the participants and identified fewer usability problems compared to both classic lab-testing and SRUT, which concluded that the method identified fewer usability problems, but also required a reduced effort, making it an attractive method for a cheap, although not complete, usability test [5] [10].

2.1 Testing physical products

The present study is set apart from the studies mentioned above in that it concerns remote testing of physical products where no published studies have been found to date. Many of the considerations and benefits when testing software products remotely are similar to testing psychical products remotely. It is potentially easier to collect a larger sample and test people despite large geographical distances, but it might also prove more difficult to keep as many factors constant

as would be the case in a physical setting, to mention a few [5]. The most significant differences might be the preceding distribution of the physical product, which evidently create logistic challenges. Another requirement is that the product (or a prototype) must have reached a certain level of maturity to be testable without on-site support, effectively ruling out early or paper prototypes. Furthermore, it is more straightforward to observe the interaction with the help of screen recorders and online usability testing tools when testing software products.

Video recordings are needed to capture the interaction with the product, when testing a physical product. Furthermore, when conducting ARUT, there is a need for a manual or instructions to guide the flow of the test or a platform that can act as a interactive moderator which, for example, can describe test tasks, collect questionnaire answers, record video, and might also provide responses and instructions depending on input from the user.

It is not sufficient to only measure usability metrics and test whether it is easy to use or not, when evaluating a product such as a LEGO set meant for play and entertainment. You need to include another dimension to measure the user experience and observe, if the product brings the intended experience to the user [11]. To collect this type of data in a remote testing scenario, video recordings of the user interacting and playing with the product are beneficial. This would make it possible to observe both verbal and non-verbal reactions from the user which could lead to valuable insights in regard to measuring the user experience [12].

3 User testing with children

When designing an ARUT testing platform for children, it is imperative to find the right balance between the efficiency of collecting user data and a fun and child-friendly design that makes it both fun and worthwhile for the child participating in the study. When doing user testing with children, one study has shown that the power balance between moderator and the child, which can otherwise prevent the child from acting naturally and speaking freely, is lowered when using SRUT compared to lab-testing [13]. When using an ARUT platform, the moderator is no longer present. We propose a combination of the child's parent or guardian and an interactive platform to act as a moderator that might help the child feel safe and relaxed.

When testing with children, recruitment of participants, task formulation and experiment design are factors to be extra aware of. It is essential to define a narrow age group when testing with children since there is a significant difference in child development within only a few years [14]. Consequently, it is not feasible going for a design that will work both for a four-year-old and a ten-year-old. The tasks also need to be simple, described with a child-friendly language, and in general the whole study design needs to be designed with children in mind [14].

To sum up, the design of a remote evaluation study of a physical product intended for entertainment must include both elements of observation and self-reporting to help collect data on interaction elements that either cannot be observed or are not reported. When the target group are children, special attention must be paid to the design of the experiment, language and engagement of the test users. The following sections documents the development of a prototype platform aimed towards remote testing children's experiences with interacting with a LEGO set.

4 Designing for children

Numerous challenges arise when the target group is children regarding usability and user experience. Most notably choosing the right tone for the language. It is important to adapt commonly used research questions and scales to appeal to the younger audience [15]. Comparably, a child-friendly platform for remote usability testing has some similarities with educational programs for children. Both need to appeal to children, explain the program's functionalities adequately, ask questions, and receive answers in a more fun and playful way than with adults. Due to this, we interviewed two elementary school teachers because of their insights into children's needs, their reading capabilities and their behaviour with computer-based learning. It became clear that an ARUT platform needs an overarching theme and a host or Avatar to deliver relevant information.

Fig. 1. The Five Degrees of Happiness scale developed for children for preference tests [16].

The platform's theme was chosen by creating 12 different theme templates and tasking children in the target group to sort the themes on a Five Degrees of Happiness scale shown in Figure 1 [16]. The preferred theme and avatar (the wizard Merlin) is illustrated in Figure 3 below.

Although the think-aloud protocol is a hallmark research method for usability studies it is not compatible with younger subjects [17] [18]. To express opinions and thoughts while using a product can induce a too high cognitive load for children. Furthermore, children are usually exposed to questions with a correct or wrong answer and are usually not asked to express their opinion freely [19]. This may cause children to not say anything at all in fear of giving a wrong answer.

Our platform aims to work around these limitations by having a parent help the child through the questions and having Merlin give instructions to show what they have built.

It is vital for the experiment that children are involved in the process from start to finish. When asking children to participate in usability testing, one needs to consider the children as ordinary people [20]. Further backed by interviews with LEGO UX researchers, the children are experts in their own lives and should be regarded as such. For example, in addition to the parent/guardian signing the consent statement, the child should personally be asked if they wish to participate and agree to be filmed during the test.

5 Platform design

The platform is designed with the overall purpose of guiding the child and parent through a usability test, where the parent to some extent will act as a test moderator. The chosen use case for the platform is a scenario where a child aged from 9 to 12 years complete a usability test while building a course with the LEGO Super Mario Starter Course[3]. LEGO Super Mario is an interactive toy for children, including an appertaining application, providing different building tasks.

Furthermore, the platform instructs the participants to show what they have built to the camera after each building session (usually the built-in laptop camera). After the last building session, they are asked to answer a short series of questions intended for evaluating their experience with the product. This evaluation will then afterwards be available for the researcher to provide insights into how the product is being perceived and used by the target group.

The design of the platform was run as an iterative, user-centred design process with focus on creating an intuitive and understandable design for children. In total, seven iterations were undertaken, with a gradual refinement and shift from a supervised test to an unsupervised one. The platform's GUI and instructions to the child were updated throughout the development based on feedback from LEGO UX experts and/or end user trials after each iteration. The process also included materials and texts such as the post test questions and the written invitations sent to the child and parent to recruit them to the test.

The platform prototype has been produced using Axure RP, which is a rapid wireframing tool. The choice is based on the requirement of getting access to the webcam through the browser and handling text inputs from the user, which is supported by Axure RP [21].

In total, the prototype comprises 29 GUI pages. These are separated into five categories; Welcome, Demographics, How to, Playing with the LEGO*, Questions* and Ending. The categories marked with (*) will need to change depending

[3] Find the product at `https://www.LEGO.com/en-gb/product/adventures-with-mario-starter-course-71360`

on the specific product. Each iteration was tested by three to five children from the target group. Thus, a total 29 of participants were part of the development.

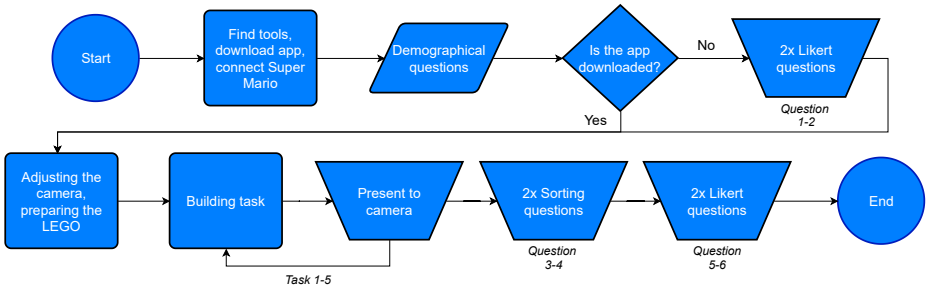

Fig. 2. The flowchart illustrates the flow of the user test. The questions are listed below.

The platform implements the flow shown in Figure 2. This illustrates the process of the user trial including the sequence of the different steps of building in the prototype. Note that a number of tasks deals with instructions for the setup, rather than the test itself. The different evaluation questions are also shown.

The tasks, translated from Danish, are as follows:
In-between the concrete LEGO tasks the child will be asked two sorting questions and two scale questions, as illustrated in Figure 2. The questions (translated from Danish) for the Super Mario case were:

1. Do you think your classmates will think LEGO Super Mario is fun?
2. How well do you know Super Mario?
3. Where would you prefer playing with the LEGO Super Mario?
4. With whom would you prefer to play with the LEGO Super Mario?
5. What do you think of playing with the LEGO Super Mario?
6. What do you think of playing with Super Mario on your own, built course?

Some GUI examples are seen in Figure 3, 4 and 5. These illustrations are chosen to give insight into some of the key features of the prototype and illustrate some of the mentioned tasks and questions. Common design features for all of them are that Merlin addresses the child with instructions and questions. This is designed in a way that aims to make the child feel like Merlin's valuable helper.

Figure 3 shows a page where the child have to find a set of tools to be used for the usability test. This page is designed using progressive disclosure, where only one item is shown from the start. The child clicks on the item when they have found it. Afterwards they receive feedback from the red check mark and a new item pops up. This feature adds some interactivity to make it more playful

Fig. 3. The necessary tools for the trial are shown using progressive disclosure

Fig. 4. Merlin shows the child how to adjust the camera

and is also designed to make sure that every item is noticed and found before continuing.

Figure 4 shows the page where the camera should be adjusted so the lens captures both the child and the LEGO at the same time. A picture of Merlin with the LEGO is shown to resemble how it should look when set up correctly.

Fig. 5. Merlin asks where she would prefer to play. The child rates the locations by placing them according to the five degrees of happiness scale.

After completing the tasks the platform will move on to evaluation questions. The child will be presented with two types of questions, where Figure 5 illustrates the first type where the child has to order pictograms representing locations they would rather play with LEGO Super Mario. Drag-and-drop interactivity is added to the pictures to relate the task to sort things in the psychical world. The idea behind the design was to get insight into a comparative fun aspect. Experiences have shown this can be a valuable method to include in addition to the absolute measures. This form is sometimes not so informative when judged by children [22].

6 Asynchronous user test of the final version

We conducted a user study to test the platform's ability to facilitate asynchronous remote user testing with children, including ease-of-use and engagement.

Even though LEGO Super Mario is the product the users are instructed to play with in the study, the focus of the experiment is not on observed usability problems with LEGO Super Mario, but instead if and how the platform is capable of collecting data regarding these usability problems. As mentioned in section 5 qualitative data are collected through the video recordings and quantitative data are collected via the responses using the built-in scales mentioned above.

We set up a number of success criteria based on the insights gained from the previous iterations. The success criteria addressed: mis-clicks (whether the child could reliably identify the interactive parts of the GUIs); whether the camera were set up correctly; if the concepts, questions and scales are understood; if the child understood the progressive disclosure scheme; and finally, if the child responded they had a good time during the session.

A post-test survey was conducted immediately following the user study to capture the experience.

6.1 Participants

Eight participants were recruited to participate in the user study. The participants were recruited by contacting their parents who, on behalf of their children, expressed interest in participating in the user study. Afterwards the LEGO set was delivered to their address one of the authors

Next, an email was sent to the parents, which included all relevant information regarding participation in the user study and was divided into two parts. The first part addressed the child and was illustrated with the wizard host, Merlin. In this part, Merlin explains directly to the participant that he needs their help with testing LEGO Super Mario. This narrative is included in the email to increase the participants' interest and willingness to participate by letting them know that they will be helping Merlin, which aims to make a meaningful participation.

The second part of the email addressed the participating parents and contained a brief introduction to the project and a description of the user study and the necessary items required to participate. In addition, this part of the email also contained the statement of consent, where both the child and parent were informed that if the child changed their mind and did not want to participate anymore, they were free to end the experiment at any time. The email also contained an invite to the online meeting where the user study was supposed to start.

All eight participants were between 9-12 years old. Their gender and time of day in which the experiment took place are listed in Table 1.

6.2 Test procedure

Microsoft Teams was used to collect video and share screens for the test, which meant that both the interaction with the prototype and LEGO Super Mario

Participant number	Age	Gender	Time of day
1	11	Boy	3:00 PM
2	9	Boy	4:00 PM
3	10	Boy	6:00 PM
4	12	Boy	8:00 PM
6	10	Boy	9:00 AM
7	11	Girl	10:00 AM
8	9	Boy	3:00 PM

Table 1. Table over demographic parameters on the participants.

could be recorded. Due to our experience from the preliminary tests, we also decided to give an introduction to the experiment to minimise the risk of other potential technical issues.

After the introduction the moderator left the Teams-meeting, and the participants and their parents were left to fully autonomously follow the instructions from the platform. However, as a backup, the participants could contact the moderator on the phone, if some unsolvable problem occurred.

7 Findings

All of the participants successfully completed the test. However, one participant mistakenly closed the video recording, and is therefore excluded from the these findings. The experiment took on average of 82 minutes to complete, with a minimum of 53 minutes and a maximum of 114 minutes. During this, no participants mentioned fatigue or boredom.
The option to call the moderator was used by two participants to fix technical issues. The problems were solved in cooperation with the moderator.

7.1 Post-test survey

For the first question, Q1; "Did you have any doubts regarding the instructions in the program?", the participants answers were overall positive. One had doubts about where to press to start the experiment. One participant said the instructions often led them to believe they were behind in the test. The six remaining participants did not report any doubts.

The second question (Q2); "How easy was it to follow the instructions in the program?" was answered using the Five Degrees of Happiness scale. The answers can be seen in Table 2. Given that all participants either answered the most or second happiest Smiley, all felt confident following the instructions.

When answering Q3; "How was it to control both the program, the application and LEGO Super Mario simultaneously?", most of the participants reported no issues managing the toy, application and platform. Three expressed some difficulties without elaborating further, two of which noted that they still enjoyed themselves.

To Q4; "How was it to present in front of the camera?", five of the participants said they where either comfortable, it was cool or fun and different. One parent noted the child thought it was fun and were not as shy as the parent had expected. Shyness and strangeness was the answer for the remaining three participants, though no one expressed worse discomfort.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**
Q2	1	1	2	1	2	1	1	2
Q6	1	2	2	2	2	3	1	1
Q8	1	2	1	1	1	1	1	2
Q9	1	1	2	1	2	1	1	2
Q10	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 2. Post test scores on the Five Degrees of Happiness scale. A low number indicates a positive attitude. The questions are listed in the text

Regarding Q5; "Did you think that Merlin was of assistance during the playtest?", five participants answered yes. However, one participant said they were distracted by LEGO which caused them to forget about him and follow the app. One participant answered that they did not use him, and one did not find him useful.

For Q6; "How did you feel about the way Merlin communicated during the playtest?", the answers were mostly positive with one giving it a "3". The answers can be seen in Table 2.

To Q7; "How did you understand the reference to TikTok and YouTube?", half the participants said they understood the meaning and that it helped them talk to the camera. Three participants had trouble understanding the reference or the meaning of it, while one participant opted not to answer at all.

Six gave the highest score while two participants scored second to Q8; "Which experience did you have by participating in the playtest?".

To Q9; "Did you understand the questions that were asked during the playtest?", five scored highest and three second.

Q10; "Do you think some of your friends would enjoy participating in the playtest?" was a binary answer with "Yes" and "No" as the options. All participants answered "Yes". The answers can be seen in Table 2.

The final question is Q11; "Who do you think would enjoy participating in the experiment the most?". Here, the participants were asked to rank *Someone younger*, *Same age* and *Someone older*. The top answers were primarily *Same age* and *Someone younger*.

These answers, along with observations from the video material, will be used to asses the success criteria.

7.2 Success criteria

The platform's purpose was to facilitate an ARUT experiment where the user was engaged and the platform was easy to use. The video recordings were analysed by the authors, with the purpose to identify usability problems and look for confirmations or violations of the success criteria. These are presented together with any usability problems occurring in connection with the particular feature.

Firstly, none of the participants mistook non-clickable objects as buttons. Therefore criteria 1 will be evaluated as "Success". There were, however, some usability problems regarding clicking. One participant was assisted by the parent. This parent was controlling the platform, which resulted in the parent skipping many pages so the participant could not process all of the messages Merlin said. One of the participants chose to go to the next pages in the prototype when their technical difficulties occurred to see if there was some helpful information. These clicks were noted as usability problems, not to infer whether it was a deliberate or miss click action.

Every participant except one was successful in adjusting the camera, so its field of view was allowed to record the participant and their LEGO creations resulting in the criteria being evaluated as "Acceptable". However, one participant was too shy to inform about their creation. One participant moved their computer, so the camera's field of view made it difficult to see the interaction between the participant and their LEGO. It is crucial to have the camera adjusted correctly, and the most significant factor contributing to this success was the parents. In five out of seven of the experiments, the parents adjusted the camera.

The other clickable elements, such as the scales, the tool list and the ranking task, was also examined to see if the design allowed the intended action. Every participant answered when asked about their demographic information. Every participant clicked the tools, and every participant understood they had to answer questions using the Five Degrees of Happiness scale when it was presented. The only interaction that was not successfully completed was the ranking task and only one participant did not complete this. This participant thought their answers should be done by clicking the card and then clicking the respective smiley. After rereading the task, the answer was given as it was intended in the design. All these features are used according to their intended function and this criteria is therefore accepted.

The platform refers to YouTube and TikTok to encourage the participant to show their LEGO creation to the camera. However, three of the eight participants did not understand the reference. One participant thought they could upload their video presentations to the mentioned websites, while others, in general were shy

and uncomfortable presenting in front of the camera. Therefore this criteria is not accepted.

8 Discussion

The findings from the user trial showed that seven of the eight success criteria were accepted which in general indicates that the prototype works as intended. This entails the participants were guided through a usability test successfully, the users were engaged and the platform was easy for them to use. The only rejected criteria was the one regarding the YouTube and TikTok references. Some problems with discomfort in front of the camera were reported. This is a general trend, that has been observed through all seven iterations. This may indicate that it can be difficult and awkward for more introverted children to show and tell about their creation. This also meant that the video recordings and presentations collected from the usability experiment varied in quality and duration.

On the other hand, some participants showed no discomfort, resulting in well-performed presentations. In addition, these participants also indicated in both the video recordings and the post-test survey that they had a good time while presenting in front of the camera. Thus, our results indicate, it will be essential to consider the children's personality when recruiting and accept that the shyness, and the openness will vary across participants, when you are working with children.

The introduction by the moderator (to ensure all the technical aspects of the experiment were working) can have affected the participants and their view on the experiment. One could suspect that it drew extra attention to the fact that the participants were recorded and that a moderator monitored their actions during the test. This could potentially affect their performance and emotions during the experiment, while *demand characteristics* could also occur. In addition, this could mean that some participants may be motivated to please the moderator, based on their expectations of the experiment.

Despite this, it is still assessed that the experiment gave insight into how the platform could facilitate an asynchronous remote usability test because the participants and their parents were on their own during the experiment.

Another potential bias towards the prototype comes from the product itself. It is speculated that the children could have difficulties differentiating between their opinion regarding the prototype and the toy.

This could mean that the participants' answer to the question regarding the overall experience, could be based solely on the experience they had playing with LEGO Super Mario. However this bias is hard to solve, due to the nature of the platform. If the children are having a good time while participating, the question of whether it is the platform or the LEGO that is the cause of the good

experience, may not be relevant. The platform should, at least maintain their positive experience, which it seems like is the result in this experiment.
The results from the post-test survey showed that all seven subjects expressed an excellent experience by ranking the question about their experience with the highest possible scale value.

It could be argued that this is because of the LEGO brand itself and because of the fact that the participants were allowed to keep the package afterwards as a gift. LEGO Super Mario has a suggested retail price of DKK 549 (€ 74), and the value may have influenced the results. The children may consciously or subconsciously have felt influenced to give more positive answers. This argument is supported by the responses in the post-test survey, where the majority of the questions were answered with the highest or second-highest grade. Another reason why the subjects have replied that they have had a good experience may be due to the fact that LEGO is considered by many as an attractive and expensive brand with a good reputation, and the subjects have rated the product higher simply because it is of this particular brand.

Due to these considerations, acceptance of the criteria regarding the users having a pleasant time was based on both the responses and the analysed videos. Both sources support the conclusion that the participants, in general, were guided by the platform through the usability test successfully and were engaged and entertained during the experiment.

Regarding the experiment's internal validity this could be affected by the location of the experiment and the surrounding environment. The experiments were conducted in different locations with different environments, where it was noted that the locations were varying across the participants where some experiments were conducted in a kitchen, or in the living room, while some were conducted in the participant's own room. These differences across the participants' locations also change the preconditions under which they conduct the experiments, which could influence the internal validity.

We also observed that assistance from the parent varied across the participants. Some parents were passive, while others were more active and instructed the participants during the experiment. While the participant's personality should be taken into consideration, so should their parent's personality. In addition, a difference in how often the participants sought assistance from their parent was observed. LEGO's UX researchers also encounters this problem, when doing situated tests.

These issues are considered difficult to control or prevent, and the recommendation is to bear in mind it may influence test results.

Our assumption is that the natural setting strengthens the ecological validity of the experiment as the location and the interaction with their parents are comparable to a natural playing situation. Therefore, the differences in locations and the parent's role may be considered to be an advantage for the ARUT paradigm, but care must be taken when analysing results.

The external validity is compromised regarding the recruitment of test participants. These are all located in Denmark, in physical closeness to Aalborg University. The sampling strategy was chosen for logistic reasons, because the LEGO Super Mario sets had to be delivered by the authors, which resulted in the selection of a convenience sample. However, it can be argued that the external validity is not the main concern of this paper due the limited sample size, but if a future final version of the platform was to be tested, the sampling of participants must be a lot more diverse.

Despite the limitations in both the experiment and platform, the results and findings are believed to be indicative to a high extent due to the agreement across the participants (at least those from a similar cultural background and reference set).

9 Conclusion

The purpose of this study was to demonstrate the feasibility of ARUT testing of physical products with children as target user group. To do this, we developed a platform that could facilitate asynchronous remote usability testing though a number of iterations. A literature review was done as well as interviews with elementary school teachers and employees from LEGO to determine the basis of the manner in which the platform should be designed. Based on this, a high-fidelity prototype was designed and developed using a wireframing tool. A proof-of-concept user trial was then conducted to collect empirical evidence. The design was tested on criteria that were based on the iterative design process as well as the literature review.

To summarise, it was possible to design a platform that can facilitate an asynchronous remote usability test. However, the collected video data vary in quality due to the participants' varying level if comfort presenting to the camera. Nevertheless, in the post-test survey the participants showed they had a great experience with the experiment. Despite these limitations and the need for further research, the potential of a platform for facilitating asynchronous usability testing with children is still considered valuable.

9.1 Future work

Further development and testing with a broader end user sample is necessary to eliminate the technical problems so the platform can run completely asynchronous and fully functional. Furthermore, an in-depth evaluation should be conducted with UX researchers from LEGO to gain insights into how well the platform fulfils their needs for data.

Lastly, the platform should be extended to easily allow for setup of new tests with different products.

10 Acknowledgement

We wish to thank all our test persons for their time and efforts, as well as the UX research team at LEGO for valuable feedback and suggestions. Finally we wish to thank LEGO for sponsoring the sets used for the tests.

References

1. Preely homepage, `https://preely.com`
2. Maze homepage: `https://maze.co/`
3. UserTesting homepage: `https://www.usertesting.com/`
4. UseBerry homepage: `https://www.useberry.com/`
5. Andreasen, M. et al. (2007) 'What happened to remote usability testing?: an empirical study of three methods', in Proceedings of the SIGCHI Conference on human factors in computing systems. [Online]. 2007 ACM. pp. 1405–1414.
6. Bastien, J.. (2009). 'Usability testing: A review of some methodological and technical aspects of the method', International journal of medical informatics. 79. e18-23. 10.1016/j.ijmedinf.2008.12.004.
7. McFadden, E. et al. (2002) 'Remote usability evaluation: Overview and case studies', International journal of human-computer interaction. [Online] 14 (3-4), pp. 489–502.
8. Andrzejczak, C. & Liu, D. (2010) 'The effect of testing location on usability testing performance, participant stress levels, and subjective testing experience', The Journal of systems and software. [Online] 83 (7), pp. 1258–1266.
9. Martin, R. & Shamari, M. & Seliaman, Mohamed & Mayhew, P.J. (2014). 'Remote Asynchronous Testing: A Cost-Effective Alternative for Website Usability Evaluation', International Journal of Computer and Information Technology. 03.
10. Bruun, A. et al. (2009) 'Let your users do the testing: a comparison of three remote asynchronous usability testing methods', in Proceedings of the SIGCHI Conference on human factors in computing systems. [Online]. 2009 ACM. pp. 1619–1628.
11. Jordan, P. W. (2000) 'Designing Pleasurable Products: An Introduction to the New Human Factors', Independence: CRC Press LLC, pp. 1-10.
12. Tullis, T. and Albert, W. 'Measuring the User Experience, Second Edition: Collecting, Analyzing, and Presenting Usability Metrics', (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2nd edition 2013.) Chapter 7.
13. Jønck, Line (2021) 'Does changing consumer testing platforms, change the product perception?', Master's thesis, Southern University of Denmark, Odense.
14. Alita Joyce (2019) Usability Testing with Minors: 16 Tips. (`https://www.nngroup.com/articles/usability-testing-minors/`).
15. Panos Markopoulos and Janet Read and Stuart MacFarlane and Johanna Höysniemi, 'Survey methods', in *Evaluating children's interactive products principles and practices for interaction designers*, 1st edition, Morgan Kaufmann, pp. 234-255.

16. Hall, L. et al. (2016) 'Five Degrees of happiness: Effective Smiley Face Likert scales for evaluating with children', in Proceedings of IDC 2016 - The 15th International Conference on Interaction Design and Children. [Online]. 2016 pp. 311–321.

17. Panos Markopoulos and Janet Read and Stuart MacFarlane and Johanna Höysniemi, 'What is a child', in *Evaluating children's interactive products principles and practices for interaction designers*, 1st edition, Morgan Kaufmann, pp. 1-18.

18. Panos Markopoulos and Janet Read and Stuart MacFarlane and Johanna Höysniemi, 'Children and interactive technology', in *Evaluating children's interactive products principles and practices for interaction designers*, 1st edition, Morgan Kaufmann, pp. 20-33.

19. Panos Markopoulos and Janet Read and Stuart MacFarlane and Johanna Höysniemi, 'Planning the evaluation study', in *Evaluating children's interactive products principles and practices for interaction designers*, 1st edition, Morgan Kaufmann, pp. 68-89.

20. Panos Markopoulos and Janet Read and Stuart MacFarlane and Johanna Höysniemi, 'Ethical practise in evaluations', in *Evaluating children's interactive products principles and practices for interaction designers*, 1st edition, Morgan Kaufmann, pp. 49-66.

21. Axure homepage, `https://www.axure.com/`

22. Read, Janet and MacFarlane, Stuart and Casey, Christopher. (2009). 'Endurability, Engagement and Expectations: Measuring Children's Fun'. Interaction Design and Children.