Tampere University of Applied Sciences

Remote User Testing for Primary School Children

With Yle Galaxi Application

Noora Nenonen

BACHELOR'S THESIS
February 2022

Degree Programme in Media and Arts
Interactive Media

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
Degree Programme in Media and Arts
Interactive Media

NENONEN, NOORA:
Remote User Testing for Primary School Children
With Yle Galaxi Application

Bachelor's thesis 39 pages, of which appendices are 2 pages
February 2022

The goal of this thesis was to research and validate remote user testing possibilities and how to adapt them to children aged from 7 to 12 years old. The research data was collected from literary sources, expert interviews, and user testing. The thesis consists of a theoretical and a practical part.

The theoretical part studies the purpose of testing in the user experience design field as well as explores the best methodologies of conducting them. The concept of children as users and their developmental stages in relation to user testing were explored and compared to adult user behaviour.

The demand for remote online solutions expanded in 2020 during the COVID-19 pandemic and this thesis aimed to find solutions for effective remote research and testing, specifically with primary school children.

For the practical part a small-scale usability testing session was conducted via Google Meet with a group of six children from Ylöjärvi, Finland. In the user testing the children performed tasks on Yle Galaxi application. The purpose of the testing was not to study the application itself but to try out remote usability testing on a free online meeting platform.

The findings of the testing were useful and educational. Despite a few problems with the tools used for the testing, it was a successful try-out for remote usability testing. All in all, further development for remote usability testing methods is needed. For this thesis there were no resources to try out commercial user testing websites or professional test participant recruiting agencies. Despite the limitations in budget, the testing proved that remote usability testing is feasible method for user research. It also proved that it is possible to conduct such tests with children.

Key words: user experience design, interaction design, children, remote research, remote testing

CONTENTS

1 INTRODUCTION

This thesis aims to research the remote usability testing possibilities with primary school children of ages 7–12. The goal is to further understand how children are as users as well as to discover the latest tools for remote testing and how to successfully set up testing sessions with children.

My reason for choosing this topic is my close involvement in children's application development and personal interest in creative user experience design solutions. Having been employed at Yleisradio Oy (Yle) children's department Galaxi at Tampere since spring 2020 my main tasks were user experience (UX) design and user interface development as well as conducting user tests with the target group children. Galaxi is a tv-show, social media and mobile application conglomerate produced by the national broadcasting company Yle for Finnish primary school children.

I started my job just before the COVID-19 pandemic hit and we had just started concepting a new application for Galaxi. I was able to conduct some in-person UX testing sessions until May 2020. Since then, the world went fully remote and so did we. I had to do a lot of research and planning in terms what was even possible to test remotely. This thesis consists of partly theory and partly my own experiences in remote testing, and an expert interview with my co-workers who also work closely with children.

The practical part of this thesis is a small-scale experiment of a remote usability testing situation with a group of 11-year-old children. For this testing I used the current Yle Galaxi application, available on app stores. Before the making of this thesis, I had done plenty of focus groups and A/B testing online, but I had never attempted usability testing remotely, with the children completing actual tasks on the app while I moderated.

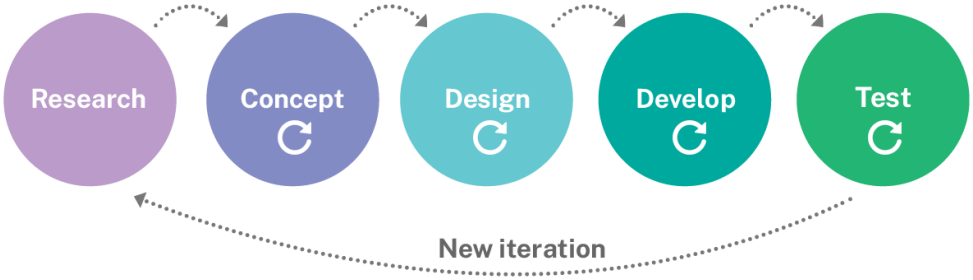
2 USABILITY AND USER TESTING

In this chapter some of the basic concepts of user experience design (UX) and user interface design (UI) are explained in relation to user testing. Oxford Learner's Advanced Dictionary defines usability as "the quality or state of being easy to use" (Oxford Learner's Dictionary n.d.). In this thesis I will talk about mostly about usability (see 2.1.1) testing and focus groups. The key difference between the two is that focus groups observes people's opinions where usability testing focuses on observing people's behaviour. (Bolt, Tulathimutte 2010, 3.)

There are many approaches such as user-centered design in the world of UX. Different methods have been developed by different designers, but they share similar principals such as emphasizing with the user, doing many iterative design rounds, and user testing early on and often. These principals can be applied to a wide age-spectrum of users, but the chapter 3 of this thesis focuses more on the aspect of how to modify them to fit user tests with children. They can also be generally applied to both in-person and remote testing. I consider user testing to be an umbrella term for all sort of testing done with users in UX.

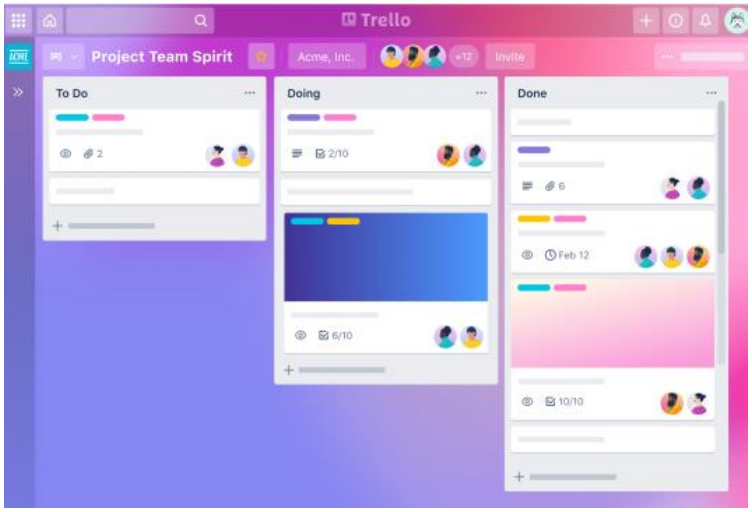
2.1 User-centered design

One of the key methodologies of UX is user-centered design (UCD). It is a design approach that drives to take the person and end-user into account since the beginning and throughout the design process. In UCD method designers involve the users in the design process via a variety of research and testing techniques, to ensure the best usability and accessible products for them (Interaction Design Foundation, n.d.). The term UCD or human-centered design as it is often called, was coined by Rob Kling in 1977 and later cultivated by Donald Norman of Nielsen Norman Group who made the concept gain popularity and attention with his book The Design of Everyday Things. Norman emphasizes the start point for good design are the needs of the users and considers e.g., aesthetics a secondary issue. Knowing your user is important and assumptions should be avoided. The second step should be finding the right problem and the right solution for it. (Norman 2013.)



PICTURE 1. The user-centered design process. (Nenonen 2022; Spring 2 Innovation 2019)

The method is outlined in Picture 1. First comes the research phase (contextual enquiry, user interviews, benchmarking) that is followed by the concept stage (building the concept from research and business insights, emphasizing with the user, ideating prototypes to design). The design stage includes sketching, prototyping, and finalizing the design in increasing fidelity as it moves through to development for coding and other implantation. Though testing stage is placed last it is usually applied as early as the first prototypes are finished, then repeated in each iteration, if possible, until the product is finished. (Norman, 2013; Moule, 2012.) After the phases starts a new iteration, with as many repeated phases as needed.



PICTURE 2. Trello application example page. (Trello n.d.)

Krug (2014) outlines his type of user-centered design as a base of web and app design. His own golden rule is to not make users think how to use the product –

the design should be intuitive and self-evident. He lists the use of conventions, effective visual hierarchies and eliminating distractions as important. (Krug 2014, 11, 29.) Norman (2013) mentions that the most essential elements to consider are accessibility, legibility, language, and visibility (Norman, 2013). Trello is a web- and mobile based project or personal management application. In Trello the use of these tips is evident. The page has simple, clean design (Picture 2). The use of visual hierarchies, the different colours marking different events and boxes highlighting them. The clickable elements are easily recognizable. The page is designed for skimming information, getting the needed information fast and effectively, with appearance customizing features.

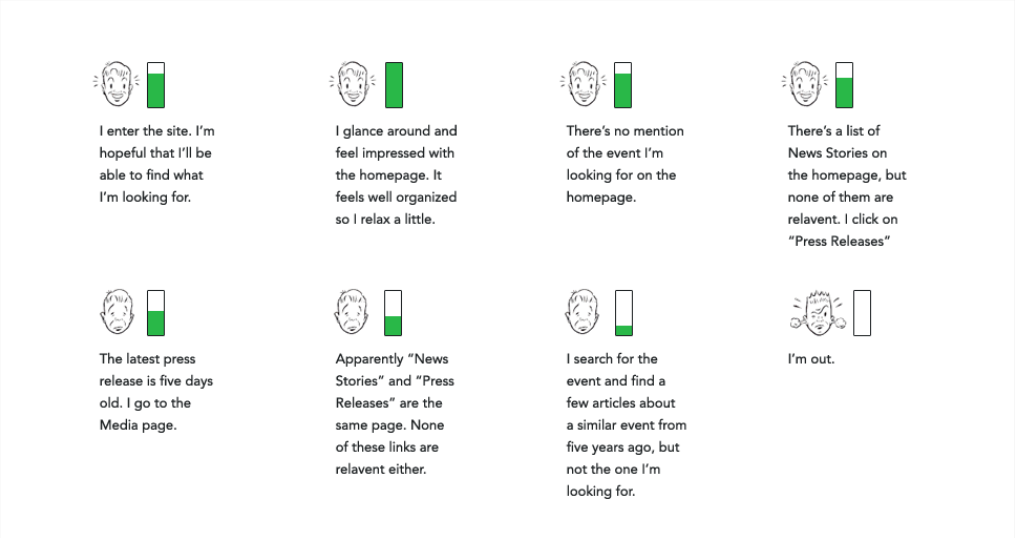
2.1.1 What is usability?

According to Jakob Nielsen (2012) usability is a standard attribute in UI design that assesses how easily used or functional user interfaces are. It is measured against five key components and design principals:

- Learnability: When first presented with the design, how easy it is for the user to accomplish tasks?
- Efficiency: Once familiar with the design, how quickly can the user perform the tasks?
- Memorability: When the user returns to the design after a period of not using it, how easily are the functions recalled?
- Errors: How many errors do users make, how severe are these errors, and how easily can they recover from the errors?
- Satisfaction: How pleasant is the user experience? (Nielsen 2012.)

In the world of interfaced products usability is a necessary attribute for survival. If an interface is difficult to use, people leave. There are plenty of other applications available; leaving is the most common act when users encounter a difficulty. (Nielsen 2012.) The term cognitive friction was coined by American software designer and programmer Alan Cooper. In his words cognitive friction is a psychological response that occurs in users when cues on a webpage or software don’t match expectations. When expectations aren’t met, the user halts,

attempting to make sense of what they have encountered. (Fike 2016.) According to Krug (2014, 166) everyone has a “reservoir of goodwill” when entering a site. That reservoir begins to drain, every time a moment of cognitive friction is encountered (Picture 3). (Krug 2014, 166.)



PICTURE 3, Example of user’s goodwill draining because of cognitive friction. (Krug 2014, 166)

The most effective way to avoid this phenomenon is to invest more in meeting the usability standards and components while designing. The best and most basic method to do that is user testing. (Nielsen 2012.)

2.2 What is user testing?

User testing and usability testing are terms often used interchangeably. Usability itself can be improved with many methods, but user testing is perhaps the most common and useful of them. (Nielsen 2012.) Testing is also the last phase of the user-centered design method. (Picture 1). Testing does not necessarily take place only at the end but all throughout the process in cycles. There are different fidelity levels in user testing to take in different stages of the design cycle. (Abras, Maloney-Krichmar & Preece 2004.)

Usability tests are essentially about watching participants trying to use the product (e.g., a UI prototype or an app) and do interactive tasks so the developers

can detect and fix the things that confuse, frustrate, or delight the users (Krug 2010, 13). The tasks in a usability test are simulating real activities that the participant might perform in their own life. Depending on the research questions and the type of user testing, they can be specific to the scope of the test or more open. Task wording and language play a part in how accurate the test results will be. Small errors or changes in the phrasing of a task can cause the user to misunderstand what they're asked to do or can influence how they perform the task, (Moran 2019.) Chapter 3.3.2 goes into depth why language choices are especially important with children, depending on their age.

Steve Krug in his book Don't Make Me Think (2014, 113) underlines the difference between focus group tests and usability testing. Focus groups are traditionally a marketing research tool, consisting of a small group of 5 to 10 people that are summoned to talk about their experiences, opinions, and reactions about the product. Focus groups are good for quickly getting a sense about the users' feelings, but require a larger quantity of participants, since opinions are more likely to vary widely across locations and demographics. They also don't usually reveal underlying problems about the actual user flow. (Bolt, Tulathimutte 2012, 4; Krug 2014, 113.)

One problem of usability testing is that it is often expensive and time consuming if done in a lab-environment, which has prompted the emergence of alternative testing techniques (Abras, et al., 2004). In 1989 at the third International Conference on Human-Computer Interaction in Boston UX guru Jakob Nielsen introduced his paper Usability Engineering at a Discount and it kickstarted the trend of discount usability testing. (Nielsen 2009.)

His paper introduced three main components to help UX designers working with smaller budgets. In 2009 he summarised them again:

- Simplified user testing, which includes a handful of participants, a focus on qualitative studies, and use of the thinking-aloud method.
- Low fidelity prototypes (e.g., paper, wireframes) that represents one problem needed to be solved at a time. This enables many iterative test rounds.

- Heuristic evaluation in which user interface designs are evaluated by UX professionals, who compare them to established usability guidelines. (Nielsen 2009.)

Krug (2010,14) also pioneers ‘do-it-yourself’ usability testing. According to Krug fast, crude, and simple user testing is effective in consistently revealing serious usability issues. (Krug 2010, 14.)

2.2.1 Why do user testing?

As stated in the previous chapter, the premise of user testing is simple: if one wishes to know whether a product is easy enough to use, one must observe the user while they try to use the product and note where they run into problems. (Krug 2014, 114.)

In mobile development, the end product, a user interface, is interactive and all the design choices directly affect the user experience. The earlier the users evaluate the designs of the product, the less likely it is the developers take a step back and rework them. Stepping in early and evaluating the design with end-users, understanding what it’s like for them to use the product before the final design or code is even considered will save a considerable amount of time and resources. (Moule 2012.) Usability testing should make a product more usable, involve actual users and real tasks, and generate results that testers can observe, record, and analyse. (Abras, et al., 2004).

“You are not your user”, states Moule (2019) in the first chapter of their book Killer UX. I think it summarises the need for user testing perfectly. As a designer, even when testing with adults, it is unadvisable to try to evaluate decisions just from one’s own or the designer team’s perspective. (Moule 2019.) When developing an interface for children it might be difficult to understand their perspective. Questions such as how they might use an interactive product or how differently they behave compared to adult users might occur. Designers might be tempted to consult experts in usability design, to do heuristic evaluation. The opinion of an

expert is an educated guess. It is sensible to involve real children in a user testing evaluation in at least one stage of the design process. (Markopoulos, Read, Macfarlane & Höysniemi 2008, 46.)

2.2.2 Prototypes

A prototype is a proof of concept for the design ideas during a project as well as a tool for user testing. Prototyping further identifies user needs and assists in defining the product's scope. This helps in risk-managing and uncovering possible problems early on. Prototypes are a great platform to showcase clients the design perspective. Prototypes enable designers to make the product concept tangible and real before moving onto the design- and development stages. (Moule 2012.)

A prototype can in its simplest form be just a wireframe or a paper cut-out. A wireframe is a schematic diagram of an app. It shows the placement of biggest building blocks of the UI such as buttons, menus, and logos. According to Krug (2010, 36, 64) the most important information to gain from wireframe testing is if things are placed where people expect to find them. Wireframes focus on interaction and confirming conventions users have about interfaces across different applications, such as placement of a back button or finding the home page. (Krug 2010, 36, 64.) When moving up in the process and fidelity the next step is to focus more on the visual design. A low fidelity prototype can be made manually or digitally. (Krug 2010, 37.) A paper prototype (Picture 4) provides the first look on how the planned components appear visually. It can be a crude cut-out with movable elements. By utilizing sticky notes or adhesive the prototype gains a level of interactivity for testing purposes.

PICTURE 4. A low fidelity paper prototype (Nenonen 2019)

A low fidelity prototype is the first guide for designers of the direction of the app visuals. In a testing situation it helps to see if the early concept design and visual elements have introduced any usability issues. It also allows a user to get a visual understanding of the product. As iterations go on the prototypes go up in fidelity. Digital, interactive prototypes that you can build in UI software such as Figma or Sketch are ideal for user testing, as they can quickly be adapted to the scope of the testing and can closely resemble the final appearance of the product in the end.

2.2.3 Roles

Depending on the scope of the testing situations it can take anywhere from two to ten people to fill the roles. A facilitator is the person conducting the test and guiding the user. An observer is a person who does not interact with the user but is observing the user's behaviour and making notes. Sometimes stakeholders (e.g., executives, investors) might want to be participate in the testing situation, too, to monitor how their product is received. (Moule 2012.) At Yle researcher Heta Mulari (2021) conducts user tests with children with two facilitators working directly with them, and another person filming the screens of the devices used (Mulari 2021). In my own work I have usually filled the role of facilitator. I have recruited a co-worker to observe and take notes while I moderate the children. It

is useful to have someone not actively participating in the testing as a silent observer. They might be able to see detail a facilitator might miss.

3 TESTING WITH CHILDREN

Nielsen and Sherwin (2019) in their research studies Children's UX: Usability Issues in Designing for Young People say that there is no such thing as "designing for children of all ages". There is a need to target narrower age groups when designing for children. (Nielsen, Sherwin 2019.) It is recommended that primary school children should be divided at least into two subgroups: early elementary kids (six to eight years old) and tweens (nine to twelve years old). Even in the range of primary school children of ages seven to twelve the changes are big year by year. (Fisher 2014.) In this thesis the focus is on children aged seven to twelve– the most typical ages of children in Finnish primary schools.

Much as when starting to develop a product for children, in a testing situation it is also important to consider the developmental stage of the children to some extent. Limitations in their physical and cognitive abilities restrict what kind of tasks on which kind of devices they are able to perform. The testing environment and learnt social behaviour might affect them despite the latter abilities. (Cantuni, 2020.) Ethical aspects such as consent, and privacy issues of testing should be considered the most of all when working with children. To effectively execute user tests with children it is pivotal to know how children behave as users. Children differ a lot from adults and testing effectively with them requires planning according to their needs in order to make the best out of the testing sessions and get useful insights. (Cantuni 2020.)

3.1 Children as users

Cantuni (2020) encourages to "think like a kid" when working with children. One of the biggest principles of UX is empathy (Cantuni 2020). According to executive producer Hanna-Mari Kauhanen (2021) at Yle Lapset, the logic of a child is different to an adult and an important point of view to examine. Children are honest and give direct feedback. It might be harder for them to communicate the issues they face as users with words, but physical expression with hands and fingers is more pronounced than with adults and easier to interpret. (Kauhanen 2021.) Constructivism – the idea that knowledge is constructed through

experiences – is an educational theory on how children learn. Most primary school children nowadays are native mobile application users, thus have empiric knowledge of basic UI concepts. They can identify what works and what does not. (Fisher 2014.) Children are more explorative than adults by nature. They are usually eager to find out things by themselves. (Cantuni 2020.)

Children are curious, playful, quick, and anything there and between. Just as in user testing with adults, the participants are individuals. Despite doing some screening when choosing participants, it is not possible to predict how anyone might react in a testing situation. I have found out that adaptability is one of the key attributes a facilitator can have when working with children. Mulari (2021) agrees that expecting to unexpectedly adapt to changing moods and diverting interest is regular when testing with children (Mulari 2021)

3.2 Development level of the target group

Developmental psychology can offer some tools for understanding the developmental level of the target group. However, according to Carla Fisher (2014) the speed and level of a child's development is influenced by many factors— gender, parenting, the number of siblings, environment, education, culture, and so on. Children might follow a generally similar path, but each individual is different. That is why psychological guidelines should not be followed nor applied too strictly in user experience design. (Fisher 2014.)

For example, gesture-control is a navigation element (tap, swipe, drag etc.) in touchscreen applications, and they are favoured in apps for children. Nowadays, children around three years old can already know how to use gesture-control instinctively, but their motoric developmental level does not allow for much accuracy or dexterity yet. (Cantuni 2020.) By age seven, children have learned most motor movements (walking, holding a pencil etc.) and start refining their skills. The way they develop speed and accuracy depends largely on how frequently they practice. A child that practices e.g., football progresses quicker than their less physically active counterpart. Playing organized sports can help develop additional skills and opportunities to practice coordination. (Fisher 2014.)

Swiss psychologist Jean Piaget (1896–1980) was the first to introduce and theorize the four stages of cognitive development in children. Before this point in history, kids were treated as smaller versions of adults. Piaget was one of the first to distinguish that how children think is not quite the same as grown-ups do. He concluded that children are not less intelligent than adults, they think differently. (Cantuni 2020.)

Piaget's four stages of cognitive development are

- Sensorimotor (birth to 18–24 months)
- Preoperational (18–24 months through age 7).
- Concrete operational (ages 7–12)
- Formal operational (ages 12 and up) (Cherry 2020.)

Focusing more on the children of primary school age, who are in the concrete operational stage. In this stage, considering cognitive and socioemotional development children are practicing emotional management, e.g., not getting angry while being asked to wait for something, having patience. Around ages of six to eight children are developing a sense of self and individuality– what they look like, what they like or dislike. (Fisher 2014.) Approaching teenage, tweens are staring to have a less egocentric view and begin to understand others' feelings and point of view (Cantuni 2020). They have increased ability to understand the motivations behind actions and the concept of moral, e.g., someone might feel one way but act another (Fisher 2014).

3.3 Issues to consider in a testing situation

Depending on the developmental stage of the child and outside of those, a few things should be considered when conducting user tests. In my own testing experience the biggest issues are the distractibility and difficulty of communication. The best way to overcome the issues is to gain experience with children. And to accept plans might change and have backup plans for situations.

3.3.1 Ethics

When testing with minors it is important to take special care of ethical aspects. It is important to regard the local legislations and laws especially in a commercial environment. My own work was done in the public sector, thus, I made sure to have detailed and written consent from the children's parents.

Children themselves must also have a right to decline to take part in the test– their wishes must be respected. If a child wants to drop out of a testing session before it is completed, they should be allowed to do so. It is important to ensure that the child is aware of this before the evaluation begins. (Markopoulos et. al 2008, 58.) Farrell (2005) lists two key questions that should be considered when planning research involving children: "Is the research worth doing?" and "Is the research explained clearly enough so the participants can make an informed decision whether they want to consent or refuse?". These questions stem from medical ethics, that urge to respect children's own view, values, and explanations. (Farrell 2005.)

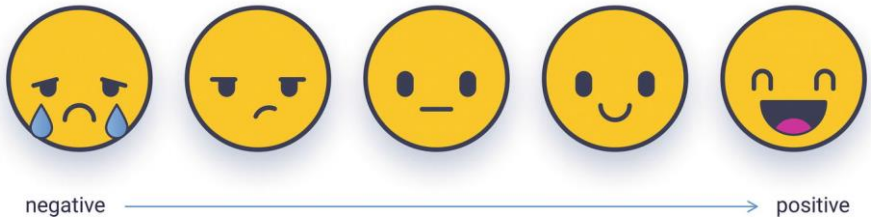
In the agreement to be photographed or recorded, it should clearly state where and how the material is going to be used. The same child that agreed to be recorded at age eight might not feel the same about the pictures or the sound of their voice being in a public work at age fifteen. (Cantuni 2020.) A good and common practice that exist and I have adapted is to blur out children's faces in the material. Collecting personal information is sometimes a must, for example, for sending out incentives. Along with carefully collecting sensitive data (full name, address, phone number etc.) it is also important to take care of it being securely archived or destroyed (Farrel 2005).

3.3.2 Language

Tailoring your language to children at different stages in their early education is a good practice. Cantuni (2020) says that it is important to have easily understandable written or verbal instructions, while not being patronizing at the same time (Cantuni 2020). For example, kids generally do not want to be

addressed or considered younger than their age (Lange 2016). The children I have worked with have verbally expressed that I should not call them ‘kids’ as they are not in grade X anymore.

Small changes in phrasing of the tasks and questions can affect the outcome of the testing (see 2.2). Asking the right questions can lead to better answers. Children are often experts at answering exactly what adults want to hear instead of what they think. This is also a sociopsychological phenomenon called the Hawthorne effect or observer effect. Its basic argument is that people change their behaviour and answers because they are being observed and not because of changes of the independent variables. In user testing this can result in faulty findings and thus can sabotage the development process (Bolt, Tulathimutte 2010, 120.) A good strategy to understandable answers from children who have trouble communicating verbally is using visual aids (Picture 5). They will help the child to visually express their feelings about the task without trying to find the right words. (Cantuni 2020.)



PICTURE 5, A visual aid/Likert scale for expressing emotions. (Cantuni 2020)

Using complex lingo or technical words is not sensible with the youngest children, but one should also not underestimate the level of knowledge either. For example, I once tested a pair of 10-year-old friends who knew a lot of UX terminology such as ‘end user’ and ‘wireframe’ and used them fluidly while testing a prototype UI. They expertly commented on the user flow and odd placement of the return button.

3.3.3 Behaviour and needs

The needs of a child and an adult vary. The needs of children, parents and teachers vary. According to Kauhanen (2021) the important thing is to empathize with the children, go to their level. To be approachable and kind. Remind them that they are the experts needed to make the product better. (Kauhanen 2021.) Mulari (2021) mentions the user tests should be carefully planned to fit the scope of the project, but also to fit the needs of the children. They should not be too long, and preferably interactive rather than speech-oriented. (Mulari 2020.)

In my own testing experience, I have noticed that it is easy to work with a school class or children that are familiar with each other. It has also been consistently preferable to let children form pairs with their closest friends and enter the testing situation together. Cantuni (2020) expresses that the “friendship pair” method works well in making the children open up and be less shy (Cantuni 2020). It adds another layer of security and I have found out that children tend to be more open and honest next to a friend that knows them and does not judge them for their answers. However, this might come at the cost of less controllability and insubordination. As a facilitator it is never easy to disentangle a situation where children are more interested in their own thing or have no respect for the rules of the testing situation such as a request to stay in their seats or to listen to the facilitator. In that case, it is good to remind them that they are here of their own will and if they cannot concentrate on the test now, they can participate next time.

The involvement of parents in a testing situation should be carefully considered, especially with primary school children. The parents might interject the child’s thoughts or overprompt them, even without meaning to. (Fisher 2014.) The parent should be instructed to be a silent observer, but they can also help explain or interpret the answers of their child (Cantuni 2020). Kauhanen (2021) and Mulari (2021) say that sometimes it is not possible to keep the parents from participating on some level and suggest giving the parents something to work with for the duration of the testing e.g., a questionnaire sheet. Most children are excited to be asked to participate. The testing is a ‘special’ moment out of the class, a new experience, and most consider talking about games and using their phones at school or kindergarten a thrilling thing. (Kauhanen, Mulari, 2021.) This is where

the incentive comes in, too– along with providing them class field trip funds we have sent them some Galaxi merchandise as a keepsake.

3.3.4 Location and environment

Some high-fidelity user tests might be conducted in laboratory conditions. A user testing laboratory is usually a “clinical” environment designed to simulate testing done in the world of scientific research, with separate rooms with a two-way mirror for the observers. (Bolt & Tulathimutte 2010.)

Often, laboratory testing is neither practical nor preferable with children. To ensure a safe and productive environment testing with children should take place in a familiar or a carefully planned, child-centered setting (Picture 6). A natural setting such as home or school can provide a natural picture of how children interact with the product. (Fisher 2014.) A quiet, separated room with child-friendly decoration such as colourful posters and appropriately sized furniture works well but being aware of not bringing in too many distractions, such as toys or music. In the end, it is as much about creating a safe space and environment through approachable actions as it is about the physical location- (Cantuni 2014.) Being approachable to children are things including but not limited to not dressing up too formally, not using formal language, cheery output, and facial expressions.

PICTURE 6, A user testing session for an interactive ball toy concept at a school. (Rodrigues, 2019)

I have held my testing sessions with the children mostly in empty classrooms. Sometimes located right above a noisy gymnasium, sometimes in a glass-walled computer room along a busy hallway. While doing remote testing, I have asked the teachers to secure a quiet, separated closet or a room for the children and the laptop via which we communicate. It is hard to control the location remotely. Yet, if need be, it is not impossible to conduct the tests in a less than ideal environment given the children still feel secure.

3.3.5 Reliability and validation

When testing children, it is important to consider the same standards of reliability as with adults. It is important to consider things such as if the test is always done the same way or are the facilitators the same. Much like in chapter 3.3.2 about language states even small variation in task phrasing might change the results. Some argue lab-environment user testing makes results and insights more reliable and validated because the specs of the testing can be recreated more or less identically each time (Bolt & Tulathimutte 2010). However, that does not equal the execution of the test being the same each time.

Sometimes with children there is no chance but to improvise on the fly (Mulari 2021). Krug (2010, 14) reminds that in “do-it-yourself” qualitative testing it is no problem that there are some mishaps, such as deviating from the script or having to stop the test all together due to behavioural or environmental issues (Krug 2010,14). This is true especially with children when situations might be even more unpredictable and ever-changing. Since in UCD method the design goes through many iterations those inconsistencies usually show themselves in the next round of testing or are easy to spot by designers. That is not to say all the appropriate measures to ensure reliability should not be applied. Deviations are a risk in quantitative testing. It is focused on finding patterns in a large amount of data, so the tasks and questions must be consistent and same for everyone.

4 REMOTE USER RESEARCH AND TESTING

Remote user testing allows user research in any environment and location. Employing tools like screen-sharing software or commercial remote usability testing platforms one can create a test situation anywhere, at any time. Remote testing can be moderated in the same manner as in-person testing or unmoderated where users complete tasks independently. (Usability.gov, n.d.)

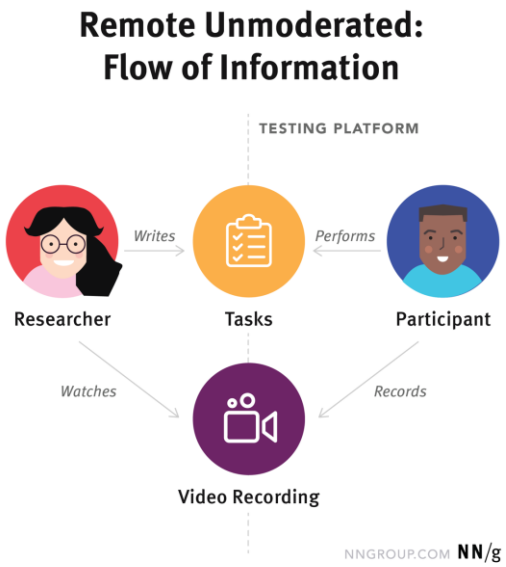
Remote research is not a fresh concept, it emerged when web-based services became commonplace. Bolt & Tulathimutte (2010, 30) estimated that in 20-30 years from the writing of their book people have grown accustomed to virtual remote meetings and environments and have learnt to read social cues from video feeds better (Bolt, Tulathimutte 2010, 30). Now, twelve years later the prediction is becoming reality. Krug (2010, 135) mentions having done remote user testing as early as 1995, before the existence of any video meeting software. He arranged testing sessions via a phone call and try to duplicate the user‘s actions on his own computer. (Krug 2010, 135.)

4.1 Benefits and challenges

Remote user testing is often considered to be more cost-effective. However, Bolt & Tulathimutte (2010, 241) remind that although remote research can save money on recruiting, space rental, and travel costs the biggest expenses – researcher salary and incentives remain (Bolt & Tulathimutte 2010). Also, according to Kauhanen (2021) in the case of application development at Yle remote testing with children is less cost-effective than in-person testing. Quick in-person testing with local children gives them time for more iterations and faster updates. This comes with the cost of not having a more diverse range of children evaluating the app. (Kauhanen 2021.)

Krug (2010, 135–136) argues that the biggest advantage remote testing is convenience. He states that all the time saved in recruiting and travelling results in a better outcome, since remote testing can produce ‘almost’ the same results as in-person testing. According to him this is dependent on if the test is monitored

or unmonitored. (Krug 2010, 135–136.) Monitored tests require more resources and planning. Unmoderated tests can run without facilitators (Picture 7). Another benefit of unmonitored remote testing is possibility to conduct time-aware and native environment research. The tests are not constricted to time, making it simple for users to participate at the time and place that is most convenient and natural to them. (Bolt & Tulathimutte 2010, 242.)



PICTURE 7. In unmoderated testing there are no facilitators, since no one is directly communicating with the participant. (Moran 2019)

An example of unmoderated, quantitative, and time-aware usability research is a prompt that users get while browsing the web or an app where they are recruited to participate in a brief test situation. The pop-up prompt can be, for example, a Likert scale questionnaire where users are asked how much they disagree or agree with a simple statement such as “I think the system is simple to use”. (Rosala 2020.) Likert is an excellent tool when working with children, too (see Picture 5).

In my quest to make remote testing feasible in our situation at work I considered is it possible to do unmoderated testing with children. I see that the session requires some level of moderation from a parent or a teacher at least with the youngest. When it comes to UI development, I consider remote testing as a staple tool alongside in-person testing. At Yle Lapset, before the COVID-19 pandemic

remote testing was practiced when there was a need to reach demographic outside of bigger cities (Helsinki/Tampere). Both Kauhanen and Mulari (2021) agree that some sort of hybrid model between remote and in-person testing would work the best, where there would be remote feed for researchers but also people on the other side helping to conduct the test (Kauhanen, Mulari, 2021).

A challenge for remote testing lay also in technology ecosystems. While testing users may encounter malfunctioning internet connection, security, or firewall issues and web camera and audio problems. Kauhanen (2021) mentions that not being able to help the users remotely with technical problems is frustrating (Kauhanen 2021). There is a question about accessibility as well. The user might not be comfortable with the level technology or have physical and mental disabilities to manage testing remotely. Children, despite being advanced mobile phone users, can still encounter problems they can't overcome.

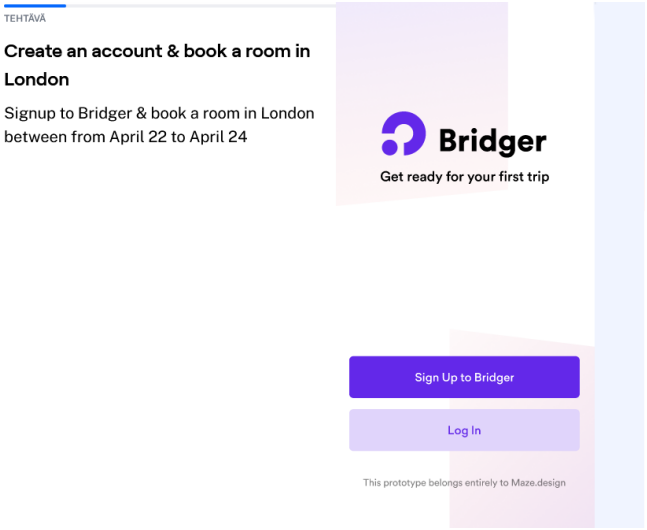
4.2 Tools

Bolt and Tulathimutte (2010, 182) list many remote testing tools and platforms that they predicted would vanish in five years of time. Some of the desired features remain regardless of time. (Bolt, Tulathimutte, 2010, 182.) Krug (2010, 137) describes that for the remote testing of an existing app or a website, the most essential tool is screen sharing. He also mentions that the deciding factor in choosing what tools and platform to use is the ease of use for participants. (Krug 2010, 137.) Nowadays, since most virtual meeting platforms work seamlessly on mobile screen, sharing is not an issue. With interactive digital prototypes of an app, monitoring user paths (user's movement through the product) is easier. Prototypes can also be customized to match the scenarios and scope of the testing.

An option for testing digital prototypes is to use the software they were built in. Sharing the interactive prototype build is a function in most interface design tools such as Adobe XD, Figma, Sketch and InVision. Usually, the function is used to share the prototype within the design team, but I have also used it for user testing sessions, too. For example, in a UI design course we used InVision platform to

build a prototype and then shared a link with our test subjects which they could open on their own mobile phone. This session was conducted in-person, but it could similarly function remotely. However, the issue would be not being able to monitor the user's actions unless they shared their screen in another program.

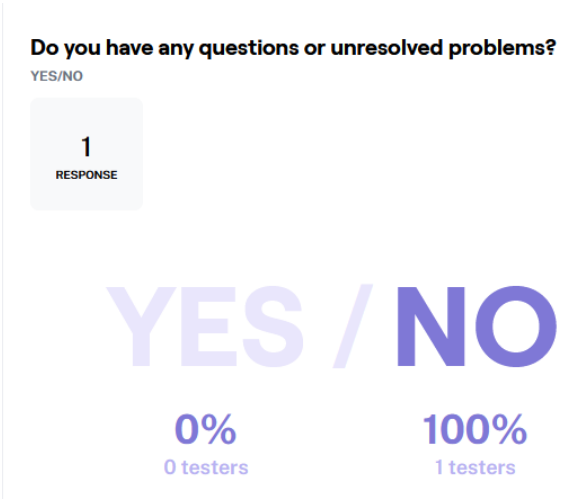
An example of a commercial tool that solves the problem of monitoring and data collecting when testing remotely is Maze. The company advertises itself as a rapid testing platform and their clients include big brands such as Uber, Klarna and Accenture. The website provides a web-based platform for usability testing as well as an automated reporting tool. On the platform researchers can customize the tasks and questions that will be given to the participants. Collaboration with etc. Figma and Sketch is available thus creating the tasks with interactive prototypes is possible. (Maze 2022.)



PICTURE 8, A task for users to complete in testing. (Maze 2022)

The user is given a simple task to complete on a mock website or a prototype (Picture 8). Tasks can also be multiple choices or open questions. The platform records their answers and maps the user paths until the final interaction. There are also duration metrics, measuring the user's time spent on a task or a particular screen. This along with other indicators such as amount of misclicks or heatmaps (where users click or hover the most) might reveal, for example, that the user cannot find what they are looking for. At the end users are given survey questions or opinion scales such as a Likert scale.

Where platforms such as Maze shine the most are the automatic reporting tools. On Maze profile you can review and share easy slideshow-like reports of all the collected user data (Picture 9). The platform does all the combining and statistics for researchers. Maze also offers a feature to work simultaneously with several team members and customize the live reports. If saving time is a priority, a service such as Maze is an easy choice.



PICTURE 9, Maze reporting tool. (Maze 2022)

A question worth considering is are these tools user-friendly for children. What I can tell from the example tests I tried on the Maze website, the tests seem to be simple and straight-forward. However, text-based questions or open answer boxes do not work well with children. I see that the using interactive prototypes and getting automated data from the usage could work but again not without some moderation or at least onboarding, familiarizing the users with the platform.

The advance of basic online meeting platforms such as Zoom or Google Meet is the availability. They are free to use, and most people already have experience using them. Their reliability is an issue. They are susceptible to many technical issues e.g., many of the platforms have a native function to limit a user's video– and audiofeed quality depending on their fluctuating bandwidth. As a user testing platform, they can be used for focus groups and interviews that happen in real-time.

5 REMOTE USABILITY TESTING THE YLE GALAXI APPLICATION

In this chapter I shortly try out the basic process of remote usability testing, where children would interact with the application and share their phone screen and activity on it. In this testing session I wanted to test out the possibility of usability testing through online meeting software Google Meet, that is used in Yle. I chose it because the platform is familiar to me and easily accessible to my test subjects. It was also a try-out in the fashion of discount usability since using it came at no cost. The aim of the test was not to get findings or insights on the usability of the application (Yle Galaxi), but to test out the actual testing method and platform. Originally, I had wanted to test out some new UI prototypes I had done for Galaxi but confidentiality reasons I chose to use the already existing Galaxi application.

5.1 Planning and recruiting

The group of children I recruited for the testing were participants of Yle Galaxi school collaboration project that I pioneered. During spring 2020 we wanted to establish a continuing partnership with a few local primary schools for user testing, but since the Covid-19 pandemic hit we had to quickly adapt to the remote work life. Luckily, the schools adapted to remote connections quickly as well, and we were able to continue the collaboration.

Keeping in mind all the legal requirements and permissions needed, for smaller or student productions with tighter budgets, schools are usually the easiest way to find groups of children. The tests can be arranged with the cooperation of the teacher staff in the middle of the school days or right after. Classes are usually willing participants if a moderate incentive is provided, e.g., some funds for a class trip. Of course, it is beneficial to do some screening, depending on the scope and reach of the project to find the best candidates for the scope of the testing. In this thesis testing, I worked with children I had had many focus group sessions with earlier. I knew that they were cooperative and most of were not afraid to express their opinions.

I started with defining the research question: “Is it possible to conduct usability testing remotely on an online meeting platform and how to do it?” Along with that I had to plan out the test script and tasks on the application. My hypothesis for the whole test was that it is possible, but not without technical issues and the lost perspective and information e.g., miscommunication because of audio or video issues. I was quite certain the platform that I chose was not the best choice for the testing, but I wanted to give it a chance.

5.1.1 Test group

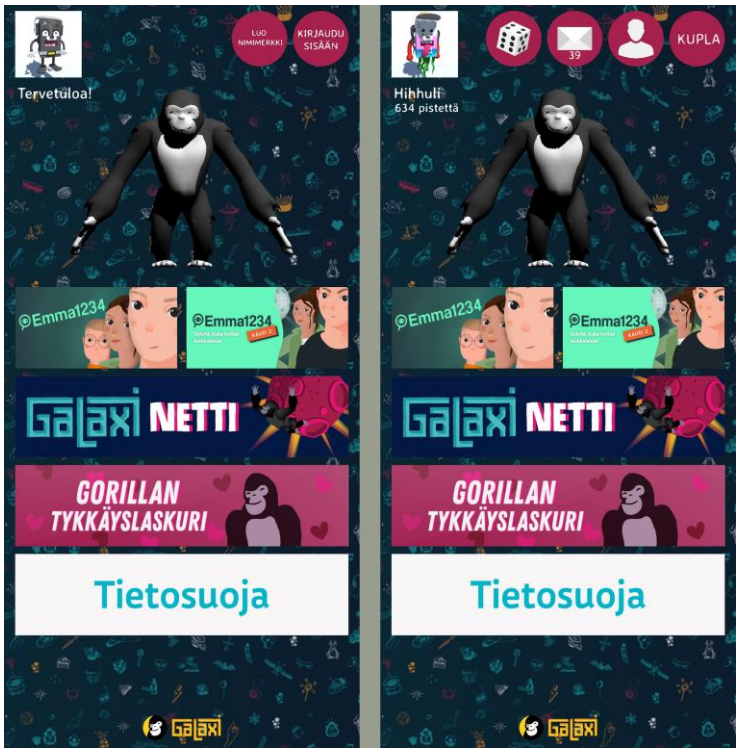
For this testing I had a group of six fifth grade students aged ten to eleven years old from Ylöjärvi. I had previously done many remotely connected testing sessions on Google Meet with almost all the students from the class. In the past sessions, I had presented them prototypes for the new Galaxi application and for example, asked them to figure out where certain things could be found. The same children also participated in focus groups for our social media development such as giving their opinions about Youtube videos. For this round I asked the class teacher to invite some students that have been working with me the least, to have some fresh perspective.

Earlier, I mentioned that I have noticed the benefit in having the children paired with their friends in the testing session and wanted to do it in this case as well. Especially being connected remotely, I in my own room, and the participants in a quiet closet somewhere around their primary school, I felt a child being alone in that situation might feel nervous and anxious. The children I recruited were classmates and close friends. Thus, I had three separate rounds, with two children with me each time. In the results I will tell why pairing up ended up being a bad choice.

5.1.2 Documents

Before my employment at Yle, I did heuristic evaluation and a user testing session of the Galaxi app as a university course project. With that and the development work I had done on the app after, I was already familiar with the issues and bugs

the app has. For the test script I chose a lot of open-ended questions that the children could ponder as they roamed freely in the app and followed them with easy and understandable tasks e.g., “Change the avatar’s hat”, “Where can you find the messages?” to elaborate. The Galaxi application requires users to create a nickname that is not linked to email or any other personal information, to access most of the functionalities (Picture 10).



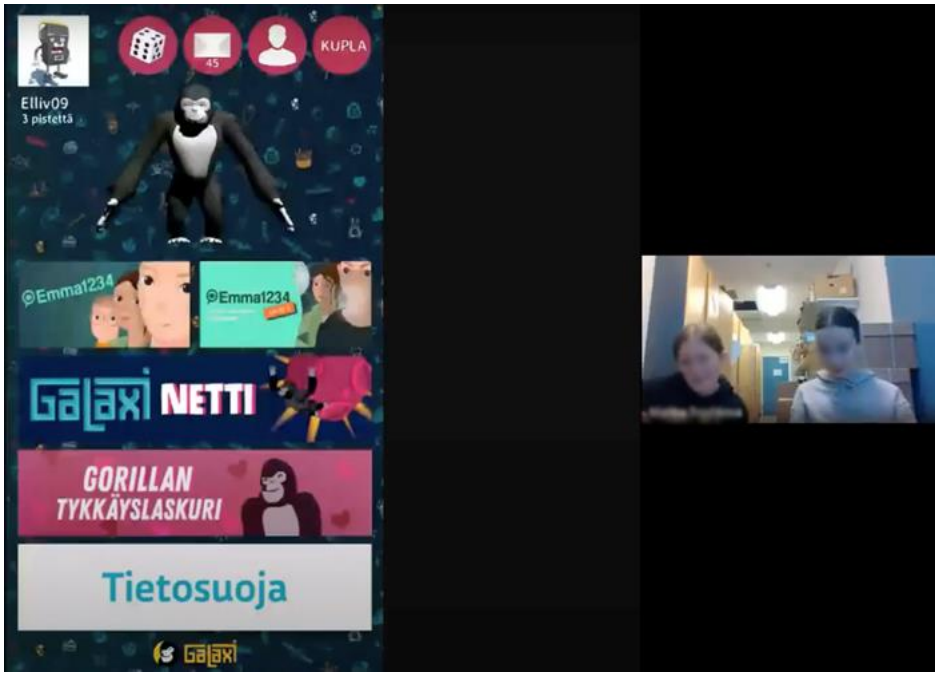
PICTURE 10, Galaxi app landing page signed out and signed in (Yleisradio 2021)

In the end I asked interview questions to get feedback. Full list of tasks is in Appendix 2.

5.2 Execution

The type of user testing I chose to do is usability testing. The testing was moderated and qualitative in its nature. According to the research question that was set I was only going to experiment with the platform and testing, and all the issues with the application itself did not matter. The session was chosen to be held in Google Meet. Although I had never done specifically application usability

testing on it, all my other remote testing sessions that were mostly focus groups were held on the platform during the Covid-19 pandemic.



PICTURE 11, The Google Meet testing view. (screenshot)

On Google Meet, the children enabled their web camera so I could see their faces and reactions (Picture 11). Earlier, a co-worker was recruited to be an observer and a notetaker. For the test itself I recorded the meeting, and the children performed tasks on their own mobile phones while sharing their screen in the meeting, so we could see what they were doing. The children had one laptop that they joined the meeting with and then joined it again on their mobile phones for screen sharing. We made the children also enable no-notification mode on their phone for privacy reasons.

5.3 Results

The first problem of the session appeared right in the beginning. Along with some connectivity and audio issues we quickly realized something mortifying– it is not possible for two people to share their screen at the same time on Google Meet. Although I had tested the mobile screensharing feature in preparation, sharing two screens was something I had failed to consider. Before, all the material I had presented to the children was coming from my end, so there was no need for simultaneous screensharing. The simple solution for this problem would have

been to have individual participants, but we could not change the set-up in the middle of the session. We continued the test, now only one child sharing their screen (Picture 12). We let both children participate in the tasks, while only the other one was operating the phone. The other participant had now potential to become essentially a back-seat driver that could influence or even annoy the one that was using the phone. I still think the presence of a friend had a positive effect on the testing. There was more dialogue, and the children were thinking out loud– that's something vital in all user testing sessions.

PICTURE 12, Sharing screen

In remote testing it is often not possible to rely on video feed to observe the user. The picture quality and the connection problems prevent that most of the time. For this test, and with children in general, it is beneficial to see their facial expressions. In this test, too, the children would often show their reactions on their faces or whisper it to their friend rather than saying an observation out loud. Seeing their reactions, allowed me to remind them to share the issues with me as well. If video is not an option, a good audio feed is a must. For an actual remote usability test of an application, I would want more reliable audio feed and

remind the participant to speak out loud what they are doing. Just looking at a screenshare of the participant using the application reveals nothing of their thoughts and possible frustrations. Google Meet is a great option for online meetings. For usability testing purposes, it is not optimized. The limits in sharing capabilities and output quality (audio, video) are its shortcoming. It would be great to use ready-made services such as Maze to get the most out of testing. Even if the testing session is moderated, an automatic reporting and user path-recording function would help. Then the facilitator could focus only on the users and tasks and save time after the testing.

In retrospect I should have asked for review opinions about the actual testing at the end. I have asked for feedback from the children through the teacher before about all the testing I had done. The children were quick to express that they would rather meet us in person and use their personal laptops at school. In the end, it was a good experiment. I consider failure a test result as well as it enables me to learn from my mistakes and to realize things that could have been done differently.

6 CONCLUSION AND DISCUSSION

The aim of the thesis was to research the possibilities and limitations of remote user testing with children. The goal was to learn more about user testing methods in application development and how those tests had been conducted remotely. Additionally, the aim was to try out a discount usability testing session on an online meeting platform. To reach the goal of the thesis it was pivotal to learn and understand how children are as users. A designer should be aware of a few developmental and ethical aspects before starting any work involving children. Acknowledging children as the experts of their own age group is important.

The aim of the Yle Galaxi app testing succeeded where the execution failed. The aim was to shortly try out the testing situation without paying attention to the content of the app. It was possible to conduct a usability test on Google Meet but not without a loss of perspective through technical issues. Google Meet is not equipped with all the necessary features to simultaneously observe, record and report on the test results. If more testing rounds would have been conducted, perhaps more of the issues could have been solved. Better and more accurate outcome could be achieved with software made specifically for remote testing. However, there is concern those platforms are not suitable or tested with children in mind.

A perfect remote testing platform for children would look a lot like apps and games directed at children. It would be colourful, playful, simple, and straight-forward. It would not include a lot of text and would rely on intuitive visual clues in the UI. For unmonitored testing, gamification could be an option. Upon completing a user task, the platform could give an engaging reward such as an animation, or imaginary progression points that would count toward a bigger reward at the end. Completing the test could give the child an access to e.g., a short mini-game or a certificate of participation. What would be complicated is the balance of making the reward gratifying enough, but not too valuable so that the children would not feel wrongly motivated to complete test.

In a few years, with the launch of 3D-world and augmented reality projects such as Facebook Metaverse, the work environments we are used to might look and function differently. For user testing, the hybrid between remote and in-person could be reached with virtual motion-tracking 3D spaces, with built-in plugins for recording and reporting. In 2021 Oculus and Facebook introduced a co-work platform application Horizon Workroom that people can use with a VR headset. It simulates an office room in which one can join a meeting as a 3D-modelled character. (Facebook/Meta 2021.) Zoom also announced the release of VR whiteboarding function to their platform, that allows collaborating and that it will be integrated into Horizon Workroom (Smith, 2021). VR platforms could also be a powerful tool with children, offering the possibility to create a familiar and safe virtual space e.g., a virtual classroom where remote testing could be conducted in. Assuming that all the participants have access to a VR headset.

In the end, the children of today are the future adult users to whom most applications are catered to. Although, the aspects of usability might not change much as time goes on, understanding the world where children are growing up now gives clues into what sort of needs people might have in the future. Especially when thinking of new, breakthrough solutions or services it is important to have a generational point of view.

REFERENCES

Abras, C., Maloney-Krichmar, D., Preece, J. 2004. User-Centered Design, In Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications

Bolt, N., Tulathimutte, T. 2010. Remote Research. Rosenfeld Media, LLC

Cantuni, R. 2020. Designing Digital Products for Kids: Deliver User Experiences That Delight Kids, Parents, and Teachers. Apress L. P.

Cherry, K. 2020. The 4 Stages of Cognitive Development
Background and Key Concepts of Piaget's Theory. Updated 31.3.2020. Read 11.3.2021 https://www.verywellmind.com/piagets-stages-of-cognitive-development-2795457

Clark, J. 2010. Tapworthy: Designing great iPhone apps. O'Reilly.

Facebook/Meta. 2021. Introducing Horizon Workrooms: Remote Collaboration Reimagined. Published 19.8.2021. Read 16.2.2022. https://about.fb.com/news/2021/08/introducing-horizon-workrooms-remote-collaboration-reimagined/

Farrell, A. 2005. Ethical Research with Children. McGraw-Hill Education.

Fike, E. 2016. Eradicating Uncertainty: Cognitive Friction and Your User. Published 28.11.2016. Read 11.11.2021. https://www.exclamationlabs.com/blog/eradicating-uncertainty-cognitive-friction-and-user/

Fisher, C. 2014, DESIGNING GAMES FOR CHILDREN: developmental, usability, and design considerations for making games for kids. Routledge.

Gladkiy, S. 2018. Article. User-Centered Design: Process and Benefits. Published 14.6.2018. Read 12.2.2021. https://uxplanet.org/user-centered-design-process-and-benefits-fd9e431eb5a9

Interaction Design Foundation. N.d. User-Centered Design. Read on 12.2.2021. https://www.interaction-design.org/literature/topics/user-centered-design

Kauhanen, H-M. Executive Producer, Yle Lapset. 2021. Interview on 17.11.2021. Interviewer Nenonen, N. Tampere.

Krug, S. 2014. Don’t make me think, revisited: a common sense approach to Web usability. New Riders, Berkeley, CA.

Krug, S. 2010. Rocket Surgery Made Easy: The Do-It-Yourself Guide to Finding and Fixing Usability Problems. New Riders, Berkeley, CA.

Lange, T. 2016. Tricks For User Testing with Kids. Published 19.6.2016. Read on 18.3.2021. https://uxdesign.cc/tricks-for-user-testing-with-kids-bd4cfaf36b12

Markopoulos, P., Read, J., Macfarlane, S. & Höysniemi, J. 2008. Evaluating children's interactive products: principles and practices for interaction designers. Morgan Kaufmann Publishers, Inc.

Moran, K. 2019. NNGroup: Usability Testing 101. Published 1.12.2019. Read 4.11.2021 https://www.nngroup.com/articles/usability-testing-101/

Moule, J. 2012. Killer UX design. SitePoint.

Mulari, H. Researcher, Yle Creative Content and Media. 2021. Interview on 17.11.2021. Interviewer Nenonen, N. Tampere.

Nielsen, J. 1993. Usability Engineering. Morgan Kaufmann, Amsterdam.

Nielsen, J. 2009. Discount Usability 20 Years. Published 13.9.2009. Read 25.2.2021. https://www.nngroup.com/articles/discount-usability-20-years/

Nielsen, J. 2012. Usability 101: Introduction to Usability. Published 3.1.2012. Read 27.10.2021 https://www.nngroup.com/articles/usability-101-introduction-to-usability/

Nielsen, J., Sherwin, K. 2019. Children's UX: Usability Issues in Designing for Young People. Nielsen Norman Group. Published 13.1.2019. Read 25.2.2021. https://www.nngroup.com/articles/childrens-websites-usability-issues/

Norman, Don. 2013. The Design of Everyday Things: Revised and Expanded Edition. Basic Books.

Oxford Learner's Dictionary. n.d. Definition of usability (noun). Published n.d. Read 8.2.2022. https://www.oxfordlearnersdictionaries.com/definition/english/usability?q=usability

Rosala, M. 2020. NN Group: Rating Scales in UX Research: Likert or Semantic Differential? Published 7.6.2020. Read 4.2.2022. https://www.nngroup.com/articles/rating-scales

Smith, J. 2021. Coming Soon: A Whole New Way to Collaborate Using Zoom Whiteboard. Published 13.9.2021. Read 16.2.2022. https://blog.zoom.us/zoom-whiteboard/

Usability.gov. N.d. How To & Tools: Remote testing. Read 28.6.2021. https://www.usability.gov/how-to-and-tools/methods/remote-testing.html

PICTURE REFERENCES

PICTURE 1
A UCD diagram. (2022). Nenonen, N. & Spring 2 Innovation. (2019). Published 18.4.2019. Read 20.12.2021. https://spring2innovation.com/design-thinking-vs-user-centred-design/

PICTURE 2
Trello. (n.d). Homepage, introduction.
Read 3.11.2021
https://trello.com/fi/home

PICTURE 3
Krug, S. (2014, 166). Example of user's goodwill draining because of cognitive friction

PICTURE 4
Paper prototype. (2019). Nenonen, N.

PICTURE 5
Cantuni, R. (2020). Designing Digital Products for Kids: Deliver User Experiences That Delight Kids, Parents, and Teachers. Apress L. P.

PICTURE 6
Rodrigues, J. (2019). User Experience Research: Interactive Toy for Kids to Improve The Fine And Gross Motor Skills. Published 23.9.2019. Read 4.2.2022. https://medium.com/john-rodrigues/user-experience-research-interactive-toy-for-kids-to-improve-the-fine-and-gross-motor-skills-6e5092973880

PICTURE 7
Moran, K. (2019). NNGroup: Usability Testing 101. Published 1.12.2019. Read 4.11.2021 https://www.nngroup.com/articles/usability-testing-101/

PICTURE 8-9
Maze. (2022). Homepage
Read 5.2.2022
https://maze.co/

PICTURE 10
Yleisradio Oy. (2021). Yle Galaxi application
Seen 10.11.2021

PICTURE 11-12
Screencaptures

APPENDICES

Appendix 1. Expert interview questions

FINNISH

- Mikä on mielestäsi tärkein asia, jonka ottaa huomioon tehdessä käyttäjätestejä lasten kanssa?
- Mitkä ovat suurimpia haasteita lasten käyttäjätestauksessa?
- Millaisia rooleja tiimissä on yleensä testaustilanteessa?
- Oletko koskaan tehnyt etäkäyttäjätestejä? Miten se eroaa paikan päällä testaukseen?
- Minkälaista palautetta olette saaneet lapsilta/vanhemmilta etätestauksen yhteydessä?
- Minkälainen luulet etätestauksen tulevaisuuden olevan? Minkälaisia ratkaisuja odotat tai olisi tarve keksiä siihen liittyen?
- Muuta mitä ei mainittu?

ENGLISH

- What do you think is the most important thing to consider when conducting user tests with children?
- What are the biggest challenges in user testing with children?
- What kind of roles does the team have while testing?
- Have you ever done remote user testing? How does it differ from in-person testing?
- What kind of feedback have you gotten from the children in remote user testing?
- What do you think is the future of remote testing? What solutions do you expect/need to emerge?
- Anything else that I failed to mention?

Appendix 2. Yle Galaxi app remote testing script

Instructions:

- Bring your own mobile phone and download the Galaxi app
- Join the meeting on your laptop
- Join the Google meet with your mobile phone and share screen
- Create nicknames on the app if not created

Tasks:

Task 1 - Landing page

1. What catches your attention right away?
2. What would you change?

Task 2 - Find the messages

1. What is in there?
2. What do you think about the messages?

Task 3 - Customise the Kamu - give Kamu a new hat

1. What else can you do here?
2. What are the colors? How do you get them?

End interview questions

1. What do you think about the app?
2. What is missing?
3. Would you recommend this app to your friends?