

#### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

## Riikka Kilpeläinen

# USABILITY TEST DESIGN FOR A NFC-BASED SEAMLESS LEARNING TOOL

Master's Thesis
Degree Programme in Information Networks
November 2014

Kilpeläinen R. (2014) Usability test design for a NFC-based seamless learning tool. University of Oulu, Department of Computer Science and Engineering. Master's thesis, 89 p.

#### **ABSTRACT**

As mobile technologies and social media are being integrated more and more to current educational practices, the notion of seamless learning has become important. Seamless learning enables learning across different contexts whenever the students are stimulated to learn. The main motivation behind seamless learning is to encourage the students to switch from one learning context into another easily with the help of a personal device as a mediator. In the most common case this device is a smartphone.

One challenge for wider seamless learning adoption is the lack of efficient tools with which teachers could create educational content that is more suitable for seamless learning. To address this challenge, the Interactive Spaces research group at the University of Oulu is implementing an editing tool, called NFC-ACT, with which teachers can create NFC based educational games and exercises for the students. Near Field Communication (NFC) is a short range wireless communication technology in which the devices initiate communication automatically once they are in close proximity. NFC is most commonly used in mobile phones. Thus NFC is a suitable technology for seamless learning applications.

In order to test the usability, usefulness and efficiency of NFC-ACT, this thesis presents a detailed usability testing plan for NFC-ACT. With thorough usability testing with teachers as test participants, the most critical usability issues with NFC-ACT can be revealed. In addition, the aim of the usability testing is to study the need for such editing tools.

This thesis consists of two parts. The first part is a comprehensive literature review about usability and seamless learning research. Then, the second part presents the current state of the application and goes through in detail the usability testing plan for NFC-ACT. Based on the literature review, the selected methodology for usability testing is an adapted combination of cooperative usability testing and SUXES method. In addition, the testing plan is constructed in a way that it takes into account the results from previously held usability demo sessions that already shed some light on the most critical usability issues with NFC-ACT. The resulted usability testing plan will be put into use once the implementation for NFC-ACT is ready.

Keywords: usability, usability testing, usability evaluation, cooperative usability testing, SUXES method, seamless learning, NFC

Kilpeläinen R. (2014) Käytettävyystestaussuunnitelma NFC-pohjaiselle saumattoman oppimisen työkalulle. Oulun yliopisto, Tietotekniikan osasto. Diplomityö, 89 s.

#### TIIVISTELMÄ

Mobiiliteknologiat ja sosiaalinen media ovat integroitumassa yhä enemmän vallitseviin opetuskäytäntöihin, minkä myötä saumaton oppiminen on noussut tärkeäksi käsitteeksi. Saumaton oppiminen mahdollistaa oppimisen eri konteksteissa milloin tahansa silloin kun oppilaat ovat stimuloituneita oppimaan. Päämotivaatio saumattoman oppimisen takana on kannustaa oppilaita vaihtamaan oppimiskontekstista toiseen mahdollisimman joustavasti käyttäen hyväksi eri laitteita. Tätä nykyä yleisin tällainen vaihdot mahdollistava laite on älypuhelin.

Eräs haaste laajemmalle saumattoman oppimisen omaksumiselle on tehokkaiden työkalujen puute, joiden avulla opettajat voisivat tehdä paremmin saumattomaan oppimiseen soveltuvaa opetusmateriaalia. Oulun yliopiston Interactive Spaces -tutkimusryhmä onkin kehittämässä tähän tarpeeseen editointityökalua nimeltään NFC-ACT, jonka avulla opettajat voivat luoda NFC-pohjasia pelejä ja tehtäviä oppilaille. NFC (Near Field Communication) on lyhyen kantomatkan langaton kommunikointiteknologia, missä laitteet muodostavat automaattisesti yhteyden heti kun ne ovat lähellä toisiaan. Koska NFC:tä käytetään yleisesti matkapuhelimissa, on se sopiva teknologia saumattomaan oppimiseen.

Tämä diplomityö esittelee yksityiskohtaisen käytettävyystestaussuunnitelman NFC-ACT:lle, jotta sen käytettävyyttä, hyödyllisyyttä ja tehokkuutta voitaisiin testata. Perusteellisella käytettävyystestauksella opettajien toimiessa testikäyttäjinä työkalun kriittisimmät käytettävyysongelmat voidaan paikallistaa. Lisäksi käytettävyystestauksen tavoitteena on tutkia millainen tarve tällaisille editointityökaluille on.

Tämä työ koostuu kahdesta osasta. Ensimmäinen osa on kattava selostus käytettävyyteen ja saumattomaan oppimiseen liittyvästä tutkimuksesta. Toinen osa esittelee työkalun nykyisen tilan sekä käy läpi yksityiskohtaisesti käytettävyystestaussuunnitelman. Ensimmäisen osan kirjallisen katsauksen pohjalta käytettävyystestauksen metodiksi muodostui kombinaatio, joka yhdistelee yhteistyökäytettävyystestausta ja SUXES-menetelmiä. Lisäksi testaussuunnitelma ottaa huomioon tulokset NFC-ACT:lle aiemmin järjestetyistä käytettävyysdemoista, jotka jo osaltaan antoivat suuntaa työkalun kriittisimmistä käytettävyysongelmista. Esitelty käytettävyystestaussuunnitelma on tarkoitus ottaa käyttöön heti, kun NFC-ACT:n toteutus on valmis.

Avainsanat: käytettävyys, käytettävyystestaus, käytettävyysarviointi, yhteistyökäytettävyystestaus, SUXES-menetelmä, saumaton oppiminen, NFC

## **TABLE OF CONTENTS**

٨	$\mathbf{R}\mathbf{C}^{T}$	$\Gamma$ $\mathbf{p}$	٨	C7
$\overline{}$	1117	1 1/	$\overline{}$	<b>.</b>

## TIIVISTELMÄ

#### **TABLE OF CONTENTS**

#### **PREFACE**

#### LIST OF ABBREVIATIONS AND SYMBOLS

1.	INT	RODUCTION	8
	1.1.	Background and motivation	8
	1.2.	Target and scope	9
	1.3.	Outline of the thesis	9
2.	USA	BILITY TESTING	10
	2.1.	Usability definitions	10
	2.2.	Designing usability	13
	2.3.	Usability evaluation	14
		2.3.1. Usability inspection	15
		2.3.2. Usability testing	16
	2.4.	Comparing the usability evaluation methods	18
	2.5.	Usability testing methods	19
		2.5.1. Think aloud testing	19
		2.5.2. Cooperative usability testing	20
		2.5.3. SUXES method	23
		2.5.4. Usability metrics	24
		2.5.5. Questionnaires, interviews and diary studies	26
	2.6.	· · · · · · · · · · · · · · · · · · ·	28
3.	NFC		30
	3.1.	Near Field Communication	30
	3.2.	NFC technology	31
	3.3.	= :	33
	3.4.		34
	<i>5</i>	The applications of the first o	٥.
4.	SEA	MLESS LEARNING	36
	4.1.	Background for seamless learning	36
	4.2.	M-learning	36
	4.3.	U-learning	38
	4.4.	NFC in mobile and ubiquitous learning	38
	4.5.	Characteristics of seamless learning	39
	4.6.	Seamless learning applications	42
		4.6.1. Authoring tools	42
		4.6.2 Mobile assisted tools for self-directed seamless learning	43

	4.6.3	. Seamless learning applications for collaborative learning	45
	4.6.4	Educational games	46
5.	NFC-ACT	APPLICATION	47
	5.1. NFC	-ACT description	47
	5.2. NFC	-ACT as a seamless learning tool	50
	5.3. Curr	ent state of NFC-ACT	51
	5.3.1	. NFC-ACT usability demo sessions	51
	5.3.2	Results from demo sessions	53
6.	USABILI	TY TEST DESIGN	62
	6.1. Struc	cture of testing sessions	62
	6.2. Testi	ng goals	62
	6.3. Test	sessions	65
	6.3.1	. Introduction session	65
	6.3.2	. Usability testing session	66
	6.3.3	Diary phase	70
	6.3.4		71
7.	DISCUSS	ION	72
	7.1. Eval	uation of the results	72
	7.2. Qual	ity of the research	72
		plan validity	74
		re improvement ideas	75
8.	CONCLU	SION	77
9.	REFERE	NCES	<b>78</b>
10.	APPEND	ICES	84

#### **PREFACE**

The process of making this thesis has been an educational one but rather long and it has required time and patience from the people who have supported me in completing this work. I now want to express my gratitude to the people involved.

First, I wish to thank Professor Jukka Riekki for providing the subject for this thesis. In addition, I want to thank Professor Riekki for his guidance and comments throughout the whole process. I also want to thank Iván Sánchez for providing me information about NFC-ACT and for his comments on this work as well. I also want to thank him for his participation in the NFC-ACT demo sessions as well as for his contributions in the actual software implementation. I also wish to thank the second examiner, Dr. Mika Rautiainen.

Warm thanks belong to the demo session participants as well as for the whole NFC-ACT development team. Finally, I want to express my sincerest gratitude to my family and friends who have been supporting me throughout my studies.

Oulu, November 27, 2014

Riikka Kilpeläinen

#### LIST OF ABBREVIATIONS AND SYMBOLS

 $\lambda$  Probability of finding any single usability problem with single

test user

 $E_{partially}$  Number of partially smoothly completed tasks

 $E_{smoothly}$  Number of smoothly completed tasks

*i* Number of test users

CSF Critical Success Factor
CTA Concurrent Think Aloud
CUT Cooperative Usability Testing
FSL Facilitated Seamless Learning
GUI Graphical User Interface
HCI Human Computer Interaction

HF High Frequency

IEC International Electrotechnical Commission
ISO International Organisation for Standardisation

LF Low Frequency

MNO Mobile Network Operator
MSA Measure of Service Adequacy
MSL Mobile-assisted Seamless Learning
MSS Measure of Service Superiority
NFC Near Field Communication
NFC CLF NFC Contactless Front-End

NPS Net Promoter Score
OTA Over The Air
QR Quick Response

QUIS Questionnaire for User Interaction Satisfaction

RFID Radio Frequency Identification RTA Retrospective Think Aloud

SE Secure Element

SUMI Software Usability Measurement Inventory

UCD User Centred Design
UHF Ultra High Frequency

UI User Interface

#### 1. INTRODUCTION

#### 1.1. Background and motivation

Technological advancements in mobile computing and wireless communication together with new platforms created by social media have initiated two major changes in the field of education [1]. The first shift involves the sense of spaces and contexts where education takes place as learning activities are being distributed among different contexts. The second shift acknowledges the importance of designing new learning opportunities created by the new contexts.

To better address these changes in education the notion of seamless learning was introduced. Seamless learning means distributing learning opportunities across formal and informal contexts [1]. It is a learning model that empowers and supports learning across a variety of scenarios whenever the students are stimulated to learn [2].

Near Field Communication (NFC) technology is one of the mobile technologies enabling seamless learning. It is short range wireless communication technology based on touching, i.e. the two communicating devices need to be in close proximity (or touch each other) in order to initiate communication. NFC is most commonly used in mobile phones, so that the NFC enabled phone is used for reading NFC tags and then an application in the phone executes according to the acquired data.

NFC enables seamless learning as with NFC learning applications can be brought to the everyday environment and thus learning is not only tied to the school environment. Moreover, NFC enables more natural learning as with NFC it is easy to create links between physical world and digital data and the students can then explore these links with the help of their NFC-enabled device. [3]

However, there still is need for gaining a better understanding on how to design the new learning experiences to support seamless learning. It is important to investigate how students interact with different technologies, learning contents, peers, teachers and parents in various contexts [1]. It is also equally important to study how seamless learning integration could be better facilitated from the teachers' point of view. From this perspective tools enabling seamless learning content creation for teachers are needed.

One such tool is NFC-ACT. The tool is being implemented by the Interactive Spaces research group at the University of Oulu and it enables the creation of NFC based learning content. It provides a web-based user interface with which the teachers can create games and exercises, write related game information to the NFC tags and upload the game to the NFC enabled phones.

In order to better facilitate seamless learning, the tools created must also serve a purpose and be efficient and easy to use. In another words, the tools must have good usability. Usability is the extent to which an application enables users to achieve specified goals efficiently and effectively [4]. Naturally, any application and system needs to have good usability but usability and usefulness play a crucial role especially in the case where the users are teachers with busy schedules.

#### 1.2. Target and scope

The aim of this thesis is to present a comprehensive usability testing plan for NFC-ACT in order to find out to find out usability and usefulness aspects of NFC-ACT. A few preliminary usability testing session for NFC-ACT have already been held but a more comprehensive usability testing session is still needed. As the software is still under development, this thesis focuses on proposing a detailed usability testing plan that follows usability testing principles and techniques.

The selected usability testing method combines cooperative usability testing with a usability testing method called SUXES. In order to provide a solid background for the usability testing plan this thesis presents a comprehensive literature review of usability and usability testing.

As NFC-ACT is a tool aimed at supporting seamless learning, this thesis will also provide a comprehensive literature review about seamless learning and its different concepts.

#### 1.3. Outline of the thesis

The introduction presents the background and the topic for this thesis. Also the target and scope of the research are defined.

Chapters 2, 3 and 4 present the theoretical background for this thesis. Chapter 2 forms a literature review about usability and usability testing. Chapter 3 gives a brief description about NFC technology and drivers behind NFC. It also presents some NFC applications. Chapter 4 focuses on a detailed literature review about seamless learning and its concepts including mobile and ubiquitous learning. Chapter 4 also presents seamless learning applications.

Chapter 5 describes the developed NFC-ACT tool and evaluates its characteristics as a seamless learning tool. Chapter 5 also presents the findings from the previously held usability testing sessions. These findings were used as background when developing the usability testing plan for NFC-ACT.

Chapter 6 presents the actual usability testing plan. The goals of the testing are defined and the different phases of the usability testing are presented in detail, including selected testing techniques for each phase.

Chapter 7 evaluates the validity of the usability test plan as well as gives a critical evaluation of the whole research. Also some future improvement ideas are presented. Finally, Chapter 8 summarises the thesis and lastly a List of References and Appendices are presented.

#### 2. USABILITY TESTING

#### 2.1. Usability definitions

The term usability is used to denote a design that is good from a Human-Computer Interaction (HCI) point-of-view. Usability brings many benefits to applications that include increased productivity, enhanced quality of work, and improved user satisfaction, to name a few. [4]

In software development, usability has been an important aspect since the early 1990s when computers started to gain a more wider user population and graphical user interfaces became the norm. Nowadays, manufactures use usability as an differentiator from competitors and user interface designs are strictly protected with patents.

However, it is important to note the difference between usability and user experience as sometimes usability is confused with user experience. [5 p. 15] defines user experience as meaning how a product behaves, how people use it, how people feel about it and do the people feel satisfaction when using, looking, and holding the product. Usability then is one of the factors defining a good user experience. As mentioned in [5 p.15, 20] "every product that is used by someone has a user experience" and usability is a crucial part of that experience and likewise, a good user experience requires good usability.

One way to demonstrate the difference between usability and user experience is through their goals. Usability goals are usually defined as measurable attributes (effectiveness, efficiency, learnability etc.) making sure that these usability attributes are met. However, user experience goals are more concerned with the nature of the user experience, for example, is the product rewarding, emotionally fulfilling, or aesthetically pleasing. [5 p. 20]

How then to define that an application has good usability? Generally speaking usable systems are often easy to learn, efficient, not error-prone, and satisfactory in use. However, despite the wide research in the usability field, an exact, agreed upon definition for usability is still missing. Thus, one cannot express usability in only one objective measure. [4]

The next chapters will go through some usability definitions. First usability is defined from the viewpoint of different usability attributes. Then the usability trade-off is presented as well as a definition for usability from the viewpoint of understanding the end-users. Last, two usability standards are presented.

Jakob Nielsen, who is one of the pioneers in usability research and usability engineering, sees usability as one of the attributes that define the practical acceptability for a system [6 p. 24]. The components of practical acceptability are shown in Figure 1 [6 p. 25].

The overall acceptability is the sum of social and practical acceptability and it tells if the system meets the different requirements of the system stakeholders. Nielsen wants to differentiate usability from utility by roughly defining that usability measures whether the users can use a functionality of the system, and utility as a measure for can the functionality of a system do what the user needs it to do. In Nielsen's model usability together with utility forms usefulness, i.e. can the user use the functionality of a system to achieve a specific goal. [6 p. 24-25]

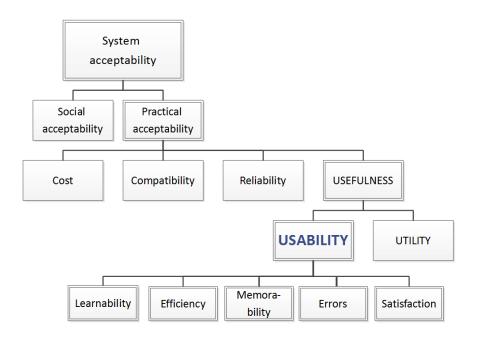


Figure 1. Usability attributes and system acceptability.

Instead of providing a precise definition for usability, Nielsen divides usability into five measurable attributes: learnability, efficiency, memorability, errors, and satisfaction. These attributes are derived from Nielsen's view of usability where it is affiliated with all the parts of a system that require human interaction. [6 p. 25-26]

Learnability describes how easy it is for new users to learn how to use the system. In most of the cases even complete novice users should be able to learn how to use the system in a reasonable time. After the users have learned how to use the system, efficiency then measures how productive the users can be while using the system. However, many systems also have casual users who use the system intermittently. Memorability measures how quickly such users can remember how to use the system after a period of time of not using it. System's usability can also be measured by analysing the errors that users encounter while using the system. Errors can be analysed based on their severity, reproducibility rate and how easy it is to recover from them. The last usability attribute, satisfaction, describes how pleasant the system is to use. [6 p. 27-33]

Whereas Nielsen sees acceptability as a sum of its components, Shackel [7] describes acceptability from a trade-off paradigm point of view where users weigh utility, usability and likeability against the cost of use. Moreover, Shackel notes that the usability of a system depends also on the dynamic interaction of the four components in any user-system situation: user, task, tool, and environment. In this framework usability is gained when the tool is designed in relation to the users, tasks and the environment so that ease of use is achieved in terms of learnability, comfort, satisfaction, effectiveness and repeatability. Based on the trade-off paradigm and the usability framework, Shackel defines usability as "the capability in human functional terms to be used easily and effectively by the specified range of users, given specified training and support, to fulfil the specified range of tasks, within the specified range of environmental scenarios" [7].

Shackel also provides a definition for usability in quantifiable terms, namely effectiveness, learnability, flexibility, and attitude. Here effectiveness means the level of performance of accomplished tasks for example in terms of speed and errors, learnability tells the amount of learning required to accomplish a task based on training and user support, flexibility describes the adaptation to variations in the system and attitude means that the user is satisfied with the system in order to continue using it. [7]

Also Desurvire goes through the difference of usability and usefulness in [8 p. 200] by mentioning that the two concepts overlap each other. Desurvire sees that usability is more focused on the user and the system, whereas usefulness is more focused on the interaction between the user and the system, i.e what does the user achieve by using the system.

Dumas and Redish on the other hand see usability as en enabler that allows the users to use a product quickly and easily to accomplish their own tasks [9 p. 4]. They base this definition on four points which describe what kind of understanding about the users the development team should have in order to design and implement usable products. These points are (list adopted from [9 p. 4]):

- 1. Focus on the users.
- 2. Users want to be productive.
- 3. Users are busy and they aim to accomplish specific tasks.
- 4. Users have the authority to dictate when a product is easy to use.

The first point is that usability means focusing on users. This point requires the development team to know and understand the potential users of the product. It also requires the development team to actually work with the users during the development phase. The second point requires an understanding of the users' performance goals by stating that people use products to be productive. The third point requires an understanding of the users' concerns with productivity by highlighting the fact that users are busy people trying to accomplish a task. The users must be able to use the product quickly and easily in order to accomplish their goals. The fourth point requires an understanding of the users learning curve by mentioning that the users have a low amount of tolerance for time spent learning how to use the system. Hence it is the users who decide when a product is easy to use. [9 p. 4-6]

In addition to definitions from different usability experts, usability has also been defined in various standards. The International Organisation for Standardisation (ISO) and the International Electrotechnical Commission (IEC) have developed various usability related standards of which the ISO 9241-11 and ISO 9126-1 standards are most widely known. Both of them define usability by using measurable attributes and as they complement each other both of them should be taken into account when designing and developing products. [10]

ISO 9241-11 (1998) emphasizes the context of use by defining usability as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Here effectiveness means the accuracy and completeness with which users achieve specified goals, efficiency means the resources expended in relation to the effectiveness, satisfaction is

seen as positive attitudes towards the use of the product and the context of use means the specific circumstances in which a product is used, i.e. the users, tasks, hardware, software, materials, and the physical and social environments [11].

ISO 9216-1 (made in 2000) is based on ISO 9216 (1991) which defined usability as a separate component of software quality by seeing usability as a "a set of attributes that bear on the effort needed for use, and on the individual assessment of such use, by a stated or implied set of users"." As usability was seen as a part of software quality the emphasis was on the usability of the user interface and the context of use was left out from the definition. The definition for usability was then refined in ISO 9216-1 by defining usability as "the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions". With this definition the ISO 9216-1 also acknowledges that usability needs to be defined always in a particular context. [10]

#### 2.2. Designing usability

Usability is ensured by incorporating it already in the design process. Depending on the source, incorporating usability to the design process is called either usability engineering or user centred design. The key component in both of the processes is to clearly define usability goals and then via an iterative process of development and testing aim to meet the goals.

In addition both processes highlight the importance of incorporating usability as early as possible to the design and development of the application. As [12] mentions by involving usability early enough it has an impact on the conceptual model of the product, the user interactions, and the user interface (UI).

Usability engineering starts with identifying users and by analysing the tasks and setting usability goals. An early focus on usability creates more usable products, enables changes before they become too expensive, enhances documentation and reduces the need for maintenance releases. [9 p. 8-9]

The key component in usability engineering is to set specific usability goals. The goals need to be clearly defined and measurable in order to set a basis for usability evaluation. Usability engineering also highlights the importance of incorporating real users throughout the engineering process. The users should be taken into account already when designing the usability goals as in usability engineering the users' need should drive the design. Also the different versions of the product should be tested with real users throughout the iterative design process. Moreover, in usability engineering the usability evaluation should be conducted by combining different usability evaluation methods. The different usability evaluation methods are described in chapter 2.3. [9 p. 9 -13]

Also user centred design (UCD) calls for direct contact with the users. Shackel [7] describes UCD as a requirement that the designers must understand who the users are and what their tasks in the system are going to be. Shackel sees UCD as one of the five fundamental features when designing usability. The other four are participative design, experimental design, iterative design, and user supportive design.

In participative design a panel of expert users evaluate the usability of the design by using e.g. mock-ups and simulations. Experimental design is used to assess ease of use

and ease of learning by conducting pilot tests with the expected users. Experimental design process should also reveal any difficulties that the pilot users have. These pilot tests should provide measurable results as well as the users' subjective reactions. Iterative design requires a regular cycle of design, test, measure and redesign to be carried out until a satisfactory level of usability is achieved. User supportive design is then used for designing the help and support features. [7]

Rubin on the other hand sees UCD from two different angles. It can be defined as a variety of techniques, methods and procedures for designing usable systems during different points in the product development phase. These techniques include for example participatory design, surveys, expert evaluations and usability testing. On the other hand UCD can be elaborated and seen as philosophy that places the user at the center throughout the entire cycle of user ownership of a product, i.e. UCD should be applied also in prepurchase and postpurchase interactions with the users instead of just during the product development phase. [13 p. 10, 19-22]

The ISO 13407 (1999) standard on human-centred design standard defines UCD as an iterative process that consists of design activities that need to take place throughout the development life cycle [14] until the usability goals (i.e. usability requirements) are met. The activities include (list adopted from [14]):

- understand and specify context of use,
- specify the user and organisational requirements,
- produce design solutions, and
- evaluate designs against requirements.

Figure 2 [14] shows how the above mentioned design activities need to be iterated until the usability goals are achieved.

#### 2.3. Usability evaluation

Both usability engineering and user centred design are based on iterative development which requires that the usability of the application is evaluated during each cycle. There are various techniques for conducting an usability evaluation. The different techniques can roughly be divided in two categories: usability inspection<sup>1</sup> and usability testing. In usability inspection the usability of an application is evaluated, either by professionals or the targeted users whereas in usability testing the usability is evaluated by letting end users use the application, most often in a controlled environment. In some sources the definitions for the two techniques overlap. However, the goal of the both techniques is the same: to improve usability by finding the problem areas in order to fix them.

Usability testing and inspection methods can also be combined for usability evaluation. This is known as triangulation. Triangulation provides different perspectives to usability evaluation and amplifies the different findings across the used techniques, thus leading to more profound findings [5 p. 293].

<sup>&</sup>lt;sup>1</sup>The term *usability evaluation* is sometimes used in literature for meaning *usability inspection* but in order to avoid confusion the term usability inspection is used in this text.

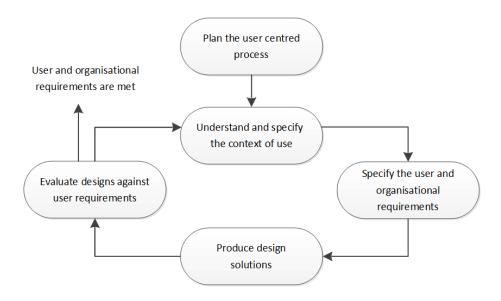


Figure 2. The user centred design cycle.

Moreover, as Nielsen points out in [15], each usability evaluation method has its strengths and weaknesses. Hence by supplementing each method with the best input from all the other techniques triangulation facilitates ruling out misleading outcomes.

[12] describes a project that used iterative usability evaluation combining different usability evaluation methods from both usability inspection and usability testing for each iteration. According to the results the combination of various usability evaluation methods yielded an enhanced user experience for the end product.

However, no matter what the chosen method for usability evaluation is, the most important thing is to clearly define the usability goals. Based on these goals one can then determine what method or methods are the best in order to reach those goals within the given development parameters [8 p. 207].

In addition, regardless of the used evaluation methods, usability evaluations can be divided into two categories: summative and formative evaluations. In a summative evaluation the overall quality of a finished product is evaluated against a set of requirements, e.g. standards whereas in a formative evaluation the product is evaluated during the development process in order to find out how it can be improved [5 p. 589].

#### 2.3.1. Usability inspection

In usability inspection usability related aspects of a certain user interface are evaluated by one or more evaluator [8 p. 1]. Depending on who is doing the inspection, usability inspection methods can roughly be divided into two groups: expert inspection and usability inspections by the end-users.

The different techniques for expert inspection include heuristic evaluation, guideline reviews, pluralistic and cognitive walkthroughs, consistency and standard inspections, formal usability inspections and feature inspections [8 p. 1].

In heuristic evaluation usability specialists evaluate each dialogue element of the UI against a set of predefined usability principles, i.e. heuristics. Guideline review is similar to heuristic evaluation as in guideline reviews the UI is also checked against a list of usability guidelines. The difference is that the guidelines can be more comprehensive and larger in number than the heuristics. [8 p. 5-6]

In walkthroughs the users evaluate their actions in the UI with evaluators. In a pluralistic walkthrough the users, developers and usability specialists together go through a scenario and discuss the problems that the user had. In cognitive walkthrough each step that the user made is evaluated in order to see if the step leads towards the next correct action i.e. the interface is evaluated against a cognitive framework to see how easy it is to learn to use the interface. Consistency inspections are used to evaluate if the UI of a family of products is consistent and in standard inspections the UI is evaluated against a standard. In formal usability inspections a group of evaluators with well-defined responsibilities inspect the UI. Feature inspections evaluate the different functions of the system. [8 p. 5-6]

Usability inspection techniques with end users include focus groups, surveys, and field studies. These techniques are aimed at obtaining information about user preferences and needs.

Focus group consists of a group of end users who discuss and evaluate an interface in a informal session [6 p. 214]. While focus groups can be used in all stages of development, they are mostly used at the very early stages in order to evaluate preliminary concepts using representative users [13 p. 20].

Surveys are used for obtaining user preferences from a broader group of users. They are most often used in early stages of development in order to gain more understanding about the potential user population. [13 p. 20]

Another usability inspection method involving the actual end users is field studies. A field study is an inspection of an application that has been placed in its actual environment of use [13 p. 23].

Field studies can be conducted after the release for follow-up study purposes but they are mainly used just prior to release for collecting data in order to refine the UI. Collected data includes patterns of use, difficulties and user opinions. Field studies conducted prior to release are also known as alpha or beta testing. Data can be gathered in two ways: the usability professionals can visit the end users and observe them using the application or the application is available for a specified group of end users who then provide feedback remotely. [13 p. 23]

#### 2.3.2. Usability testing

Usability testing is the most widely used usability evaluation method and it has been used ever since it gained popularity in the early eighties. It is also regarded as one of the most effective contributors in strategic focus on usability. [16]

In usability testing end users test an application with the aim to find out whether the application is usable by its intended users in order to achieve the tasks for which it

was designed for [5 p. 646]. Usability testing can be used for collecting quantitative data, for example performance, made errors and time used for navigation. Similarly, usability testing can also be used for collecting qualitative data in order to find out what the end users think about the product: are the symbols and icons clear, how intuitive it is to use the system, do they think the product is useful etc. No matter what the goal of usability testing is, it is best suited for diagnosing usability problems in order to improve the overall quality. [5 p. 646]

Dumas and Redish define usability testing as a "systematic way of observing actual users trying out a product and collecting information about the specific ways in which the product is easy or difficult for them" [9 p. 12]. They elaborate this definition by providing five characteristics that usability testing should always fulfil (list adopted from [9 p. 22]):

- 1. The goal is to improve usability.
- 2. The test participants represent real users.
- 3. The test participants do real tasks.
- 4. The actions of the participants are observed and recorded.
- 5. Results are analysed.

As one can see from the above list the primary characteristic of usability testing is to improve the quality of the product as well as improve the process how the product is designed and developed. However, it is good to note the difference with usability testing and quality assurance: quality assurance is used for evaluating whether the functional requirements of the system are fulfilled and usability testing evaluates if these functional requirements fulfil the needs of the end user. The second and third characteristic highlight the importance of understanding the end users. It is crucial to test usability with real end users and it is equally important to design the test tasks so that they are relevant to the end users. In addition, it is important to design the test tasks so that they relate with the goals set for the usability testing. The fourth and fifth characteristics note that data analysis is equally important to data gathering. [9 p. 22-25]

Also Rubin points out the importance of iterative nature of usability testing by defining that the overall goal of usability testing is to identify and rectify usability deficiencies in order to ensure that the application is created so that the product or system is easy to use and easy to learn how to use, is satisfying to use, and provides utility and functionality [13 p. 26].

Brooks in [8 p. 265] defines usability testing as a mean for providing help for development decisions. Brooks also mentions that usability testing can be utilised to indicate how acceptable the UI is and indicate how the UI will perform in a competitive market.

Usually a usability testing session is based on test tasks. As one cannot test every possible task the users can do with the application, the test tasks guide the user in the usability test. The test tasks should be selected so that they probe the potential usability issues with the application. The task set should also reflect the usability goals defined before usability evaluation is initiated. [9 p. 160]

Instead of providing the user with a list of the test tasks that are expected to be carried out in the usability test, the test tasks are formed into scenarios. The scenarios tell the users what they are expected to do during the test without revealing the exact steps [9 p. 172]. Scenarios can also be modified and adapted in order to test more than one version of a certain design [6 p. 18].

There are several factors contributing to the popularity of usability testing but the most prominent ones are the facts that it resembles the everyday use of interactive systems and that it is practical to manage. Resembling everyday use ensures that usability testing has high face fidelity and usability testing is practical to manage as it usually consists of a small number of evaluation sessions that on average lasts about an hour. In addition usability testing is generally seen as a facilitator for the communication between development teams and usability professionals as the end users' experiences from usability problems may be more convincing to the developers than the results from a usability inspection. [16]

There are many different usability testing methods available. These include among others think aloud usability testing and cooperative usability testing. Most often different usability testing methods are also complemented with different questionnaires and surveys. Section 2.5 presents different usability testing methods.

#### 2.4. Comparing the usability evaluation methods

Karat in [8 p. 207-210] presents some studies on comparing usability inspection against usability testing. In the first one usability testing, guideline evaluation, heuristic evaluation and cognitive walkthroughs were compared. The results indicated that heuristic evaluation identified the most usability problems and usability testing was the second best choice as it found serious and recurring problems. The second study compared usability testing and heuristic evaluation and as a result it was found that heuristic ratings were effective in predicting tasks where problems could occur but usability testing was better in finding the actual problems in the UI. Another study also came to the conclusion that even though heuristic evaluation was effective in predicting the spots for severe usability problems, it was able to predict only 44 % of the total usability problems identified in the usability testing. One more study that Karat describes compared usability testing, individual walkthrough and team walkthroughs and found that usability testing identified the largest number of usability problems as well as the largest number of usability problems than usability inspection.

Also Brooks came to the same conclusion in [8 p. 258] by stating that usability inspections are not as effective as usability testing in understanding what are the most important dimensions for users in the particular environment where the system is used, i.e. usability inspection is good at identifying the potential usability issues but usability testing is required for finding a solution for that problem.

Cost-quality trade-off is another factor to be taken into account when choosing between usability inspection and usability testing. [17] points out that as usability inspection methods requires experts, the cost of using them might become high. For example the best results from heuristic evaluation are achieved when it is conducted by double experts, i.e. persons who are experts in both usability as well as an expert in the par-

ticular type of interface that is being evaluated. Obtaining such evaluators for usability inspection may increase the cost of evaluation drastically and such evaluators may be hard to find. Moreover, usually more than one evaluator is needed for conducting the heuristic evaluation, thus increasing the cost even more. [17]

However, as usability evaluation methods can be applied already in early development, they enable rapid iterations by pinpointing usability issues earlier and thus the the overall cost of usability evaluation is less than the overall cost of usability testing. As Brooks mentions in [8 p. 256] it is less expensive to identify and remove major problems earlier than later in the development. But when weighing the cost of the usability inspection and usability testing, the output of the evaluation i.e. the quality should also be considered. As previously mentioned, usability testing reveals more severe issues than usability inspection and thus it cannot never fully substitute usability testing. [17]

Another aspect when comparing usability inspection against usability testing is to look what kind of results can be achieved with each technique and how the results can be communicated onwards [8 p. 201]. As Brooks mentions in [8 p. 255-258] an appropriate evaluation method cannot be chosen until it is fully understood that who will use the data provided by the evaluation and more importantly, what decisions will be made from the provided data. Usually usability inspection is effective when the results are used for identifying usability problems or choosing from among different design alternatives during early development. However, results from testing with real users provide better data for assessing the important factors in the interface design as well as data for finding out what factors correlate with product success in the market.

Lastly, finding usability problems is not enough as it is equally important to deliver a system that is more usable than the competing solutions available. Usability testing is more accurate in comparing different interfaces than usability evaluation. [8 p. 260]

#### 2.5. Usability testing methods

#### 2.5.1. Think aloud testing

The most simple usability testing method is the observation method where information for usability evaluation is collected by observing users' experiences while they are using the application under testing. The observation method has been proven by many researchers as an efficient method for usability testing [4].

It is important that the test monitor does not interfere the test user during an usability test that uses observation method. However, this prevents the test monitor from asking any clarifying questions when the user makes an error or expresses frustration. The questions can of course be asked after the user is finished with the test but then the user might not recall all that happened during the test.

Think aloud testing is an elaboration of the observation method. It was developed to address the above mentioned challenges. In a think aloud usability test the test user is asked to use the system and verbalise their thoughts simultaneously. Thinking aloud is a widely used method in usability testing since the gathered data reflects on the actual usage of a system instead of focusing on the test user's opinion about its usability. [18]

When the test subjects verbalise their thoughts they provide insight to their understanding of the system and thus they provide more understanding to their misconceptions with the system [6 p. 195], which is the major advantage of think aloud testing.

Think aloud testing has also its drawbacks and [18] presents some of these disadvantages. First of all, thinking aloud does not come very natural to most people and participants may require constant remainders to keep thinking aloud. Moreover, participants may work differently in a think aloud session as compared to normal use which may lead to false results. For example thinking aloud may have an affect on the way how the test participants handle the test tasks, the time for carrying out the tasks may be different compared to normal use or the test participants may even be so distracted by thinking aloud that they are unable to complete tasks. Due to this think aloud testing is not suitable for performance measures.

Think aloud tests can be divided into two categories: concurrent and retrospective thinking aloud. In concurrent thinking aloud the participants think aloud simultaneously while they use the system, and in retrospective think aloud the participants first carry out their test tasks silently and then after the test they verbalise their thoughts based on a recording (most often a video) of the test session. One advantage of retrospective think aloud testing is that it should decrease reactivity compared to concurrent think aloud, i.e. the participants should not be distracted by thinking aloud. Retrospective think aloud testing also enables performance measures. [18]

In addition, also retrospective testing might produce false results as the participants may provide biased version of the thoughts that they had while carrying out the test tasks. The test participants may for example forget specific things, invent thoughts that they did not have, conceal some thoughts on purpose or modify their thoughts. There is also risk for bias in verbalising thoughts in concurrent think aloud as well but in retrospective think aloud the probability is higher as the participants have more time to reflect on their thoughts. [18]

[18] compared retrospective think aloud (RTA) and concurrent think aloud (CTA) methods in order to find out which method would be more suitable for usability testing. The study found out that test participants in RTA testing sessions had more time to verbalise their thoughts. Also as they could fully concentrate on the test system while using it they were able to also point out additional problems whereas the CTA participants had to use the system and at the same time verbalise their thoughts and hence they only pointed out problems related to their actions. In fact, in CTA 93 % of the comments made by the test participants corresponded to a problem that could have been observable also otherwise whereas in RTA 54 % of the comments pointed to an observable problem. In addition, the extra workload caused by the thinking aloud while simultaneously working caused the CTA participants to experience additional problems. Nevertheless, there was no significant difference on the types of errors found by using either CTA or RTA as 89 % of all the problems found included problems experienced by both CTA and RTA participants.

#### 2.5.2. Cooperative usability testing

Cooperative usability testing (CUT) [19] extends the retrospective think aloud testing by engaging the test user even more to the usability test session. In cooperative usabil-

ity testing the test user and evaluator are brought together via constructive dialogue in order to reveal usability issues. The dialogue is based on the expertise of both test users and the evaluators: the test users have expertise concerning the work domain and the evaluators have expertise in usability related issues. Moreover, usually the test users also have insight in alternative systems.

Usually in a usability test the aim is both to observe the test subject's undisturbed task completion for example for performance and other measures as well as gain more understanding about the test subject's opinions and reflections about the system under testing. CUT was developed to solve the conflict between the two aims. [16]

The two drivers behind CUT are context of use and the users' work-domain. As [20] mentions "products can only have quality in relation to their intended purpose". The intended purpose is defined by the context of use, i.e. the characteristics of the users, the tasks the users are going to perform, and the organisational as well as physical environment in which the product is used [20].

CUT is best used for providing deeper understanding of the users' work-domain. A work-domain is a clearly defined context of use, for example the work-domain of teachers or emergency response personnel. When designing and developing systems knowledge about that work-domain in question is crucial in usability evaluation in order to prepare and design relevant task scenarios, to understand the interaction between the test user and the system under testing, and to provide appropriate solutions for the revealed usability issues. In fact, as [16] mentions "usability problems identified by work-domain experts have been found to cause greater impact on the development of the system than problems identified by non-work-domain experts". However, using work-domain professionals for evaluating usability may become too costly not to mention that the findings do not necessarily mirror the experiences that the user might have with the system. Hence usability evaluation needs to take into account also the end users' work-domain knowledge and cooperative usability testing is one method for this. [16]

A test session in CUT consists of two phases, namely interaction and interpretation phases. The test user carries out the testing activities in the interaction phase, usually executing according to a specific usability testing method, for example think aloud. The interaction phase is then followed by the interpretation phase which is aimed at identifying the most important usability issues as well provide more insight to them. In the interpretation phase a video recording of the interaction phase is evaluated in cooperation between the evaluator and test user. The difference to retrospective think aloud and cooperative usability testing is that the users do not only verbalise their thoughts based on the video recording, instead, in CUT the interpretation phase is aimed to be a dialogue between the evaluator and the test user. Moreover, instead of just listening to the test users verbalising their thoughts as in retrospective think aloud, in CUT the evaluator leads the conversation during the interpretation phase by pointing out sequences in the video recording that should initiate a discussion about the usability issues. [19]

[19] compared CUT and think aloud usability testing by having usability professionals use both of the techniques. Also input from the test users was gathered to find out if the test users had preferences over one of the testing methods. Both the evaluators as well as test users saw the dialogue used in the interpretation phase as natural and meaningful as both of them were able to share their experiences from the interaction

phase. Both of them also saw the interpretation phase facilitating the discussion of solutions for the found usability issues. The results revealed also some weaknesses in CUT, the main one being if the discussion in the interpretation session became too general and not focusing on the interaction session. Also the interpretation phases sometimes lead to discussions that had more focus on the interaction processes and not on the interpretation of the usability issues.

The usability evaluators felt that the interpretation phases of CUT were valuable in identifying usability issues and they saw the interpretation phase as an effective tool for improving their knowledge and understanding of users and their performance. The test users provided more positive comments about CUT testing sessions than about think alouds. Moreover, the test users of CUT reported that they liked how CUT enabled them to reflect and comment upon their interactions and many of the test users found it interesting to participate in the interpretation of the test. In addition many of the test users felt the discussion of the most severe usability problems as satisfying because the discussion also provided some insight to them on why some of the test tasks proved to be difficult or remained even unsolved during the interaction phase. [19]

[16] presents an adaptation of CUT with two differences to the original CUT. Firstly, the adapted method uses interchanging interaction and interpretation phases, i.e. once one scenario is finished it is immediately followed by and interpretation phase before continuing to the next task. The purpose of the interchanging interaction and interpretation phases was to find out the test users' immediate reflections for each scenario in the test so that the participant would not have so much time for rationalising their own behaviour. Likewise, the purpose in the adapted model was also to place more emphasis on the work-domain knowledge of the test user.

In addition, in the adapted model the interaction phase reviews were based on scenario walkthroughs instead of a video recording since a walkthrough is more flexible and it provides a better opportunity for an interactive dialogue. For example the test user might have some questions regarding some UI elements that were unclear during the interaction phase. In the walkthroughs the usability evaluator initiated the discussion based on points in the interaction phase where the test user had experienced difficulties or concerns. The participant was motivated to reflect any possible causes for the difficulties as well as provide some design changes that could help solve the difficulties. Every interpretation phase in the adapted model also ended with the evaluator asking the test user a set of predefined questions regarding the test users experiences with the UI. [16]

[16] also conducted a study where the purpose was to find out how the inclusion of interpretation phases affected the outcome of usability testing of work-domain specific systems. The study made a qualitative and quantitative analysis on issues that were observed only during the interaction phase; issues that were observed during the interaction phase and then elaborated in the interpretation phase; and issues that were only observed during the interpretation phase. The results from the study indicated that the understanding of usability issues evolves between the interchanging interaction and interpretation phases. About one third of issues encountered in the interaction phase were elaborated in the interpretation phase which provided more insight and suggestions. In addition, the results noted one clear difference between the type of issues found in interaction and interpretation phases. In the interaction phase the issues were more or less concentrated on interaction design whereas the interpretation phase also

identified issues related to needed content and functionality and other requirements for use. Thus the interpretation phase adds more value to the usability testing results as together with interaction phase a more comprehensive set of usability issues can be generated. Moreover, the interpretation phases also appeared to reveal more high-priority issues than the interaction phase.

#### 2.5.3. SUXES method

One relatively new method for usability testing, called SUXES, has been developed at the University of Tampere in Finland in 2009. SUXES [21] is aimed at gathering subjective metrics from user experiences by capturing both user expectations and user experiences of the UI under testing. In SUXES the test participants fill out a question-naire before and after the use of the application. The questionnaires are named as the expectations questionnaire and the experience questionnaire. Both of them contain the same set of statements regarding the feel and use of the application. The results from the two questionnaires are then compared in order to analyse the different usability factors of the application.

By collecting also the pre-test user expectations the SUXES method provides a more user-driven context for interpreting the usability test results as the expectations can show how significant the different factors are. Moreover, the results from the expectations questionnaire alone can already provide more understanding to the usability evaluators how the test participants perceive the system under testing. [21]

In both of the questionnaires the statements are followed by a seven point scale where the test participants mark their expectations and experiences [21]. In the expectations questionnaire the test users mark two values for each statement: an acceptable level describing the lowest acceptable quality level and a desired level. These two values form the zone of tolerance. An example SUXES statement is depicted in Figure 3 [21].

		1	2	3	4	5	6	7
1	This color of this font is easy to read.		х				х	

Figure 3. An example SUXES statement where 2 stands for acceptable level and 6 stands for desired level.

After having used the application the test participant then fills out the experiences questionnaire but now they only mark one value: a value describing the perceived level of quality. Thus the gap between the expectations and experiences can be measured using two measures: the Measure of Service Superiority (MSS) that describes the difference between the perceived level and the desired level and the Measure of Service Adequacy (MSA) that describes the difference between the perceived level and the accepted level. When the experiences fall into the zone of tolerance the MSS values are negative and MSA values are positive. [21]

#### 2.5.4. Usability metrics

Usability testing can also be used for measuring usability with different metrics. These metrics include NPS, performance measures, success rate and the severity ratings.

Net Promoter Score (NPS) was presented by F. Reichheld in [22] for measuring customer loyalty and growth. In order to calculate NPS the customers are asked the question "How likely is it that you would recommend [company X] to a friend or a colleague". When NPS is used for usability calculation, the question is rephrased so that is asks how likely the users would recommend an application. NPS then tells the difference between the users who would "very likely" recommend the product and the users who are "very unlikely" to recommend. The main point with NPS is that it does not take into account the users who do not have strong feelings either way.

NPS is calculated by letting the users answer to the question on a 10-point scale where ten means "extremely likely", five is neutral, and zero means "not at all likely". As depicted in Figure 4 [22], the answers are divided into three categories, namely a) the promoters who gave ratings of nine or ten, b) the passively satisfied with ratings from seven to eight, and c) the detractors who gave scores from zero to six. NPS can then be calculated by subtracting the percentage of the detractors from the percentage of promoters. [22]

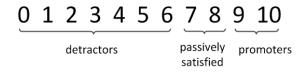


Figure 4. Detractors, passively satisfied, and promoters in NPS.

After NPS was introduced in 2003 it has gained wide popularity, despite the criticism it has faced. For example [23] declared NPS as "invalid, unreliable and non-robust". Whatever the case may be with regarding NPS's validity as a customer loyalty measure, the basic question behind it (would you recommend this product to a friend or a colleague) is a good indication of usability. It can be argued that if the test participant during a usability testing session does not find the usability as good, they would not recommend the product to a friend, and vice versa. Naturally, in order to calculate a NPS during usability testing, the test should be carried out with a sufficient amount of test users in order to calculate a valid value. For example with only three test participants there would be no point in calculating NPS. Nevertheless, the answers from just three participants can however already reveal test users opinions on how usable and/or useful they find the application under testing.

User success rate, i.e. the percentage of tasks that users complete correctly, can be used for providing numerical representation of usability [24]. User success rate provides two values: effectiveness and efficiency. However one should note as these measures only provide numerical values they do not reveal any reasons on why the users fail or how well they performed the tasks that were completed.

Effectiveness measures the "capability of the user to complete a task within the application" [24]. During the usability test session, the test monitor observes the user and evaluates the completion of each task. If the test user completes a task successfully, it

	Twell It Entirely single several real distriction precions.					
Level	Severity	Description				
1	Usability blocker	Prevents the completion of a task				
2	Major usability problem	Creates significant delay or frustration				
3	Minor usability problem	Minor effect on usability				
4	Cosmetic/subtle problem	Only requires an enhancement				

Table 1. Example single severity rating scale for usability problems.

is marked as a 'success' and the task is marked with 100 % credit. Likewise, if the user is unable to finish a task or gives up while trying, the task is marked as a 'failure' and marked with 0% credit. In addition, a task can be marked as 'partial' success based on the decision of the test monitor, i.e. if the test monitor feels like the user gave an effort and was able to complete a part of the task. A partial success task is given 50 % credit. [24]

Effectiveness can then be calculated using the following formula 1

$$Effectiveness = \frac{S + (0, 5xP)}{Total}x100\% \tag{1}$$

where S is the number of successfully completed tasks and P is the number of partially successful tasks (tasks marked as failure are ignored as they denote 0% credit). [24]

Efficiency, on the other hand, measures the "smoothness of completing a task" [24]. By denoting task execution with E, a task that was completed smoothly can be denoted with  $E_{smoothly}$ , a task not completed smoothly with  $E_{failure}$ , and a partially smoothly completed task with  $E_{partially}$  the following formula for efficiency can be derived [24]:

$$Efficiency = \frac{E_{smoothly} + (0, 5xE_{partially})}{Total} x 100\%$$
 (2)

Once again, failures are ignored from formula 2.

Another way for using usability metrics in order to analyse the usability of an application is to use severity ratings. Each problem that the user experiences during testing is marked with a severity rate that can be used for categorising and prioritising the problems. [6 p. 103]

There are two approaches for severity ratings: using a single scale or using an orthogonal scale [6 p. 103]. Table 1 (adopted after [6 p. 103] and [9 p.325]) presents an example of a single severity rating scale for usability problems.

A level 1 problem could be for example when a test participant is in a situation where they either cannot proceed anywhere or they give up trying all together because of the problem. Level 2 problems could be inconsistent terminology or unclear icons or for example if there is no feedback for the user from their actions and they are unsure if they have completed a task or not. Level 3 problems might include unnecessary icons and level 4 problems are usually the features that the user mentions as "nice-to-have".

However, as Dumas and Redish point out in [9 p.325], the test monitor needs to pay attention while rating the problems. There is a difference if a user mentions that "It would be nice if this application had the feature X" instead of mentioning "I find feature X essential for these types of applications". Based on the first comment, the

problem might be marked as a Level 4 problem where as the latter comment might result in a Level 1 or 2 problem.

In orthogonal scales the usability ratings are rated on two (or more) dimensions [6 p. 104]. The most common two dimensional orthogonal scale is where the usability problem is rated on the "frequency with which the problem is encountered by users and the impact of the problems on those users who do encounter it", as depicted in Figure 5 [6 p. 104]. Both of the dimensions in the table can be elaborated at finer resolutions, thus increasing the number of dimensions.

Severity ratings based on one test user are too unreliable. However, by using a mean set of ratings from just three evaluators produces satisfactory usability metrics for many purposes. [6 p. 103].

		Proportion of users experiencing the problem		
		Few	Many	
Impact of problem on the users who	Small	Low severity	Medium severity	
experience it	Large	Medium severity	High severity	

Figure 5. Orthogonal scale for estimating the severity of usability problems.

Performance measures provide quantitative data about the usability of an application [9 p.184-185]. Performance measures are gathered by observing the test user. The test monitor may for example log time for completing tasks, how many errors were made, and how many times an error was repeated. Other performance measures include among others (list adopted from [9 p. 185]):

- time to recover from errors,
- number of wrong menu choices,
- number of wrong icon choices,
- number of references to help or manual,
- number of observations of frustrations,
- number of observations of confusion, and
- number of times, when the test user expresses satisfaction.

#### 2.5.5. Questionnaires, interviews and diary studies

A good way to obtain the test users' subjective opinion about the application is to use questionnaires and interviews. In fact, many usability tests consists of a particular usability testing method that is then supplemented with questionnaires and/or interviews. Questionnaires are used for validating data, i.e. making sure that all the participants answer a same set of questions in order to form measurable results whereas interviews are conducted for gaining a better understanding of the results from each individual

test user as the questions can be adjusted for each participant based on their actions during the test. Interviews also enable interaction with the test participant. [13 p. 199].

Questionnaires can be used either before or after a usability test. The most common use case for pretest questionnaire is the background questionnaire for gathering information about the test user in order to help the interpretation of the results [9 p. 209]. Usually the questions in a background questionnaire include questions about the demographic information of the test participant as well as questions related to the technical profile of the participant.

The post-test questionnaires are aimed for gathering the test users' subjective opinions about the application. The questions in the post-test questionnaire can relate to any application available or to the application that the test user just finished testing [9 p. 211].

The post-test questionnaire usually comprises of a combination of open and close-ended questions. In open-ended questions the participants may answer in their own words and there is no answer limitations whereas the answer choices are limited in close-ended questions. There are many types of close-ended questions such as Likert scales, semantic differentials, fill-in questions and branching questions. In Likert scales the participants rate their agreement or disagreement with a statement, usually on a five or seven point scale. In semantic differentials the participants choose a degree to which they favour one of two adjective pairs on a scale (most often a seven point scale). In check-box questions the participants choose the answer from a list of options. Fill-in questions are similar to open-ended questions as the participants fill out the answer in their own words. However, the length of the answer is usually limited and the participant is usually asked to prioritise their answers. Branching questions are used to control the path of the participant when they are filling out the questionnaire in order to address certain questions to certain participants. [13 p. 203-205]

Another option is to use predefined usability questionnaires such as the SUMI and QUIS questionnaires. The SUMI (Software Usability Measurement Inventory) is an internationally standardised 50-item questionnaire aimed at measuring the user's satisfaction with the software via attitude rating scales [20]. SUMI questionnaire results provide a usability profile that is divided into sections by affect, efficiency, helpfulness, control, and learnability.

The QUIS (Questionnaire for User Interaction Satisfaction) is also used for measuring user satisfaction [5 p. 310]. The questionnaire consists of 12 parts that can be used in total or in parts. The parts include system experience, overall user reactions, screen design, terminology, learning, and system capabilities to mention a few.

Usability testing can also be extended with diary studies where the application under testing is provided to a set of test users. While using the application the test participants keep a diary and log for example the mistakes they made, what they have learned, and how often they used the product [25 p. 369].

One advantage with diary studies is that even with minimal analysis they can provide a source of feedback that reveals patterns that would be difficult to identify otherwise [25 p. 369]. Moreover, by allowing the users to track their own experiences, a more profound understanding of the user's experiences with the application can be achieved.

Diary studies can be divided into two groups: unstructured and structured diaries. Unstructured diaries are more participant-driven as the diary's loose structure leaves it to the test participants own judgement that which experiences they log. Structured

diaries, on the other hand, can be organised as surveys, problem reports, or even as usability tests where the test participants are asked to perform specific tasks that go through certain features of the application. [25 p. 371, 377]

#### 2.6. Number of test users

In addition to selecting the right technique for conducting usability testing, another issue to be taken into account is to choose the right amount of test participants. Generally the recommended number of test users for usability testing varies from three to five.

In [26] Nielsen and Landauer provide a formula that approximates the relationship between the number of usability issues found with the number of test participants:

$$N(1 - (1 - \lambda)^i) \tag{3}$$

where N is the total number of usability problems found,  $\lambda$  is the probability of finding any single usability problem with any single test user, and i is the number of test users.

According to [26] the typical value for  $\lambda$  is 31 % and the mean value of N is 41. Based on these values, [26] argues that the best pay-off ratio between the benefits and the costs of usability testing will be achieved with three test users. Moreover, as there is always initial, fixed costs when conducting a usability test, it is recommended to test with more than one test user in order to get the maximum amount of gain from the costs used for setting up the usability test.

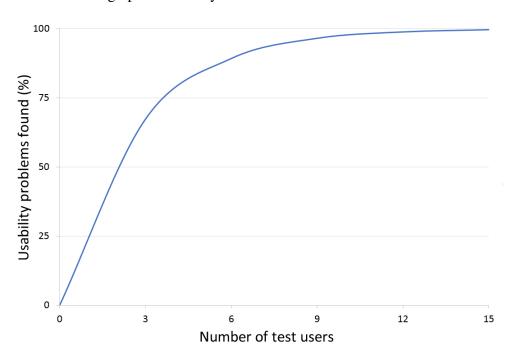


Figure 6. Probability for finding usability problems.

In addition, with  $\lambda$  equalling to 31 % [26] plotted the ratio between the number of test users and the probability for finding (unique) usability issues. Figure 6 depicts this ratio<sup>2</sup>.

The first and foremost observation from the curve in Figure 6 is that with zero test users zero usability problems are found. The second observation is the almost exponential growth between zero and one test user as with only a single test user almost third of the usability problems can be found. Then by adding more and more users the amount of new problems found decreases and after five users the the growth stabilises and closes to zero. In another words, after five users, increase in the number of test users does not result in revealing new usability issues as after five test users the same issues are observed again and again. This is due to the fact that even though each test user is individual and may reveal unique usability issues, there is always overlap in the issues that different test users will find.

<sup>&</sup>lt;sup>2</sup>The source for figure 6 is an article by J. Nielsen from March 19, 2000. In the article the curve from [26] is generalised. Article available at: http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/ (accessed 18 November 2014).

#### 3. NFC

#### 3.1. Near Field Communication

NFC (Near Field Communication) is a short range, high frequency, low bandwidth wireless technology based on RFID (Radio Frequency Identification) [27 p. 1]. Due to the short range communication the NFC interaction is based on touch, i.e. the two communicating devices have to touch each other in order to initiate data transfer. Hence "NFC can be used to replace manual typing, menu selections and other user interface actions with touching" [28].

NFC was designed to combine several technologies such as wireless communication and mobile devices and applications. As such it is one of the enablers of ubiquitous computing where computing devices are completely integrated in the everyday life of users. [27 p. 2-3]

NFC is most commonly used with mobile devices, usually mobile phones. As mobile phones have both computational and communicational capabilities in addition to a user interface they are suitable for building stand alone as well as networked NFC applications [3]. A NFC-enabled mobile phone can read NFC tags placed in the user environment and the application in the phone then reacts to the data read from the tag.

NFC was developed by Philips and Sony in 2002 as a method for contactless communication and it was accepted as a standard by ISO and IEC in December 2003. In 2004 the NFC Forum was founded in order to promote NFC usage in consumer electronics and applications by developing specifications, ensuring interoperability and educating about NFC. In addition for standardising NFC technologies, the NFC Forum has also developed a trademark for NFC called the N-Mark which can be achieved for a product after passing NFC Forum's certification tests. [27 p. 68-69, 76-78]

In addition to being an extension of RFID, NFC also uses interfaces from smart cards [27 p. 60]. The following sections will quickly go through RFID and smart cards.

RFID creates an link between the real world (world surrounding us) and digital world (applications and computer devices) [3]. The RFID system consists of a RFID reader, a RFID tag and a host computer. The reader consists of an antenna, an unit for processing and controlling the signals, a transmitter, and a receiver module and a serial data interface to the host system. The antenna in the reader reads and writes data from the RFID tag via a RF signal. The tag consists of a coupling element and a low-power IC chip, that contains the non-volatile memory where the data is stored. [29 p. 133]

There are three types of RFID tags: active, semipassive and passive. Active tags have their own battery that supplies power and semipassive tags contains a battery that powers only the IC chip but not the communication functions. Passive tags do not contain any source of power and rely on the reader in order to power its functions. Hence they are cheaper and more reliable than active tags. However, active tags have a greater read range than passive tags. Also the size of the memory in a tag is related to the power source: the greater the memory the more power is required, i.e. active tags have the most amount of memory. [29 p. 133, 155]

RFID tags are small as they most often are embedded or attached to different objects e.g. identification cards and other printed labels. The data content on a RFID tag can be either in read-only or read-write mode. The read-only tags contain a small amount of data that is transmitted when queried by the reader. Tags in read-write mode

usually contain a tag identifier in order to change data in specific tags within a given application. [29 p. 153, 155]

The functional range of RFID systems varies from few centimeters up to hundreds of meters. The operational range is determined by the used frequency band and the power limit associated with the frequency. The region where the system is used dictates the frequency bands for RFID. The used frequencies range from 130 kHz up to 5 GHz and they include low frequency (LF), high frequency (HF), ultra high frequency (UHF) or microwave bands. [29 p. 112-113]

There are three processes that are required in order for a RFID system to function, namely energy transfer, downlink and uplink. How these processes are implemented determines if the RFID system is a full duplex, half duplex or a sequential system. In full and half duplex systems energy is transferred constantly whereas in sequential systems energy is transferred first by the reader and only after that the tags respond. The flow of the data is also different in each system: in full-duplex data is transferred continuously and bi-directionally where as in half-duplex systems data is transferred in turns by the reader and the tag. In sequential systems first the energy is transferred and then by utilising the received energy the tag can receive the data, executes according to the data and then send a reply to the reader. [29 p. 182]

RFID tags are most commonly used for tracking of goods [3]. Other applications utilising RFID are for example animal tracking (RFID implants) and ski lifts that can be used with a RFID enabled key card [29 p. 58].

Smart cards are plastic cards with embedded IC that contains a memory with or without microprocessor. Smart cards were first introduced to large public as telephone cards in the 1980s and the usage of smart cards expanded in the 1990s as microchips were integrated on debit cards and with the introduction of smart card based SIM cards that were used in mobile phones. Today smart cards are used in the field of finance, communications, identification and physical access control, transportation, loyalty, and health care. [27 p. 60, 67]

A smart card system includes smart cards, readers and a backend system. The system operates at 13.56 MHz. The smart card communicates with the reader either in physical contact or in a contactless way when the card and the reader are in close proximity (less than 10 centimeters). The reader is connected with the backend system that manages and stores data. Based on their implementation smart cards can be divided into two categories: memory based and microprocessor based. Memory based smart cards can store any type of data but since they do not have microprocessor they do not have the capability to process data. Microprocessor based smart cards on the other hand contain a memory and a microprocessor so they can record, modify and process data. They also have a greater capacity for storing data compared to memory based cards. As microprocessor based cards also have an operating system, they can act as multi functional cards by combining different applications to the same card. For example the same card can be used for access control as well as payment. [27 p. 4, 58, 60-61]

#### 3.2. NFC technology

As previously mentioned, NFC is based on both RFID and smart cards. In NFC the initiating device creates a low frequency radio wave and when another NFC device

Table 2. NFC tag types.

Type	Available mem	Expandable mem	Data speed	R/W config. by
1	96 B	2 kB	106 kbps	user
2	48 B	2 kB	106 kbps	user
3	variable	1 MB	212 kbps	manufacturer
4	96 B	32 kB	424 kbps	manufacturer

gets to close proximity of the initiating device, magnetic inductive coupling transfers energy and/or data between the devices. [27 p. 69-70]

Whereas RFID can operate on a range of various distances up to several meters, NFC can only operate on short-range distances (a few centimeters). Also, the tags used in NFC are only passive in contradiction to RFID where the tags can be either active or passive. [27 p. 1]

NFC operates on the 13.56 MHz band and has a data transfer rate from 106 up to 424 kbps. Hence, NFC is fully compatible with RFID tags working in the HF band [3] as well as contactless smart card interfaces [27 p. 71].

NFC devices are categorised into readers and tags. In addition the NFC communication requires a NFC interface.

NFC tags are passive RFID tags, i.e. they do not contain any internal power source. When an active NFC reader comes into the proximity of a NFC tag, the device creates a RF signal that boots up the tag and the tag transfers its data back to the NFC device. NFC tags can contain any type of data but the size capacity is limited. Table 2 defines the different types of NFC tags available. [27 p. 75, 101-102]

It is also good to note the difference between NFC and quick response (QR) codes. QR codes are 2D barcode matrix graphics that can contain information such as URLs [30]. Like NFC tags, the QR codes can also be applied to paper and other materials and then they can be read with QR barcode readers, most commonly nowadays with a smartphone camera. However, in comparison to NFC, the QR codes need to be read by using a specific application whereas NFC operates autonomously. Also rewriting the QR codes is not so easy as NFC tags. [30]

NFC reader can be an independent device or it can be embedded on a mobile device (usually on a mobile phone). NFC readers contains their own power source so they are active devices. NFC readers can communicate with NFC tags or with each other by using a bidirectional link. [27 p. 75]

The three components in a NFC Interface are the NFC Controller, NFC Contact-less Front-end (NFC CLF) and the NFC antenna. The NFC controller enables NFC transactions by acting as modulator/demodulator between the NFC antenna and the RF signal. The NFC CLF includes the logical interface between the NFC controller and the secure element (SE) which provides secure storage for data as well as secure execution of NFC services. The SE can be implemented as embedded to the hardware, as a removable secure memory card or as a physical smart card, i.e. a SIM card. [27 p. 82, 85-86]

NFC devices can operate in three different communication modes: reader/writer, peer-to-peer and card emulation mode. The reader/writer operator mode is used between the NFC reader and the NFC tags. In this mode the NFC reader can read and

alter the data stored to the NFC tag and execute according to the retrieved data. Applications operating in reader/writer mode usually do not require a SE. Reader/writer mode applications can be used for improving mobility and hence they are suitable e.g. for remote shopping applications. [27 p. 12, 31-32]

In peer-to-peer mode two NFC readers communicate with each other by using a bidirectional half-duplex communication. Peer-to-peer operation mode provides easy data exchange and it can be used for device pairing, social networking and file transfer. [27 p. 12-13, 32]

In card emulation mode one NFC reader acts as a contactless smart card and other NFC readers can read data from it. It is mainly used for eliminating physical objects, e.g. by embedding credit cards to mobile phones. One of the advantages of card emulation is that there is no need for separate NFC tags and the stored information in the NFC reader can be used for further operations [31].

As card emulation mode uses SE most efficiently, it can be used for secure communication for example in payment and ticketing applications as well as in access control. One NFC reader can also store multiple smart card applications simultaneously. Card emulation operation mode is also compatible with existing smart card applications. [27 p. 12-13, 32]

#### 3.3. Drivers for NFC

Maybe one of the greatest advantages of NFC is that it is simple and easy to use: the users do not need to have any prior knowledge of the technology as all they have to do is bring the two communicating devices together [31]. NFC can also produce much richer user experiences, especially in context sensitive applications [28]. Hence, NFC is one of the enablers for physical browsing as it can be used for creating simple and easy to use user interfaces for applications that are embedded in the user environment [3].

Classical graphical user interfaces (GUIs) present interactive elements (e.g. menu and icons) on a screen and the user actives these elements by using I/O (input/output) devices like mouse or keypad, or most recently their fingers due to the wide spread usage of touch screens. However with physical browsing the menus and icons are replaced by touchable objects which are placed in the environment. [3]

NFC is suitable for physical browsing also due to its implicit pairing capability [27 p. 70]. The automatic pairing of NFC enables that the installed application in the NFC enabled mobile phone is immediately launched when a matching pair is found and the application can perform all the wanted processes until user intervention is required.

The most simplest use case of physical browsing would be if the user environment contains individual NFC tags which the user can read with a NFC device, e.g. a phone [28]. When the phone touches the tag it reads the command and executes according to it, for example a phone call is made or a web page is opened. The phone might also send the command to other available resources such as displays. These types of environments are called interactive spaces as the user interface is composed of physical objects and because they are based on user interaction. In interactive spaces the environment does not have to be able to predict user behaviour and needed services since it is the user who initiates the interaction.

The user control in NFC is also an advantage from a privacy point of view. Even though NFC applications are executed automatically once a matching pair is found, no actions are started until the user brings the devices to close proximity. As the user makes a decision whether to launch the NFC communication, no unwanted actions are executed without the user knowing. Another security advantage in NFC is the short range of the communication. The shorter the distance between the communicating devices, the lower the probability for signal interception becomes. [27 p. 70]

However, it is good to note that NFC, like any other technology, is vulnerable to security threats as well. Security is an important aspect especially in the use case when the NFC system stores sensitive user data. NFC security can be improved by improving the security of NFC tags, NFC readers and the used communication methods. More information about improving NFC security can be found from [27] pp. 241-282, for example.

Another advantage of NFC is that the tags can be easily re-written. Moreover, when the NFC application is in a mobile phone, the application can be remotely controlled via over the air (OTA) technology [32]. OTA provides support for remote download, management, upload and deletion of the application.

Lastly, the growing smartphone market is another key driver for NFC as out of the 345 million NFC-devices shipped in 2013 smartphones accounted for 80 % and NFC was widely used especially in the mid-tier smartphone segment where almost half of the shipped decives were NFC-enabled<sup>1</sup>. According to IDC<sup>2</sup> the worldwide smartphone market grew by 38 % from 2012 to 2013 when smartphone shipments in 2013 exceeded one billion units and thus smartphones accounted for 55.1 % of the overall mobile phone shipments in 2013.

In fact, one analysis predicted that the amount for the overall handsets including NFC could increase up to 1.6 billion by 2018<sup>3</sup>. Moreover, as analysts predict also dramatic growth for NFC adaptation in other consumer electronics such as television sets, cameras and other household appliances<sup>4</sup>, the volume of interactive spaces could increase even more.

#### 3.4. NFC applications

Today in NFC is used in many applications from various fields. [31] reviewed around 50 different commercial and non-commercial NFC applications and found that NFC was used in applications from fields such as education, data management and tracking, payment and ticketing, smartposters, health care, social networking, games and multimedia as well as guides and other tourist information applications.

According to [31], commercial NFC applications such as payment, ticketing and electronic keys mostly use the card emulation mode of NFC. However when taking

<sup>&</sup>lt;sup>1</sup>Statistics available at http://www.nfcworld.com/2014/01/08/327447/abi-smartphones-accounted-80-nfc-devices-shipped-2013/ (accessed 17 November 2014)

<sup>&</sup>lt;sup>2</sup>International Data Corporation, full statistics available at http://www.idc.com/getdoc.jsp?containerId=prUS24645514 (accessed 17 November 2014).

<sup>&</sup>lt;sup>3</sup>Analysis accessible via http://www.nfcworld.com/2014/01/21/327571/1-6bn-nfc-handsets-2018/ (accessed 17 November 2014).

<sup>&</sup>lt;sup>4</sup>Article available at http://nfctimes.com/news/consumer-electronics-continue-incorporate-nfc-pairing-and-other-connectivity/ (accessed 17 November 2014).

into account also the non-commercial applications, most NFC applications are using the reader-writer mode. Reader-writer mode is applicable to a variety of different use cases and it is relatively easier to develop applications in reader-writer mode than in the other operating modes.

Smart Posters are one of the most used types of NFC applications [33]. They are objects such as posters, billboards or three-dimensional objects that have NFC tags placed upon them. Smart posters can be used in applications such as tourist information, taxi ordering, meal ordering in restaurants or shopping assistants.

Payment and ticketing applications are another typical use case of NFC. The adaptation of NFC enabled mobile payment applications in the market however has been somewhat slow. One analysis<sup>5</sup> argued that the reasons behind this slow adaptation are the slow pace at which operators have been pushing NFC payments and the slow adoption of contactless payments by the retailers. However, various joint ventures for NFC enabled mobile payment were announced in late 2013 and early 2014 and the rollout of these ventures could rapidly increase mobile payment growth. In these ventures banks, credit card companies, mobile network operators (MNOs) and device vendors have joined forces for creating mobile payment channels across the globe in projects like the SmartPass<sup>6</sup> in Germany, La Caixa<sup>7</sup> venture in Spain, and the collaboration of Commonwealth Bank, Mastercard and Samsung in Australia<sup>8</sup>.

NFC is also used in healthcare applications for example in applications for monitoring, management, tracking, data capturing, prescription and safety [34]. NFC can also be utilised in applications for helping elderly people, such as a meal service systems described in [35] and [36].

<sup>&</sup>lt;sup>5</sup>Analysis available at http://www.nfcworld.com/2013/10/29/326568/strategy-analytics-lowers-nfc-payments-forecast/ (accessed 17 November 2014).

<sup>&</sup>lt;sup>6</sup>http://www.nfcworld.com/2013/12/13/327366/vodafone-launches-smartpass-nfc-payments-germany/ (accessed 17 November 2014)

<sup>&</sup>lt;sup>7</sup>http://www.nfcworld.com/2013/12/17/327378/la-caixa-partners-telefonica-vodafone-orange-visa-europes-largest-nfc-payments-service/ (accessed 17 November 2014)

<sup>&</sup>lt;sup>8</sup>http://www.nfcworld.com/2013/12/13/327343/mastercard-samsung-introduce-embedded-nfc-payments/ (accessed 17 November 2014)

#### 4. SEAMLESS LEARNING

#### 4.1. Background for seamless learning

In the recent years there has been a change in education where "the traditional view of the learning process as a 'transfer, remember and recall' process has been shifted towards a view that sees the learning process consisting of 'create, discover, interconnect and understand ways to apply knowledge' "as Nordmark and Milrad point out in [37]. This change is advocated by the rapidly growing integration of new and more user-friendly mobile technologies and social media into people's everyday lives. When combining this integration with an educational environment Nordmark and Milrad note that it could create "a seamless and integrated flow of formal learning activities and informal play that may result in new practices of interaction and group dynamics" [37].

Also Sollervall et al. in [1] highlight that the educational field is facing changes. The first change is in the context where education takes place as different learning activities are more commonly distributed across various contexts. The second change acknowledges the growing importance of designing different learning opportunities.

The above described changes can be addressed by seamless learning which is a learning style "where a learner can learn in a variety of scenarios and in which they can switch from one scenario or context to another easily and quickly with the personal device as a mediator" [38]. As the personal device nowadays in most cases is a mobile device, seamless learning incorporates mobile technologies tightly together with a ubiquitous learning environment. Hence the following chapters will briefly cover different aspects of mobile and ubiquitous learning before seamless learning is introduced more thoroughly.

#### 4.2. M-learning

M-learning is an expansion of e-learning where the mobile device is seen as pervasive tool in the learning context, however the definitions for m-learning still vary [3]. Most often the 'm' in m-learning refers to mobile or mobility.

Mobile technologies nowadays are an integral and well-established part of our daily lives and children today are accustomed to use different mobile devices. Thus, by adding a mobile dimension to education would allow the the young, native mobile users to participate, create and share their 'digital voice' by using a familiar tool that they consider their foremost mean for communication [37]. In another words, a mobile platform in learning enables self-expression and other learning accessibilities.

M-learning enables flexibility in terms of time and place as well as offers learners the possibility to interact with fellow students and educators [39]. As such the term mobility does not only stand for the mobility of the student but also for the mobility of the content.

Over the past 10 years the interest towards m-learning has increased [3] and many studies on how to utilise mobile technologies in education has been conducted. [40] describes different use cases for m-learning. Mobile devices can be used for augmenting the learning capability of the students by extending the learning skills such as memorisation, reasoning, deduction and knowledge sharing. Mobile devices may also

be used for enhancing the learning experience by enabling the students also to create content themselves and more value could be added to the content creation if the mobile devices are used in an ubiquitous way. Mobile devices also provide possibilities for teachers for example in evaluation and providing personalised feedback and support.

Despite the various use cases for m-learning, its adoption in the field of education has been slow [39]. In 2013 [41] studied opinions about m-learning among a student population where e-learning was a standard element and even among these students only 6.5 % of the respondents preferred to follow e-learning course via a mobile device instead of a PC. As such only 6.6 % of the respondents used mobile devices intensively for studying and third of the respondents responded to never have using mobile devices for learning purposes. However, almost half of the respondents agreed to have occasionally used mobile devices for m-learning and 63 % of the respondents believed that m-learning will have an important role in the future for learning purposes.

In another study [42] only 4% of the respondents used only their mobile devices for accessing learning materials whereas 45 % of the respondents solely used PCs for accessing learning materials. 49 % of the respondents used either PCs or mobile devices. According to the study, there is also a clear need for m-learning from the students perspective as almost 70 % of the respondents said that mobile applications are an important part of their day to day lives.

Reasons for the slow adoption of m-learning include concerns toward security and privacy, educators not having sufficient amount of understanding on how to utilise the technology and some educators might still be uncomfortable with the technology itself. In addition, from the learners' perspective one of the major barriers for adopting m-learning is the price of the mobile devices. [39]

Also [41] mentions the price and privacy concerns as one of the weaknesses of mlearning. Another hinderer for wider m-learning adoption is the various operating systems and devices in the mobile market that make it difficult to create a unified solution for m-learning.

[39] did a meta analysis of the critical success factors (CSF) for m-learning based on 19 studies in order to find out from the learners' perspective what are the factors affecting m-learning usage. Based on the results, learner's perception was the most critical success factor i.e in order for m-learning to gain more popularity among learners it needs to have favourable user perception. Other CSFs for m-learning were content, user friendly design, flexibility, and the ability to connect with other learners and educators.

[39] goes also through the benefits of m-learning. The listed benefits include high efficiency, flexibility, interactivity, portability and the fact that it enables rich content creation and sharing. In addition, m-learning allows the learner to choose the time and place of studying as the users can access material remotely as well as store the materials to their mobile devices, or nowadays to the cloud. It is also easier to create applications for a mobile platform than for desktop systems as the platforms are open and do not have the complicated characteristics of desktop operating systems. Also mobile devices can connect to the internet via other networks than LANs and WLANS as they support mobile networking via 2G, 3G and 4G networks.

# 4.3. U-learning

Ubiquitous learning (u-learning) can be seen as an extension to m-learning as it enables learning anytime and anywhere through wireless networks by incorporating context aware technologies and m-learning [43]. In another words, m-learning is transformed into u-learning by letting the learners interact with their learning environment through incorporating the learning environment with learning objects than can detect and collect information about the learners' location.

[44] on the other hand defines u-learning as a learning environment that provides an interoperable, pervasive, and seamless learning architecture that integrates learning collaborators, learning contents, and learning services together. More specifically, u-learning provides the right collaborators, contents and services in the right place at the right time. Hence, the effectiveness of u-learning depends on the context where the learning takes place.

[44] defines the learning context in u-learning from two perspectives: the learners and services. The learners' context is defined by the surrounding environment that affect the learners' preferences and the services' context is defined by the environment that affects the delivery and execution of different learning services.

[43] describes a few examples of u-learning applications. One of them is a context aware ubiquitous English learning system with which the students could associate English course materials with their surroundings in a campus environment. Another example was based on RFID. The learners could identify learning objects in their surroundings and use them for their learning activities. Both of the applications yielded positive results in terms of improved learning efficiency and interest.

#### 4.4. NFC in mobile and ubiquitous learning

As NFC incorporates both mobile and context aware features it is a suitable technology for enabling m- and u-learning applications. As usually a NFC based learning application incorporates both of the features, the following examples do not differentiate NFC applications for m-learning from NFC applications for u-learning.

Mostly NFC has been used as a tool for attendance monitoring in all education levels from primary schools [45] to universities [46], [47]. In one example<sup>1</sup>, NFC was utilised in campus ID cards, where the NFC based campus cards combined access control with meal purchases and copying, laundry service and printing. Also [48] and [49] describes different scenarios how NFC could be utilised in the the university environment. The described scenarios include information retrieval about courses and contact information for the faculty staff as well as administrative applications e.g. for scheduling appointments or enrolment for courses and exams. In the university environment NFC could also be utilised for accessing bibliographic sources [50].

All the above use cases use NFC for aiding and assisting activities related to learning. None of them however specifically utilise NFC for educational and learning purposes. [51] describes 4 different applications for NFC enabled learning. The first two applications, Touch & Learn Languages and Touch & Learn Reading have the same

<sup>&</sup>lt;sup>1</sup>Article available at http://www.nfcworld.com/2013/11/04/326670/us-universities-test-nfc-campus-cards/ (accessed 17 November 2014).

principle: NFC tags having information that support the learning are placed in the environment and the students can use the tags in two modes: explore and practice. Touch & Learn Languages was aimed for learning vocabulary of a foreign language and Touch & Learn Reading was aimed for three-to-five-year-old children for supporting their efforts to learn how to read. The third application, Touch & Share, was aimed at children between the ages of 9 and 13 and the application was used for biology learning so that the NFC tags were placed in a zoological museum and the children used the application in order to learn about the animals. The fourth application, Touch & Run was a NFC application encouraging children to do physical exercise as well as support teamwork.

Other NFC-based applications for learning can be found e.g. from [52] where a NFC-based learning environment targeted to children between the ages of 3 and 8 was described and from [53] where a game based on NFC was used in a university environment for supporting learning motivation. Another example comes from a British college, where the students whose native language was not English used NFC for learning English<sup>2</sup>. Smartposters were placed on different objects around the campus and the posters contained the English definitions of the objects. The students found the interaction with NFC more pleasant as compared to QR codes as they did not need to scan the tags. Also the teachers preferred the NFC tags over QR codes as the tags could easily be re-written.

Using NFC in m-learning could also expand the popularity of m-learning. [41] lists way of navigation and usability, content organisation and social interaction as drivers for more wider m-learning adoption.

Way of navigation and usability are important because if the navigation methods in the UI are cumbersome the time spent on actual device usage is always away from time spent on the actual learning activities. NFC enables fast and easy-to-use navigation. With content organisation the learning content is divided into meaningful and complete units and NFC provides fast access to these units. Moreover, when the NFC tags are placed in the learning environment, NFC also helps in providing quick associations with the content units hence enhancing the learning experience. NFC also enables the social interaction aspect of m-learning as the students can share information between their NFC phones.

# 4.5. Characteristics of seamless learning

"The basic rationale behind seamless learning is that it is not feasible to equip students with all the skills and knowledge they need for lifelong learning solely through formal learning" [2]. Instead, learning should move towards enhancing the learners' capability to learn seamlessly in both formal and informal learning environments.

Formal learning is traditionally regarded as pre-planned learning that takes place in a formal setting (e.g. a classroom). Formal learning is dictated by a curriculum that is followed in textbooks and other learning materials provided for the students. Correspondingly, informal learning has been associated with unplanned and incidental learning that happens most often outside formal learning. The main idea behind seamless learning is to bridge the gap between formal and informal learning. [54]

<sup>&</sup>lt;sup>2</sup>Article available at http://www.nfcworld.com/2013/09/10/325847/uk-college-uses-nfc-help-teach-english/ (accessed 17 November 2014).

By integrating formal and informal learning together via the use of a mobile device, seamless learning can be characterised as the third generation of mobile learning [2]. The first generation of m-learning was about transferring learning content onto mobile devices and the second generation focused on pedagogical design. In the third generation of m-learning, i.e. in seamless learning, the focus is on increasing the learners' capability to move their learning environment as the learners themselves move from place to place. Hence, seamless learning enables a seamless flow of learning across different contexts and it makes the learning experience more personalised for the learner.

It is also good to note that m-learning and u-learning are only aspects of seamless learning and they should not be treated as synonyms. Where m-learning is mobile enhanced learning and u-learning is about learning in the right way, in the right place, and in the right time, seamless learning is a mindset for the learners (as well as for teachers) which may or may not need to be mediated or supported by technology [2].

In addition to m- and u-learning another important aspect defining seamless learning is the notion of 1:1, 24x7 setting for learning [2]. Originally the 1:1 notion referred to the ratio of one computer to one student but with pervasive use of mobile technologies the meaning of 1:1 has now switched to one learner per one (or more) personal computing device that is mobile, enables multimedia input and output, and that is wirelessly connected. As with seamless learning the learning takes place also outside of school the 1:1 notion is elaborated with 24 (hours) x 7 (days) i.e learning can take place anywhere, any time. [2]

The mobile device (or devices) play a vital role in seamless learning. In fact, a device that is available for the learner 24/7 and that integrates all the personal learning tools, resources and self-created artefacts is seen as a 'learning hub' in seamless learning [38]. By using the learning hub the learners can then manage their own seamless learning across different learning activities and contexts. A learning hub also does not need to consist of physical hardware as nowadays the hub can exist in a cloud-based service.

[2] did a comprehensive literature analysis in order to find out the common features of seamless learning. The analysis was based on 54 papers about seamless learning that were published between 2006 and 2011 and differentiated 10 features. Table 3 (adopted after [2]) lists the features<sup>3</sup>.

The following section presents the different MSLs according to [2].

The difference between formal and informal learning has already been discussed but it is good to note that [2] highlights with MSL1 that mobile communication and and online communities should not be seen as a threat to formal education, as they should be treated as means for transforming learning into the mobile age through a dialogue between formal and informal learning environments. MSL2 likewise calls for a balance between individual and social learning. Social learning in its many forms (such as small groups and peer review work) is important but educators and learners alike should not forget the importance of being able to work also alone.

As has already been defined, seamless learning can happen any time (MSL3) and anywhere (MSL4). However it is important to note the difference with learning 'anywhere and any time' with learning 'everywhere and all the time'. The goal of seamless

<sup>&</sup>lt;sup>3</sup>MSL stands for Mobile-assisted Seamless Learning

Feature Characteristic MSL1 Encompassing formal and informal learning MSL2 Encompassing individual and social learning MSL3 Learning across time MSL4 Learning across locations MSL5 Ubiquitous knowledge access MSL<sub>6</sub> Encompassing physical and digital worlds Combined use of multiple device types MSL7 MSL8 Seamless switching between learning tasks MSL9 Knowledge synthesis MSL10 Encompassing multiple pedagogical or learning activity models

Table 3. Seamless learning features

learning is to support learning whenever and where ever the students are stimulated to learn, not to require students to learn all the time and every where. Equally important is to note that seamless learning cannot always happen simultaneously anywhere and whenever, i.e. seamless learning also can happen in predefined times and at selected locations as many systems today still have requirements and limitations that may not allow completely learning by the 1:1, 24x7 notion.

MSL5 is a combination of context aware learning, augmented reality learning, and ubiquitous Internet access. In another words, MSL5 calls for pushing and/or pulling information from ubiquitous learning resources such as materials supported by the teacher, student cooperation and the Internet. MSL6 then calls for a balance between encompassing digital and physical worlds for example by making use of sensors, smart rooms, and ambient environments that capture information from the physical world and transform it into a digital form.

The main idea in MSL7 is to encourage the use of multiple devices that complement each other in order to fully enable seamless learning. For example a mobile device support seamless learning while the learner is also moving but a laptop or a PC supports more efficient learning in fixed locations. Likewise, physical hardware and cloud-based services complement each other.

Whereas MSL7 calls for switching between different devices, MSL8 calls for switching between different task types, like data collection, analysis, and communication. Not only do different task types make the use of different devices more efficient, MSL8 also provides the learners with multiple opportunities for reflecting the transitions between different activities and interests. This way the learners could integrate their ideas with the collected data and observations.

MSL9 happens in three different forms: the synthesis of prior and new data; the synthesis of different thinking skills; and the synthesis of multi-disciplinary learning. MSL9 is important in order for the learner to acquire data and knowledge in different domains and forms as well as in order to record, organise, process, and reflect upon the knowledge. Lastly, MSL10 calls for a seamless flow of different learning models and activities such as self-regulated learning, collaborative learning and activity learning.

The above described features can roughly be divided in three subgroups by technological focus (MSL5 and MSL7), pedagogical focus (MSL8 and MSL10) and learner

focus (the rest) [2]. In another words, by combining the technical resources (i.e. MSL5 and MSL7) with pedagogical means (MSL8 and MSL10) to support learners' learning models (MSL9) by stimulating learners with new information anytime (MSL3) anywhere (MSL4) within any context (MSL1, MSL2 and MSL6). This way the different MSLs enable holistic seamless learning. [2]

# 4.6. Seamless learning applications

#### 4.6.1. Authoring tools

One way to expand the use of seamless learning is to provide authoring tools for the teachers. Seamless learning calls for the integration of new technologies and methods into traditional educational environments. However, due to the complexities of educational policies as well as other practical issues, the integration faces challenges [37], namely, technical, pedagogical and theoretical challenges.

The technical issues include acquiring the right equipment, infrastructure and support whereas the pedagogical issues include the renewal of the curriculum and teaching methods in order to incorporate the new technologies. Pedagogical issues also revolve around questions like how to evaluate the learning gains and how to deal with copyright, ethics and other digital content related matters. Theoretical issues then are about how to remodel the current approaches to learning and teaching. [37]

In order to overcome these challenges the teachers should be provided authoring tools for exploring and experiencing the new technologies and learning models [37]. An authoring tool is a system that enables teachers to create digital learning systems [55]. The tools should be easy-to-use so that even novice computer users are able to create educational content with them. In addition, the tools should be flexible so that teachers can easily modify and import previously created educational content into the tools [56].

Moreover, the authoring tools should also be quick and efficient to use. The work-load for teachers is constantly growing as the teachers are given more and more tasks to complete but the amount of pupils nor the demands on the teaching hours are not compensated with the new tasks [37]. When the use of authoring tools is not time-consuming the teachers are provided with sufficient time to incorporate the new teaching models and technologies better into the everyday teaching activities thus improving the quality of teaching.

Educational authoring tools can be divided in two categories: systems for creating digital content and systems for pedagogical scripting [55]. The tools aimed for content creation are used for creating content for lessons and courses whereas pedagogical scripting tools are also used for planning the courses and their pedagogical strategies.

There are also many authoring tools meant for collaborative content creation. These tools enable interaction among content creators. However it is good to note that some collaborative tools only enable content reusing, i.e. users can search and use content created by others whereas other collaborative tools are aimed at collaborative content creation, i.e. the users create and enhance content together by combining their knowledge and insight. Either way collaborative tools can improve the process of authoring educational content as well as enhance the quality of the created content. Major hin-

derer for collaborative content creation is that the different authoring tools used today are not interoperable i.e. each systems produces specific content that cannot be reused in another system. [57]

One example of authoring tools is a tool aimed at learning Jawi language [58]. The authoring tool was developed in order to help the teachers to transfer manual games used for language learning into games that could be played on a computer. In this tool the teachers made use of ready-made game templates that describe a particular type of a game. These templates dictated the type of game artefacts and rules that could be used in the game and the teachers could then just add content to the game artefacts.

There are also authoring tools meant for creating m-learning content. One example is [56] which is an authoring tool used on top of Microsoft PowerPoint. The teachers could use the tool for creating mobile content as well as analysing the usage of the created mobile content. Another example is [59] which is a web-based m-learning system that enables content creation for teachers as well as study supporting tools for students.

Activity Pad [60] is an authoring tool aimed for creating content for NFC enabled m-learning. The Activity Pad is a pad that combines an A4-sized sheet of paper with 24 programmable NFC readers. The paper placed on top of the NFC pad represents the UI of the application and the pad is configured by placing tangibles, i.e. physical objects of the learning environments with NFC tags, on top of the paper. This way the Activity Pad supports teacher-driven content creation as the tangibles can be configured to be application-specific. The student then places the sheet of paper on top of the pad and places the tangibles on top of the paper in order to solve the problem or exercise that the teacher has created. The pad reads the NFC tags and reacts accordingly. The pad can also give feedback for the student by flashing LEDs or playing sounds via an internal speaker. The pad is mostly aimed at children between the ages of six and eight as well as children with learning disabilities.

#### 4.6.2. Mobile assisted tools for self-directed seamless learning

[43] argues the importance of self-directed learning in seamless learning environments. Self-directed learning enables the students to learn according to their own pace. Moreover, self-directed learning also provides the students with the possibility to obtain answers to questions immediately and the possibility the discuss the issues with other students immediately through different channels like social media and mobile devices. [43]

Mobile devices advocate self-directed seamless learning as they are portable and versatile. Thus they enable the students to become active participants instead of passive receivers in learning activities [61].

There are many reasons why some learners prefer self-directed learning. [43] lists the following:

- Learners want to adopt their own learning pace.
- Learners want to match the learning with their own learning style.
- Learners want to keep learning free.

- Learners want to put their ideas into their learning plan.
- Learners are unsure if regular courses will suit them.
- Learners want to learn according to their own schedule.
- Learners are unsure if traditional learning environment will suit them.
- Learners have personal reasons why not to attend traditional courses.

However, self-directed learning does not exclude the teachers from the learning scenarios. Instead, teachers are needed in self-directed learning in order to provide the learners suitable assistance according to the each learners learning motives, prior knowledge and learning character. Moreover, as each learner has an individual capability for adopting self-directed learning, the teachers are needed for providing the students with sufficient amount of assistance in order to suit the different levels of adopting self-directed learning. [43]

[43] describes an example of a mobile assisted self-directed seamless learning tool that divides the learning process into three strategies, namely arrangement strategies, teaching strategies and estimation strategies. Arrangement strategies occur before the actual learning and include the building of a stimulating learning environment, establishing learning motivation and responsibilities, and setting clear goals. The teaching strategies include that the teachers assist the learners with the different arrangement strategies as well as provide a setting for interaction and feedback among the learners. The estimation strategies then provide the means for self-evaluation and sharing the learning experience. The tool was tested in a university campus environment with 43 students and from the results it could be seen that most of the participating students were satisfied with the self-directed learning activities and that the students felt that they had impact in their own learning pace.

[61] describes two case studies of mobile assisted seamless learning tools aimed for language vocabulary learning in a primary school. The other tool was aimed at learning English prepositions by letting the students to take photos with their handheld device. In the photos the students "acted" out different prepositions in their surroundings. The other solution was aimed at learning idioms. The students were encouraged to take photos with their smartphones of their daily lives so that the photos depicted the learned idioms. The students then posted the photos to a wiki-page for peer review. Both of the tools place the language learning in the real world setting in order to emphasize the proactive association of language vocabulary in the context where they are used.

[38] continued the work presented in [61] by inputting the two language learning use cases<sup>4</sup> through a facilitated seamless learning (FSL) framework that consists of four processes: learning engagement, personalised learning, online social learning, and in-class consolidation.

Activities in learning engagement usually set up the learners for the up-coming activities in the following processes and they usually are teacher-driven. Examples of learning engagement activities include classroom instructions and group activities. In personalised learning process the individual learners observe, record, draw connections

<sup>&</sup>lt;sup>4</sup>The preposition learning case study was incorporated into a bigger 2-year seamless learning project that consisted of multiple seamless learning activities.

and reflect upon their daily encounters and associate those with their prior knowledge and then during social online learning the learners carry out peer reviews and discussion online. In-class consolidation then enables group or class-wide discussions that is facilitated by teachers. [38]

Based on the results in [38] the students learning motivation was improved as they saw the learning activities carried out according to the FSL framework more as a fun activities instead of seeing them just as school assignments. This notion was reinforced by the positive feedback given in the post-study surveys as well as by improved exam results.

[38] also provided guidelines for a more effective adaptation of self-directed seamless learning, namely learning processes should be developed so that they aim to change the mindset of students; so that they reuse created learning artefacts; and so that the students should be able to carry out learning activities without teacher assistance.

The first point should encourage the students to carry out more open-ended learning activities also outside of school assignments based on their own interests. The second point emphasises the importance of linking different learning activities together by re-using previously created content for new activities and the third point encourages teachers to conduct more activities for supporting also self-guiding learning skills. [38]

# 4.6.3. Seamless learning applications for collaborative learning

Collaborative learning is another important aspect of seamless learning. Collaborative learning is a process in which problem solving is supported from multiple perspectives and where the results do not necessarily need to be pre-defined [37]. Usually collaborative learning is coupled with creative tasks in order to emphasise the importance of creativity and innovation as facilitators for learning. The pairing of collaborative learning with new technologies could create new practices of interaction and group dynamics.

[37] describes a use case where digital storytelling was combined with collaborative learning. The basic concept behind digital storytelling is to provide the students a way for conveying a story using digital video editing tools. In the case study primary school students were divided into small groups and each group was equipped with a mobile device (iPod Touch) and their objective was to capture and create their own stories in a local museum. From the results the study highlights most the sheer joy and motivation that the students showed when carrying out their task. The students expressed joy in using the technology for a school assignment and the collaborative aspect of the task genuinely boosted the motivation and work-moral of the children. Also the participating teachers and museum personnel agreed that the children's level of motivation and participation exceeded all their expectations.

[62] on the other hand goes through a more general framework that provides assistance for collaborative learning in a seamless learning environment. The framework combines different hardware devices (e.g. smartboards, smartphones and desktops) with different communication infrastructure (i.e. physical spaces and data transfer between devices) and software (including a specific toolkit for teachers).

# 4.6.4. Educational games

Mobile and social networking games are nowadays becoming as popular as traditional PC and console games. In 2014 46 % of teens played games on mobile devices and 71 % used dedicated gaming devices. In the US alone 36 % play games on their smartphone and out of the 91 % of all people who have a mobile or smartphone spent 80 % of their time inside applications or games.<sup>5</sup>

The trend towards mobile gaming is also appearing in education. Mobile games are well-suited to be integrated with education as they enable seamless evaluation of student performances if the game is tied together with in-game evaluation of the learning outcomes. Also, games do not restrict the learning (i.e. gaming) to a specific location. These features also make mobile gaming a part of seamless learning. [63]

One of the advantages of educational mobile games is that mobile devices include additional hardware, such as GPS sensors, accelerometers and NFC readers, that normal PCs do not have. The games can make use of the hardware and improve the learning situation. [63]

In addition, mobile games provide a seamless way for changing the context which can increase the time used for playing the games and thus improve the learning outcomes. Moreover, wireless connectivity enables using the mobile games for in-class evaluation or as support material e.g. for lab or field activities. [63]

However, developing mobile games introduces new challenges especially due to the variety of different mobile platforms available. "Each platform introduces particular user interfaces, interaction paradigms and specific development tools – and different game versions must be developed for each target mobile platform" [63]. Not only do these new technical requirements increase the development costs but they also limit the possibilities for educators to create their own mobile gaming content.

To address the technical issues, [63] proposes an authoring tool tailored for educators for editing games that allows both adapting pre-existing educational games to mobile games as well as the creation of completely new educational games.

[64] on the other hand goes through the results of a study where an educational game was introduced to a mobile vocabulary learning system. Using the system the students could log newly learned vocabulary to the system and then the game provided the students with quizzes using the words. In order to retain the vocabulary even better the quizzes also included also "older" vocabulary content. The system then showed each student their current degree of advancement, and using the system the students were also able to see what vocabulary their fellow students were learning. Even though the results indicated no clear statistical difference for whether the vocabulary learning was more efficient using the game, at least the game added a fun factor for vocabulary learning thus increasing study motivation.

<sup>&</sup>lt;sup>5</sup>Statistics available at http://www.bigfishgames.com/blog/2014-global-gaming-stats-whos-playing-what-and-why/ (accessed 17 November 2014).

# 5. NFC-ACT APPLICATION

# 5.1. NFC-ACT description

NFC-ACT is an authoring tool aimed for creating NFC enabled learning games and exercises for mobile phones. NFC-ACT provides a web-based GUI with which the teachers can create games, write related game information to the NFC tags and upload the game to the NFC enabled phones. NFC-ACT is being developed by the Interactive Spaces research group at the University of Oulu.

The major driver behind NFC-ACT is the lack of tools for creating engaging mobile content for educational and learning purposes. As [41] mentions, the lack of the authoring tools is one of the main hinders for wider m-learning - an thus also seamless learning - adoption.

NFC-ACT is targeted for teachers who desire to create their own applications to be used by their students. Due to the yet relatively simple form of games that can be created using NFC-ACT, the main focus group today is kindergarten and primary school teachers. However, should the future versions of NFC-ACT enable more complexity to the game creation process there is no reason why NFC-ACT could not also be used by teachers in higher education levels.

The first versions of NFC-ACT were developed for creating games for Nokia mobile phones running the Series 40 3rd edition OS but the upcoming version (which is currently under implementation) is supporting game creation for the Android OS version 4.1.2<sup>1</sup> but the games should support also earlier Android platforms from version 2.3.2 and onwards.

All games created using NFC-ACT follow the same basic principle, namely, the game in the phone proposes challenges that are solved by touching NFC tags with the phone. The NFC tags are placed in the learning environment, e.g. on furniture in the classroom or different objects in the school environment.

Currently NFC-ACT provides two game patterns, explore and practice, that dictate the game flow. The teachers cannot change the game flow of the two patterns. Future versions of NFC-ACT could also support more game patterns or even enable the teachers to create their own game patterns, if demand for such feature is seen.

In the explore mode, the players are instructed to find different tags (i.e. challenges) that are related to the game and touch them with their phones. After a tag is touched the game in the phone provides feedback. For example the students could search for NFC tags that are attached to post-it notes showing names of animals in a foreign language and upon touching a tag, the phone shows the corresponding image for each animal. In the practice mode on the other hand, the game flow requires the students to solve a challenge in order to proceed with the game. For example, a game in the challenge mode could display the picture of an animal and the student has to touch the corresponding tag in order to proceed to the next animal.

In another words, the game flow in the games created with NFC-ACT depends on the selected game pattern. In the explore mode the game waits until a tag is touched, provides the feedback according to the tag and then the game proceed to wait until the next tag is touched. If an unrecognised tag is touched, the game provides feedback

<sup>&</sup>lt;sup>1</sup>a.k.a Jelly Bean (API level 16), http://www.android.com/versions/jelly-bean-4-3/ (accessed 18 November 2014)

Table 4. Game responses after touching a tag.

Action	Next action
Correct tag is touched	Game proceeds to the next challenge
Wrong tag is touched	Game remains in the current challenge
Unrecognised tag is touched	Game remains in the current challenge

about that as well. In the practice mode the game provides a clue of some sort with which the student needs to find the correct tag. Once a tag is touched the game displays feedback and proceeds according to actions listed in table 4.

In both of the modes the feedback can be in textual, visual or audio format; or it can be a combination of them. Both modes also support timer events, i.e. the game can proceed from one challenge to another if a specified timer expires.

Basically, the game in the phone consists of different screens that are shown for the user based on their actions, i.e. touching different tags (or timer activated events). The teacher uses the NFC-ACT to configure the different screens as well as for configuring which tag is associated with which screen. The first step in the game creation process in NFC-ACT is to select the game mode. Then each challenge in the game is created: the user configures the screens that show the challenge and associate each challenge with corresponding feedback for right, wrong and unrecognised tags. The user can also specify if the game flow shows the challenges in a specified or random order. After the game creation is finished, the user then writes the challenges to the NFC tags using a NFC reader and the last step is uploading the game to the phones. The UI in NFC-ACT is divided into three screens, namely, game creation (i.e. main screen), tag writing and game uploading (i.e. deploy screen).

As can be seen from Figure 7, the game creation screen in NFC-ACT is divided into three parts. The left column is used for creating the game and adding challenges to the game. User can then select a certain challenge from the challenge list in the left column and it is then opened in the middle column that represents the flow of the challenge on a grid that visualises screens, events and feedback related to that challenge. Once the user selects a certain screen from the grid, it is opened in the right column of the UI where the user can specify the wanted content (text, image, audio) with the corresponding screen.

In the first version of NFC-ACT it was also possible to modify a "general challenge" <sup>2</sup>. The idea behind it was that before modifying each challenge individually the user could modify the general challenge from which the content was automatically copied to each individual challenge. For example if the user wanted to associate for each challenge the same screen for a correct answer, they could define the screen in the general challenge screen instead of setting the screen individually for each challenge.

In the top-left corner of the UI there is a button that enables switching between main screen, tag writing screen and the deploy screen. In order to write the tags the user opens the tag writing screen, connects a NFC reader to the computer and places a NFC tag to the reader. As can be seen from Figure 8, the tag writing screen shows all the available challenges (represented as tags) in the left side of the screen and on the right

<sup>&</sup>lt;sup>2</sup>This feature was in fact dropped out from the upcoming version of NFC-ACT. It is however mentioned here as it was still available in NFC-ACT when the demo sessions were carried out.

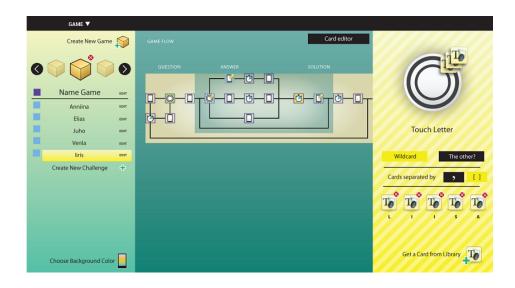


Figure 7. The game creation screen in NFC-ACT.

side of the screen a NFC reader is displayed. First the user places a tag to the reader. Then by using the drag& drop UI of the tag writing screen the user can simply drag the wanted tag and place (i.e. drop) it to the reader. As soon as content is placed to the reader in the UI, the information is written to the tag automatically. User can then place another tag to the NFC reader and repeat the same steps for each tag.

The same drag & drop functionality is also used in the deploy screen where the user can transfer the game to the phone via Bluetooth, see Figure 9. By dragging a game icon over a phone in the UI, the computer creates a Bluetooth connection with the phone and transfers the game to the phone. The newest version of NFC-ACT is planned to use WiFi connection for uploading the game to the phones.

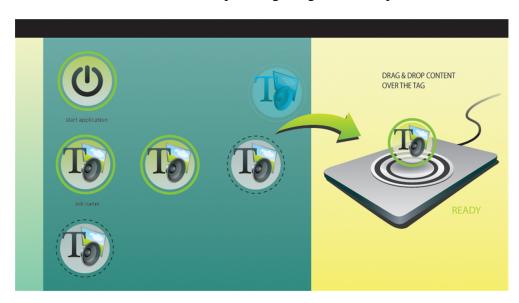


Figure 8. The tag writing screen in NFC-ACT.

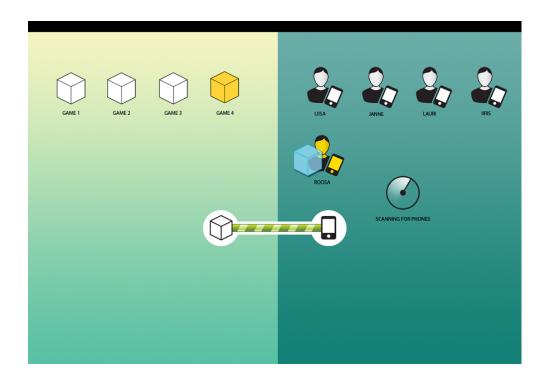


Figure 9. Uploading the game to phones.

# 5.2. NFC-ACT as a seamless learning tool

In addition to being an authoring tool, thus embodying seamless learning, the games created with NFC-ACT also follow the seamless learning features (MSLs) described in chapter 4.5. Games created using NFC-ACT combine formal learning in informal settings (MSL1) and the games enable teacher-driven learning in groups and/or individually (MSL2). Even though NFC enabled games can only be played in pre-defined areas and usually they are played according to a pre-defined schedule, NFC-ACT still follows MSL3 and MSL4 as the time and location for the games solely depends on the teachers decision, i.e. the teacher can also configure the games so that they allow learning across time and across different locations.

Currently, the games created with NFC-ACT only work on a specific mobile platform, but the games can be designed so that they fulfil the requirement for multiple device type usage (MSL7). For example the students could collect information with the NFC game that they can re-use on another device.

Games created with NFC-ACT also enable MSL8 (switching between different learning tasks) as the games combine data collection and analysis. Also when the challenges in the game are carried out in small groups, the games also enable communication. Moreover, the games also enable synthesising prior knowledge with new (MSL9) either by requiring prior knowledge in order to solve the challenges in the game or by requiring that the challenges are solved in order to continue with the ongoing learning activity.

In addition, MSL10 requires the deployment of different learning models like selfguided learning, collaborative learning and activity learning. NFC-ACT enables all of these as the games can be designed so that they support one or all of the mentioned learning models.

Lastly, MSL5 highlights the importance of using context-aware technology for which NFC games are more than suitable, and MSL6 requires combining physical world with the digital world. By placing the NFC tags in the learning environment, NFC games stimulates the students to establish relations with the game and the environment. Moreover, games created using NFC-ACT can also combine the digital content from the game to a physical world learning activity, for example, the students need to collect data from the NFC tags in order to complete a reading comprehension exercise.

Also [1] points out the importance for providing the students with learning activities that combine the physical and digital world via the help of context-aware technology by mentioning that when carrying out such activities, the students are able to learn and explore a topic in authentic setting, reason and argument in order to come to the solution of a problem, as well as collaborate and reflect upon relevant matters and support abstract thinking.

By using NFC-ACT in order to create games that provide the students with the above mentioned possibilities, the teachers advocate the use of seamless learning even more.

### 5.3. Current state of NFC-ACT

In March and April 2013 two demo sessions were held where the purpose was to test the usability of the current version of NFC-ACT at the time being. Hence, the following section goes through the different usability aspects of NFC-ACT at the time being as that was the version used during the time of the testing. The findings from the demo sessions were then used as a basis for the modifications implemented for the upcoming version of NFC-ACT.

# 5.3.1. NFC-ACT usability demo sessions

The main focus of the demo sessions held in 2013 was on usefulness and time consuming aspects of NFC-ACT. Satisfaction with the software and future improvement ideas were other focus areas as well as ease of learning and ease of use. There had been one previous usability test session in January 2013 and based on the feedback, changes to NFC-ACT were still being implemented when the two demo sessions were held. This caused some bugs in the software during the session.

During the demo sessions participants were asked to create a game using NFC-ACT. In the first session the participants created a game using the exercise mode and in the second one a game was created using the practice mode. The participants had created the content for the games before the demo sessions.

The main usability evaluation method used in the demo sessions was to observe the users. However, in contrast to normal user observation the participants were provided a short introduction to the editor itself and its concepts (flow of the editor, what is a challenge, practising mode vs. exercise mode etc.) and they were provided help if they required some. The sessions were then concluded with a discussion with the participants

and after the first session the participants also responded to an online questionnaire. In the questionnaire the participants were asked to rate agreement on statements about the user experience of NFC-ACT on a 5-point Likert scale, ranging from strongly agree to strongly disagree. There were also a few open text questions in the questionnaire. Both of the sessions lasted about an hour and they were recorded with an audio and a video recorder.

The participants of the demo session were five English major students with four or five years of studying. All of the participants were males with ages between 21 and 29 years and all of them have had educational or pedagogical studies as they were going to be teachers. All but one had also prior teaching experience. Moreover, they did not have a high technical profile as they did not have prior experience from other edition tools than Microsoft Office, Photoshop and MovieMaker. Also NFC as a technology was not familiar to them. As such they were suitable candidates for the demo session.

The participants participated to the demo session because it was a part of their course work where their goal was to create future concepts for teaching and come up with ideas on how to utilise them in the future. Before the first demo session an introduction meeting was held with the participants where the concept of NFC-ACT was explained to them. The participants were then asked ideas about what type of games they would like to create with NFC-ACT. The participants had many ideas, however most of them were too elaborate for NFC-ACT in regard of the state of the software at the time being as well as restrictions that the NFC tags themselves had.

After a discussion, it was decided that that the participants idea about a listening comprehension targeted for schoolchildren between ages of 13 and 15 years would be the best option for the first demo session. The game would work so that each of the tags would contain an audio file where parts of a story is read aloud and the tags would be placed somewhere in the school environment in a random order. The children would then have to search the tags, touch them with the phone, listen to the audio and write it down. After having collected all the audio files, the children would then have to arrange them in the correct order in order to form a coherent story. It was then decided that the participants would prepare the material (parts of the story as well as the audio files) ready for the demo session and they were asked to think in advance that what would be written to the tags and what would be shown in the phone.

The first demo session began by giving an explanation of the different parts of the editor to the test participants. After the introduction it was decided that the participants would create a game in the exercise mode as it suited better the listening comprehension exercise that the participants had created. In the demo session one of the participants used the editor at a time and the others observed the game creation via projector screen. This set-up proved to be quite useful in regards of getting information about the usability of NFC-ACT as this way the participants had to discuss the game developing procedure as a group.

Writing the tags was also demoed in the first session. The participants did not however upload the game to the NFC-enabled phone as at the time there was still some problems with this step in the software.

As only the exercise mode was demoed in the first session, the participants agreed to come for a second demo session to which the participants would prepare material for a game in the practising mode. The participants came up with a grammar exercise targeted for teenage students where the students would fill out sentences with selecting

the right alternative. The main focus of the second session was of course the usability of the editor but with more emphasis on whether the participants still remembered how to use it and on the game creation process now that the participants were more familiar with the software. In addition, the creation of the physical content, i.e. material besides the tags and the game itself, was demoed.

### 5.3.2. Results from demo sessions

The overall feedback from the participants about NFC-ACT was positive. This came clear in the discussions with the participants as well as from the questionnaire results where four of the participants agreed that they were satisfied with NFC-ACT and three of them agreed that NFC-ACT was fun to use, see Figure 10.

However, as can also be seen from Figure 10 only two of the participants would recommend NFC-ACT to their friends and only two also agreed that NFC-ACT was pleasant to use. This might have something to do with the bugs that occurred in the software during the demo sessions as in the questionnaire answers two of the participants actually mentioned the bugs as being the main problems that they encountered when using the software. The bugs were also raised by the participants during the discussion in the first session as one of them stated "perhaps if there were to be a lot of errors and bugs in the software then that would be a problem for most teachers because none of, well at least many of the English teachers don't have knowledge as far as technology goes".

#### Ease of learning and ease of use

In the questionnaire all of the participants agreed that they quickly learned how to use NFC-ACT, see Figure 11. Moreover, all of them agreed that it was easy to learn how to use the editor and one of them stated in the questionnaire responses that "even with minimal guidance it was really easy to learn how to use the application". In the discussion the participants assumed that there was no need for an elaborate training for the software and they speculated that a day or even a half a day training where the teachers would be shown how to operate NFC-ACT would be adequate.

Even though the results support that using NFC-ACT is relatively easy to learn, the participants still required some instructions and help during the game creation. Especially realising the concept of a challenge as represented in the editor proved to be difficult for the participants even though the concept was explained to them before they started creating the game. They required help in order to understand the different components of a challenge and during the discussion they mentioned that "everything was not as self-explanatory in the editor as it could have been" and that "you need to know what each item means so it [NFC-ACT] would be easier to use". Also in the questionnaire responses one of the participants reported "unfamiliarity with the symbols" being problematic. Especially the path for an unrecognised tag was something that the participants had difficulty in realising and they did not add any properties for the path when they were creating the games. The participants also had some difficulties in realising what the delay [the delay after a tag is selected] was and what happens after that

# Q12 Please select, how strongly you agree with the following claims about the satisfaction of NFC-ACT. 1= Strongly agree and 5=Strongly disagree

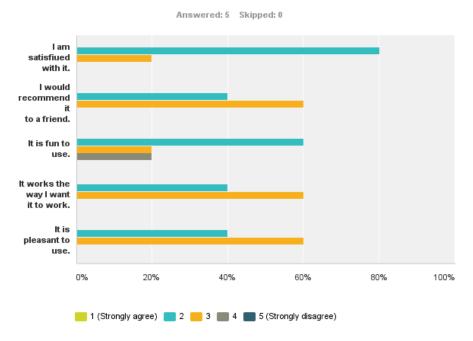


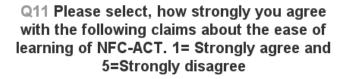
Figure 10. Demo session participants' satisfaction with NFC-ACT.

time in a challenge. These aspects of the UI are going through major modifications for the new version of NFC-ACT.

Nevertheless, the participants managed to create the game without major problems already during the first session and mentioned that creating the game was quite straightforward and simple. This was the general opinion among the participants also during the second session as one of them mentioned that "it's been many weeks since we used this and once you do one, it's pretty straightforward to remember how to use it".

The questionnaire results from the first session also supports the ease of use as can be seen from Figure 12. All of the participants agreed that NFC-ACT was simple to use, easy to use and that the UI was user friendly. Moreover, from Figure 13 it can be seen that the majority of the participants agreed in the questionnaire with the claim that "in general, using the application is intuitive".

However, there were two questions regarding the ease of use of NFC-ACT in the questionnaire that had dispersion in the results, see Figure 12. The other was the statement whether NFC-ACT required the fewest steps possible to accomplish what the user wants to do with it. The bugs in the software might explain some of this dispersion as some of the participants might have felt the bugs adding more steps for the game creation. The other question with dispersion in the responses was whether both occasional and regular users would like NFC-ACT. The dispersion in the responses indicates that NFC-ACT would be more enjoyable for users who frequently use it. The



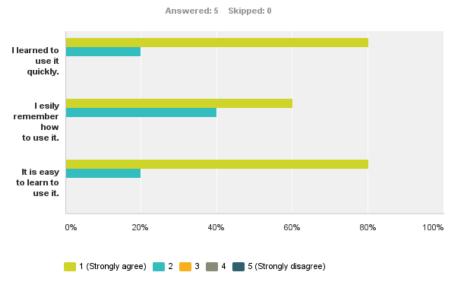
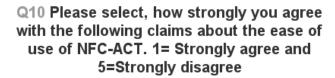


Figure 11. NFC-ACT learnability.

responses also indicates that in order to achieve more out of NFC-ACT the user first has to learn how to use it properly.

The participants also had difficulties in grasping the difference between the general flow and the flow of each individual challenge even though it was explained to them twice during the first demo session: once before they started and one time during the game creation they were recommended to add content to the general flow if they wanted that content to be copied to all of the challenges. Despite of these instructions they still added the same content (e.g. the same images and instruction texts) to each challenge individually when they could have just added content to the general flow. One of the participants also mentioned this in the questionnaire responses by stating that "having to add the same written texts for tags several time, even though all of them were identical on that stance" was problematic. The participants also experienced this problem during the second demo session but this time they were interrupted and the feature was explained to them once more and finally the participants were able to use the feature correctly. Based on these findings, the feature will be removed from the upcoming version of NFC-ACT.

One explanation for the difficulties the participants had in the different game flow templates was the fact that they still were not completely clear on what would be displayed in the phone and what data would be written in the tag. Also as the participants had created the material for the games before the demo sessions so some modification was needed in order to transport the different game elements to the NFC-ACT editor.



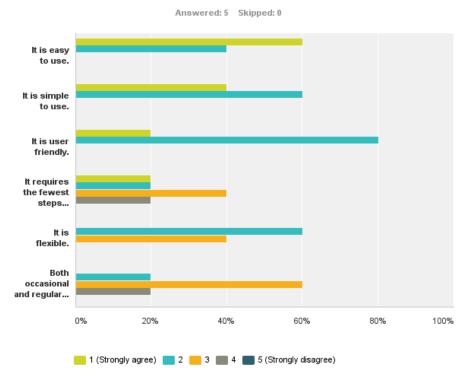


Figure 12. NFC-ACT ease of use.

As previously mentioned, in the first demo session the tag writing was also demoed. Writing the created challenges to the tags using the tag reader went smoothly and the drag-and-drop UI of the tag writing was quite intuitive for the participants to use. They also quickly understood that they needed to readjust the position of the tag in the reader if the tag was not first recognised.

#### Time consuming aspects of NFC-ACT

One of the main goals of the demo sessions was to find out how time consuming creating exercises with NFC-ACT would be. Based on the questionnaire results the participants felt that using NFC-ACT was not too time consuming as four of them disagreed with the claim that "creating a game was too much time consuming", as shown in Figure 13.

The time consuming aspect was discussed widely in the first demo session. The participants mentioned that using NFC-ACT was "light" and "fast". However, the participants were in agreement that if using the software would be too time consuming, teachers would not bother to use it. They mentioned that the ideal case for using the software would be that the teachers could just add content to the editor so fast that they

# Q8 Please select, how strongly you agree with the following claims. 1= Strongly agree and 5=Strongly disagree

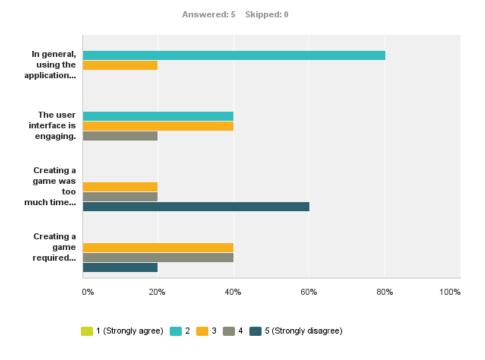


Figure 13. Time consuming aspects of NFC-ACT.

could do it during the breaks between classes and some of the participants appeared to have doubts if that would be realistic for NFC-ACT. However, one of them mentioned that "this was our first time using it [NFC-ACT] and we have never even heard of this before and it was still pretty fast".

Based on the discussion during the sessions, the majority of the participants felt that creating content for different games would actually be more time consuming than creating the game with NFC-ACT (assuming that the user is familiar with the software already). During the discussion in the second session they identified that both coming up with an idea for a game or an exercise as well as actually planning the content for the exercises could take some time.

However, the participants mentioned that the most likely scenario would be that the teachers would come up with ideas when they are planning the material for the lessons, so finding different exercise ideas would not be the major time consuming aspect. Moreover, in the first session they pointed out that many teachers are going to create extra content for classes anyway as one of them said that "the existing material that comes from the publishers just isn't that innovative or that fun for the kids". For such teachers NFC-ACT would not create extra-work, but instead it might even save time as one of the participants mentioned that "this [NFC-ACT] might even save time if you are one of those teachers who want to make his or her own material". They also pointed out that it would be even better if NFC-ACT would be provided to the teachers'

personal computers, so that the teachers could use it also at home as according to the participants teachers do much work outside working hours anyway.

The participants also felt that coming up with ideas for exercises get easier the more one uses NFC-ACT as one of them mentioned that they already came up with some ideas for exercises already while using it. They also mentioned that one way for facilitating the process of coming up with ideas would be if the editor contained ready-made game templates. By following the templates the teachers could then more easily come up with their own ideas for what kind of games and exercises they would want to create with the editor.

During the discussion in the second session, it came clear that the participants were content also with the feature that the previous games are easily accessible via the UI of NFC-ACT. This feature could be a time-saving feature in many aspects. The teacher could use the older games as a base for new games. Moreover, the previously created games could also serve as a help in learning how to use the editor as the teacher could use them as a reference when creating new games.

Based on the discussion, the participants also felt that the quality with the exercises and games created with NFC-ACT would improve with time once the user is more familiar with the software. Already during the first session one of them mentioned that "what I think would've been nice is that we would have had this session before developing our game so we would know exactly what to do and then we would have had better ideas". Also in the second session they mentioned that their ideas for the games were still quite simple and that they actually had had problems with coming up with useful exercises that they could create with NFC-ACT for language learning. However, they felt that the main reason behind this was that they were still a bit unfamiliar with the software, or as one of them said that "it's more about us not being able to use the program as well as we can".

In both of the sessions the participants mentioned that a cloud-based service or a database where teachers could upload and download content would be a time saving feature which they would like to include in NFC-ACT. Teachers could use the database for inspiration when coming up with ideas for exercises and if there already would be a certain game available that the teacher had in mind, it would definitely save time since the teacher could just use that game instead of creating a new one from scratch. As one of the students mentioned "collectively this would work very well".

According to the participants one problematic area in NFC-ACT in regards of time consuming aspects was the tag writing and uploading the game(s) to the phones. The participants mentioned this concern in both of the sessions as well as in the question-naire feedback. They wondered if in the future there would be available devices (they used the word "dock") with which the teacher could write many tags simultaneously. They also suggested that instead of uploading the game to a phone one-by-one it would save time if many phones could be updated simultaneously.

# Creating tangible content for the games

In the second demo session the participants also demoed how they would create tangible content for the games, i.e. the physical content with which they would represent the game and its tags in the classroom. They decided to write the different choices for each question on post-it-notes and then place them in a random order around the classroom as they felt it was the best option for the grammar game that they had created.

Creating the physical content proved to be a good exercise as the participants had difficulties in transporting the game to the post-it-notes, meaning even though the game flow for them was quite straight-forward when they did the game using the editor they were still not completely sure on how the game would be represented using the tags. For example they still were not sure on how the students using the game would proceed from one challenge to another so they were still under the impression that NFC game in the phone would require user interaction when it goes from one challenge to another.

Naturally as this was only the second time that the participants were using the editor (and the first time that they were creating tangible content), transporting the game into physical content does not go as smoothly as for more experienced users. This would indicate that any future trainings for NFC-ACT should also take into account the creation for tangible content. As tangible content for the game is created, the users have to think the game in its real-life context and thus the users might learn the different concepts of the editor more quickly. In addition, once the participants were clear on what information would be represented on each tag, the creation of the physical content went quite fast.

#### **Usefulness of NFC-ACT**

In both of the demo session one focus area was to find out the usefulness of NFC-ACT. The participants brought up many use cases where NFC-ACT would be useful to the teacher and the overall feeling after the sessions was that there was a need for a tool such as NFC-ACT.

In the questionnaire results all of the participants saw NFC-ACT being useful for teachers, see Figure 14. As the participants mentioned in the first demo session, games are played a lot in the classrooms so NFC-ACT in that sense would be useful. Also, in the second demo session one of them also mentioned that between the demo session he had had some lessons with students where he felt that NFC-ACT would have been useful.

The participants felt that NFC-ACT could also bring something new to the classes, as one of them mentioned that "I think everything is done the same way for the past few decades" and that "help and new ideas are always welcome" which indicates a need for NFC-ACT or equivalent technologies. Even though the usage of new technologies might depend on the teachers' knowledge of technologies and their eagerness to use them, one of the students mentioned that "teachers might not always even realise how much [new technologies for teaching purposes] they would have available".

As the participants were all English majors they did feel that it might be an issue to come up with ideas that would be relevant for English language learning. However, when thinking the games in a more broader context the participants came quickly up with different ideas on how to utilise the games for students of different ages which supports the usefulness of NFC-ACT. One of the participants also mentioned during the second demo session that if the users (i.e. the teachers) would put more effort to the planning process of the games then they could create with NFC-ACT useful learning material for the classes.

# Q13 From the viewpoint of a teacher, please select how useful do you see NFC-ACT. 1 = Very Useful, 5 = Not useful at all

Answered: 5 Skipped: 0

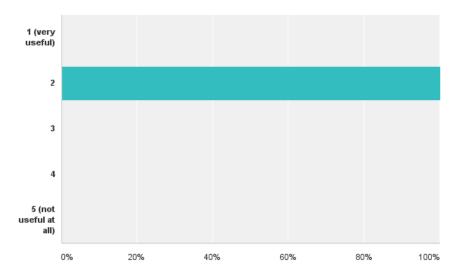


Figure 14. Usefulness of NFC-ACT.

Despite the consensus among the participants was that NFC-ACT would be useful they also saw some problematic points of using it, the most obvious being the cost of the phones as every student should have a NFC-enabled phone. In addition to the cost of the phones, the participants were seeing the cost of the tags also as a problem. However, once they were told that the tags were not so expensive, especially when they are bought in bulk, and that they can be reused, they did not see the cost of the tags as a problem any more.

One solution for the cost related problems could be that the students would be able to use their own phones for the exercises. However, the participants felt it problematic as even though in the future all the students might have NFC enabled phones themselves the teachers would not be able to monitor what the students are doing with the phones. According to the participants students using their phones for e.g. social media, gaming and other applications during classes is a common problem. The participants felt another solution for the cost related problems might be some other device that would be designed only for the use of NFC-based exercises, as they might be more affordable than phones.

The participants also saw problems in the scenario where the schools would be responsible for acquiring the phones and then giving them to the students for the duration of the classes. As one of the participants mentioned "it would add a lot to the teachers' responsibility for the hardware if we just give it to them [the students] and then just let them run loose so it might give some extra pressure or feeling of responsibility for the hardware". In addition, this scenario could create some problems to the teachers if they would have to carry all the phones to the classrooms.

The participants felt that one solution for all of the problems with the devices might be that if in the future the students would have some sort of device that could be used for many different learning purposes. As an example they mentioned an "all-around tablet". This way the students would already have a device that would also support NFC-based exercises.

During the demo sessions the participants also provided some future improvement ideas for the editor. One of them was the already mentioned ready-made game templates and the online database. According to the participants one other improvement would be if the games could provide interactive feedback for the students, for example if the games could provide information for the students that why the answer they chose was incorrect or correct.

Another improvement idea would be if the games could provide the teacher some sort of log on how the students executed the games. From the log the teachers could for example see what kind of mistakes the students have made and how many tries they used to get the correct answer. This sort of log could then pinpoint the areas where a student would require more help and thus make it easier for the teacher to provide more personal feedback to the students. If it would be possible to integrate the games with interactive feedback and logs the participants felt that NFC-ACT could also be useful for evaluation and exams.

# 6. USABILITY TEST DESIGN

# **6.1. Structure of testing sessions**

Based on the feedback from the demo sessions, many changes were introduced to NFC-ACT and the implementation of those changes is still under work for the upcoming version. Once the implementation is ready, the next step is to undergo thorough usability testing with actual teachers. This chapter will present a detailed plan for the usability testing for the next version of NFC-ACT.

The testing will be carried out with teachers, who already have teaching experience, preferably for more than five years. The main reason behind this decision is the fact that NFC-ACT has not yet been tested with experienced teachers and hence input from teachers is missing.

Based on the approximations presented in 2.6 as well as taking into account the relatively small size of the NFC-ACT project, the usability testing will be carried out with three test participants. There is some leeway for testing with four test users as well, if four suitable test candidates are found. However, four is the absolute maximum amount of needed test participants and three test participants will be more than sufficient.

The usability testing for NFC-ACT will be carried out in four phases. By dividing the testing for separate phases the test sessions are not too demanding for the test users and the focus for each phase can be shifted thus maximising the output gathered from each phase. The four different test phases are presented in Figure 15.

The first and last phases are more for sharing and gathering information and the two middle phases are then focused for actual usability testing. The first phase is an introductory session and the second phase is then the actual usability test session. The second test phase uses an adapted combination of the cooperative usability testing (CUT) and the SUXES method. The session consists of three stages, stages A-C. The third testing phase is a diary phase where the test users use NFC-ACT by themselves and in the last phase a debriefing session is held.

# 6.2. Testing goals

At this stage of development there is more need for qualitative than quantitative results. In another words, the main focus of the usability testing is to find out if the users can learn how to use NFC-ACT and whether the teachers find NFC-ACT usable and useful. Usable here refers to easy-to-use and clear UI and a smooth user experience and useful refers to whether the teachers find meaningful use cases for NFC-ACT in their every day teaching activities. Useful also refers to whether the teachers find NFC-ACT supporting and complementing their work.

All in all five usability testing goals are defined for NFC-ACT usability testing. Table 5 lists the goals. The first four usability goals are aimed at finding out how usable the test users find NFC-ACT and then usability goal no. 5 will cover the usefulness aspect of NFC-ACT.

Table 6 defines the focus and the NFC-ACT screen(s) under testing for each goal. The first two usability goals mainly focus on the game development, i.e. the main

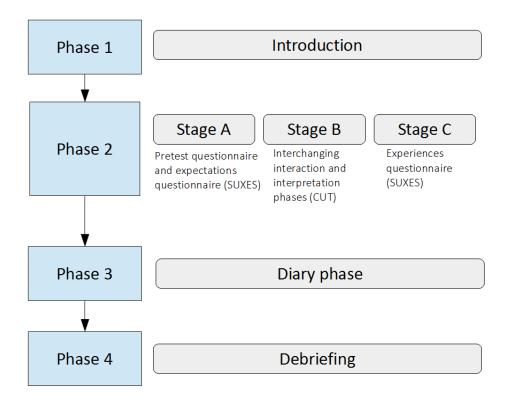


Figure 15. Testing phases for NFC-ACT usability testing.

screen of NFC-ACT. The game creation process is the most complex task in NFC-ACT as it requires a considerable level of abstraction: how to transform a game idea into a working game using NFC-ACT and how to represent different phases of the game using challenges? Thus the UI for this screen is the most troublesome from usability point of view. Moreover, the main screen will go through major modifications for the upcoming version of NFC-ACT and for this reason requires the most attention.

Usability goals 3 and 4 then are aimed at finding out usability issues when writing the game to tags and transferring the game to the phones. Lastly, usability goal 5 is related to the usefulness of NFC-ACT and hence it does not have focus on any of the screens.

In addition, during the second phase (i.e. the actual usability testing session) task success rate and severity rates will be measured during the testing in order to provide more detailed information for the first usability goal (is the UI clear and easy to follow).

Performance measures are not in the scope of this usability testing. Of course one of the main goals that the development team has is that the users find using NFC-ACT not too much time consuming and that the software reacts fast for user input, but the actual measurement on how fast the users complete the different tasks or the software reaction time will not be measured. Having said that, during the testing the time to complete tasks will naturally be measured but within the context on finding out how fast will the users grasp the different icons, structures, and other elements of the UI.

As from the results of the demo sessions it was clear that getting to know the UI of NFC-ACT requires some time, especially as the test users in the usability testing also

Table 5. Usability goals

Goal	Description
1	Is the UI clear and easy to follow?
2	Will users quickly learn how to use NFC-ACT?
3	Will users be able to write the tags?
4	Will users be able to upload the game to phone?
5	Do the test users find NFC-ACT useful?

Table 6. Focus for usability goals.

Goal	Focus	Screen
1	Can users create a game template?	
	Can users create a challenge?	
	Is the flow of a challenge clear?	Main screen
	Are the icons clear?	
	Is the terminology clear?	
	Is the UI clearly structured?	
2	Explore vs. practice game patterns	All screens
	Navigation in the UI	
3	Ease of use in drag & drop UI	Tag writing screen
4	Ease of use in drag & drop UI	Deploy screen
5	Identifying use cases for NFC-ACT	
	Effectiveness and efficiency of NFC-ACT	_
	Time consuming aspects	

will not be previously familiar with NFC-ACT. Moreover, the demo sessions revealed that already on the second try the users were able to use NFC-ACT more swiftly, it is assumed the once the user is familiar with the different UI elements of NFC-ACT, using the edition tool is not too much time consuming. Hence performance measures regarding task completion are not of interest in the scope of this usability test session.

Moreover, as section 6.3.2 will show, the nature of the actual usability testing session is such that there is little time for logging the number of errors made, wrong icon choices, or keeping a count on occasions when test users display signs of frustration. In order to gain such data one test monitor would solely have log these measures during testing or they could be measured post-session from a video recording or an on-screen video capture of the test user using NFC-ACT. Either way, the measurements would be too much time consuming in regards of information gained and hence performance measures are not in the scope of this usability test plan. One way to measure such data would be to use some sort of automatic logging software that would measure and analyse mouse movements, selections and time to complete tasks in the UI automatically but currently the budget for usability testing of NFC-ACT does not have room for acquiring and setting up such software.

Lastly, it can be assumed that usually performance and preference results correlate. In another words, if the users find the software easy-to-learn and easy-to-use, the soft-

ware performance (time to complete tasks, number of occurred errors and wrong menu choices etc.) is also at an acceptable level.

#### **6.3.** Test sessions

The following test plan is designed so that all testing activities can be conducted by only one test monitor. Naturally, the preferable option would be that the testing is carried out by two test monitors but nevertheless, the following usability testing is designed so that also one test monitor is able to conduct it.

#### 6.3.1. Introduction session

The aim for the first session is the give an introduction for the test users about the main concepts behind NFC-ACT. The UI nor the game creation process is not revealed but the users are provided with a verbal explanation about the usage of NFC-ACT. In addition, an introduction about NFC is given to the test users. During this session the test users can also ask questions about NFC or NFC-ACT.

After the introduction, the participants play a demo NFC game on a mobile phone. The game is an anatomy game with the purpose of teaching anatomy in English. In the game the players try to connect the correct picture of a body part with the correct English word. For example the game in the mobile phone shows the player an picture of an arm and the player must find the corresponding tag with the word "arm" on it and touch the tag with the phone.

After the session the teachers can take the NFC enabled mobile phone and the tags home with them so that they can get to know the game even more and even maybe play it with their students in the classroom, should the game suit the teachers' lessons. However, this is not a mandatory part, e.g. if there is not enough mobile phones available, but should there be enough phones and tags to be distributed there is no reason why the teachers could not take the game with them.

Not only does the game introduce the NFC as an technology for the test users, it could also already inspire the test users to think about how they could utilise NFC in their teaching. The test subjects are of course not yet instructed to think about how to include NFC games to their teaching, but the aspiration here is of course to spark interest in the test users towards NFC enabled learning. Moreover, as the anatomy game is an integral part of the upcoming usability testing session as will be shown in the following section, it is good that the test users are already familiar with the flow of that particular game.

The introduction session should last somewhere between 30 minutes to an hour at the maximum per each test participant. Besides the game related hardware no other specific material is needed for this session. If time allows the test monitor may prepare some presentation materials (e.g. PowerPoint slides) about NFC and/or NFC-ACT but this is not mandatory.

# 6.3.2. Usability testing session

The second phase consists of the actual usability testing session. The selected methodology for usability testing is an adapted combination of cooperative usability testing and the SUXES method, described in chapters 2.5.2 and 2.5.3 respectively. The reason behind this adaptation is that the results provided from both of the methodologies serve best the usability goals defined for NFC-ACT usability testing. In addition, as chapter 2.3 already described, combining different usability evaluation methods usually yields in better results.

The main focus of the second session is the game creation process. As the game creation requires the participant to form abstractions of the game to be created, the UI for this part is the most challenging in NFC-ACT. This part of the UI is also the part where the user spends the most of the time while using NFC-ACT so it is very likely that if the user finds the game creation UI of NFC-ACT unclear and confusing, the user will not bother to use NFC-ACT.

It was also clear from the demo sessions results that the game creation UI was not as clear as it should have been, and many modifications are going to be implemented to this part of the UI for the next version of NFC-ACT. Thus, the aim of this usability testing session is to test whether the modifications will improve the usability.

Uploading the game to the phone or writing the tags will not be covered in the usability testing session. Instead they will be covered during the third phase. As the demo session already pointed out, the tag writing was quite straight-forward and did not reveal any major usability issues, so there was no clear need to re-test in this usability testing session. The decision was made also in order to optimise the time used for the usability testing session: by leaving out the testing for tag writing and the game uploading the session can focus more on the actual game creation process, and the session duration is not too exhausting for the test participant.

The second session consists of three stages: stages A, B and C. Stage A is the first part of the usability testing session where the preparations in order to proceed to next stage are done. During stage B the test user uses NFC-ACT and stage C summarises the session.

#### Stage A

Stage A begins by letting the participants fill out a pretest questionnaire. The questionnaire is solely meant for gathering demographic data and for drawing a technical profile about the test participants. The questionnaire covers the following areas:

- age,
- gender,
- what subject do the test participants teach,
- teaching experience (in years),
- how much do the test participants create educational content by themselves,
- previous experience of using editing tools, and

• the test participants familiarity with NFC.

The pretest questionnaire (as well other questionnaires during the second session) are printed to paper instead of using an online survey. This allows the test monitor to conduct the test anywhere that suits the test participants' schedule the best without having to worry about internet access in the test location. Moreover, as there is going to be only three test participants, transforming the data to a digital format for better analysis after the session should not take too long from the test monitor.

After the pretest questionnaire, the test participant proceeds to filling out the expectations questionnaire as per the SUXES method. In the questionnaire the participant is provided with nine statements about using NFC-ACT and they are asked to mark two values for each statement: the acceptable level (i.e. the lowest acceptable quality) and the desired level (i.e. the uppermost level of desired quality). The values are selected from a 7-point Likert scale where 1 describes low importance and 7 describes high importance. Figure 16 depicts the nine quality statements in the expectations questionnaire.

		1	2	3	4	5	6	7
1	Using NFC-ACT is intuitive/straightforward.							
2	It is easy to learn how to use NFC-ACT.							
3	The user interface is clear and easy-to-follow.							
4	Information is presented in an organised way.							
5	Symbols and icons in the user interface are clear and self-							
	explanatory.							
6	Using NFC-ACT is quick.							
7	Using NFC-ACT is pleasant.							
8	NFC-ACT provides enough help while it is being used.							
9	NFC-ACT benefits my work.							

Figure 16. The usability statements in the NFC-ACT expectations questionnaire.

# Stage B

After the test participant has filled out the questionnaires, the test moves on to stage B. This part of the usability testing is carried out using the cooperative usability testing methodology, and more specifically cooperative usability testing with interchanging interaction and interpretation phases.

During stage B the test participant is instructed to use NFC-ACT in order to recreate the anatomy game that they already are familiar with from the first session. The main reason behind the decision not to let the test users to create a game of their own choosing is based on the results from the previously held demo sessions. In the demo sessions the test participants had trouble coming up with ideas for a suitable game that they could create using NFC-ACT as they had not used NFC-ACT before. Moreover, after an idea was ready, the participants had difficulties in realising what would be displayed in the phone and which information should be written on the NFC tags etc. In another words, the test participants did not have a clear mental model on how to transform their game idea to a NFC game. By eliminating the effort for making an

abstraction of a NFC game, it is ensured that the usability test focuses only on the UI of NFC-ACT.

In order to guide the test participant through the game creation process the test participant is provided with a document describing different scenarios. Stage B will go through three different scenarios in the game creation process, namely, a scenario for creating a new game template and two scenarios for adding a challenge to the game. The main reason with using the scenarios is to structure the testing session into clear components, thus clearly dividing the interaction phases from the interpretation phases. In fact, the test participant is instructed both verbally as well as in a written form in the scenarios document not to continue to the following scenario until permission is granted by the test monitor.

In the first scenario the test participant is asked to create a new game template for the anatomy game and they are asked to stop when they feel they have reached the point when they are ready to start adding different tasks to the game. The word 'task' is used instead of challenge in order not to give away the terminology of the UI too easily in order to see whether the test participant understands the different terms used in the UI. If the test participant still appears to proceed too far in the game creation process, the test monitor may stop the test user. This is the first interaction phase.

After the first interaction phase, the test monitor initiates the first interpretation phase. In the interpretation phase the test monitor and the test user go through the actions that the user made in the interaction phase with the emphasis on the main problem areas. The test monitor initiates the conversation and motivates the participant to discuss the possible causes of the situation and encourages the test participant to provide improvement ideas, if they have some.

In addition, during each interpretation phase the test monitor also asks the test participant the following set of questions (if an answer is not already provided during the conversation):

- 1. Where did the user feel he/she experienced the most difficulties?
- 2. What is the user's experience of the layout and interaction mechanisms (buttons, icons, terminology) i.e. what was clear and what was not?
- 3. What were the positive parts?
- 4. Does the user have any improvement suggestions?

In the second and the third scenario the participant is asked to create two challenges to the game (e.g. a challenge for finding the word 'arm' and a challenge for finding the word 'head'). The interaction and interpretation phases change turns accordingly. The reasons for having two scenarios for creating a challenge is to find out a) how easily can the participant create a challenge to the game template; and b) is there difference in creating a challenge after the participant is already familiar with the icons and the terminology of the UI.

There is, however, one difference with the last interpretation phase as with the first two. During the last interpretation phase the question set is elaborated with the following questions:

1. Does the user see NFC-ACT suiting their daily work?

- 2. Would NFC-ACT improve the users daily work?
- 3. Does the user see NFC-ACT as too time consuming?
- 4. How much training would be needed for NFC-ACT?
- 5. Other comments?

Test monitor observes while the test user carries out each scenario and takes notes especially on the problems. In order to facilitate the note taking the test monitor is provided with a template for taking notes for each scenario.

In the note template the test monitor is instructed to list the main problems that the participant experience during each scenario. The test monitor is also asked to rate each problem based on severity ratings described in 2.5.4 i.e. is the problem a level 1, 2, 3 or 4 problem.

In addition for each scenario the test monitor is asked to provide the success rate as described in chapter 2.5.4. In another words, the test monitor marks in the note template if each scenario was a success, partial success, or a failure. There is also room for providing a short rationale why the scenario was a success or a failure, should the test monitor have time for that.

By marking each problem with a severity rating serves for two purposes. First, based on the severity rating, the test monitor can quickly check which problems were the most troublesome during each scenario and guide the emphasis of the discussion during the interpretation phase towards the major problems. The severity ratings can then be plotted against an orthogonal scale after the testing session is ended for a more deeper analysis of the problems encountered. The success rates can also be used post-test for calculating effectiveness and efficiency.

# Stage C

After the test participant is finished with the scenarios and the discussion for the last interpretation phase is ready, the test participant is asked to fill out the experiences questionnaire needed for SUXES. It includes the same statements than the expectations questionnaire but this time the test participants are asked to select only one value, i.e. the value describing the experience they had while using NFC-ACT.

After the usability testing session has ended, the test monitor can compare the values from the two SUXES questionnaires in order to determine the Measure of Service Superiority (MSS) and the Measure of Service Adequacy (MSA). MSS describes the difference between the perceived level and the desired level, and MSA describes the difference between the perceived level and the accepted level. When the experiences fall into the zone of tolerance the MSS values are negative and MSA values are positive.

After having finished the SUXES experiences questionnaire, the test participants are demoed on how to write the tag as well as how to upload the game to the phone so that the they are able to do this during the diary phase. The test participants are also provided with NFC-ACT "quick instructions", i.e. a guide going through the basic steps of creating a game template, how to write the NFC tags, and how to upload the game to the phone so that they will not stop using NFC-ACT during the diary phase if they for example cannot remember how to write the tags. The quick instructions

should also include a "quick help" guide for basic problems the test participants might face, e.g. if they are unable to upload the game to phones they are instructed to check that the connection settings in the phone are correct. In addition, the test participants are encouraged to contact the development team should they face any problems that they are unable to solve by themselves and the problems prevents them from using NFC-ACT.

Stage C - as well as the whole usability testing session - ends by providing the teacher with the needed material for conducting the diary phase.

#### **Practical issues**

The whole usability testing session will be recorded with one or more video recorder and/or audio recorder for post-test analysis as well as for recording the discussion during the session. The video may also be used for measuring the time to complete the different scenarios.

In addition, stage B can also be recorded by using an on-screen video capture for recording the selections the test user made in the application during different scenarios but only if such software is available. The selections can also be recorded by having a dedicated video recorder recording the screen as well but even this is not mandatory as if there is only one video recorder available it should be used for recording the discussion and the expressions the test user displays while using NFC-ACT.

The test monitor should reserve 2-3 hours for each participant for the whole session. Materials needed for the second session include:

- laptop running NFC-ACT,
- video and/or sound recorder(s),
- questionnaires (in paper format),
- all hardware needed for the game (phone, tags, NFC reader),
- test monitor note templates,
- equipment to be provided for the test user for the diary phase, and
- NFC-ACT quick instructions to be provided for teachers.

# 6.3.3. Diary phase

Another aim of the usability testing for NFC-ACT is to gain information on how useful do the teachers see NFC-ACT in their work. Some conclusions can be drawn already from the results of the second session, but in order to really answer the question, NFC-ACT will be provided for the test participants for test use for two weeks. During this time the participants will create games using NFC-ACT, write content to tags and upload the game to provided phones.

The diary phase is carried by letting the test participants to fill out an unstructured diary. The diary will loosely guide the users to log their experiences, whether they be

positive, negative or neutral, while they are using NFC-ACT but the diary will not ask any specific questions or satisfaction ratings. The users are asked to describe the use cases where they used NFC-ACT and provide feedback on how the tag writing and uploading the game to the phones went. They are also asked to describe briefly what kind of tangible material they created or used for the games. The test participants may choose if they want to fill out the diary in paper format or online. Both options will be presented to the test participants and they can choose the one that suits their schedule and preference better.

For the diary phase NFC-ACT will be provided with a couple of ready-made game templates in order to help the teachers to get started. Another reason behind this decision is to find out whether the teachers find such a feature helpful or not.

Should there not be enough mobile phones to be distributed for the test users, the diary phases for different users should then be arranged adjacent to each other. However, this would prolong the testing quite a lot so the aspiration here is to have enough hardware available so that each test user can conduct the diary phase testing simultaneously.

# 6.3.4. Debriefing session

After the diary phase, there will be one last session. During this session the test participants return the equipment from the diary phase as well as their diaries if they have not used the online diary form.

This session is aimed to be an open discussion, meaning that except for some basic questions (what was the most positive or negative aspects etc.) the participants are encouraged to describe their experiences with NFC-ACT in their own words. If the test participant has filled out an online survey, the test monitor may naturally go through the feedback in order to see some points that could be discussed during the session. However, since the test participants are also provided with the option not to use the online diary version, the test monitor may not have access to the diary feedback until the last session. Thus the session will not be based on a predefined question list as it can be assumed that the most troublesome - as well as the positive - points will be pointed out by the test participant anyway.

During the debriefing session the test participants are also asked would they recommend NFC-ACT for a colleague. With only three test participants there is no need to calculate NPS as described in 2.5.4. Instead, the test participants are only expected to provide a yes/no answer and elaborate the reasons behind their answer.

The test monitor should reserve one hour per each participant for the debriefing sessions. Also these sessions will be recorded with a video and/or audio recorder for further analysis.

# 7. DISCUSSION

The objective of this research was to develop a comprehensive usability test plan for carrying out usability testing for NFC-ACT. This chapter evaluates the results of this research and goes through the quality of the research and the validity of the test plan. This chapter also presents also some future improvement ideas for the test plan.

#### 7.1. Evaluation of the results

The presented usability testing plan provides value to the NFC-ACT development team once it is taken into use as with thorough usability testing the most critical usability issues with the upcoming version of NFC-ACT can be revealed. The original intention for this research was indeed to also conduct the testing itself but as the implementation of NFC-ACT was delayed it was agreed that the testing is postponed until the implementation is ready.

Another impediment for actually conducting the testing was that it proved quite difficult to find suitable test participants. The testing should be carried out with teachers who already have working experience in order to gain the best possible results from the usability testing. However, finding teachers who would have the possibility to participate to the testing was difficult as teachers are quite busy and have really strict curriculum and schedule that do not give leeway for extra work.

The results from the usability testing will also provide crucial information on how useful do the teachers see NFC-enabled seamless learning. The results will also reveal if there is demand for such editing tools as NFC-ACT.

The test results could also shed some light on the need and usefulness of other seamless learning tools as well, for example if the feedback from the test users during the test sessions provides new ideas for the development team that are then implemented and studied. Thus, the proposed usability testing plan could provide the seamless learning research community new information.

The presented usability testing plan is also easy to adopt for usability testing of other tools than just NFC-ACT. As this test plan is available for public use, it can provide benefit to other development teams planning usability testing. The teams can either use the plan as it is presented here or modify it according to their needs.

Lastly, the proposed usability testing plans presents a new adaptation of usability testing as it combines cooperative usability testing with the SUXES method. Once the testing plan is put into use the efficiency of this type of usability testing method can also be studied.

# 7.2. Quality of the research

Yin lists in [65 p. 40-41] four criteria that have commonly been used to assess the quality of research. These criteria include construct validity, internal validity, external validity and reliability.

Construct validity is used for identifying whether correct operational measures have been used. Construct validity can be assessed by using three steps. The first step is used to indicate that multiple sources of evidence have been used for data collection and the second step establishes a chain of evidence related to the collected data. The third step is related to the data composition phase by assessing if the research has been reviewed by key informants. [65 p.41]

According to Yin, the six most commonly used sources of evidence include documentation, archival records, interviews, direct observations, participant observations and physical artefacts and as these sources are complementary, the quality of the research increases when more than one source is used [65 p.101].

This research has used documentation, interviews, direct observations and participant observations as sources for evidence. As Yin notes, correct use of documentation plays a vital role in increasing the construct validity of a research as documentation is used to corroborate and augment evidence from other sources [65 p.103].

The used documentation for this research includes a thorough literary review of usability research as well as technical literature about NFC. Moreover, as NFC-ACT is aimed to be used for seamless learning purposes, a literature review of seamless learning is also presented.

In addition to documentation, the usability test plan produced by this research is also based on the feedback from the demo sessions and interviews held in 2013. Thus this research has also used interviews, direct observations and participant observations as sources of evidence. According to Yin, interviews and observations are an essential source of evidence as they produce important insights into human and behavioural events related to the research as well as help in providing information for understanding the context or phenomenon under study [65 p.108, 110].

Second step in determining the construct validity for a research is to establish a chain of evidence by presenting the collected data with supporting and challenging viewpoints. By establishing a chain of evidence the research also presents enough evidence that the writer has enough knowledge about the subject. Thus, the chain of evidence allows the reader of the research to follow the derivations from documentation to research conclusions. This guarantees that the readers are able to draw independent conclusions about the validity of interpretations presented in the research. The chain of evidence also should be established so that it takes into account the selectiveness aspect of used data. In another words, the used data should be selected so that the report is limited to the most critical evidence and that presentation is not cluttered with secondary data. [65 p.122, 189]

This research ties the produced usability test plan tightly with the presented data from documentation and the demo sessions. Moreover, as the test plan follows the techniques used in cooperative usability testing and the SUXES method, an adequate chain of evidence for this research has been presented.

The third step in measuring construct validity is to assess if the key informants of the study have reviewed the report. The review helps in corroborating the essential facts and evidence presented in the research. Moreover, the corrections made through the review process can enhance the accuracy of the research by reducing the probability for false reporting. [65 p. 182-183]

This research has been reviewed by the project members involved in the NFC-ACT development and they have provided feedback for modifications. The development team has also expressed their approval for the proposed usability testing plan. However, the demo session participants have not reviewed the results. This decreases the

construct validity for this research a bit as the demo session participants have not been able to provide comments about the results. Nevertheless, when taking into account all three aspects of construct validity, it can be argued that the construct validity for this research is quite good.

Internal validity is related with the causal relationships of a research and can be tested by using analytical tactics such as pattern matching, explanation building or by using logic models. However, internal validity is mainly assessed for explanatory studies where causal relationships play a crucial role when determining the results and as Yin mentions, the above described tactics are not suitable for descriptive or exploratory studies [65 p. 43]. Hence, the internal validity of this study cannot be assessed.

External validity is used for assessing how well the results and findings from the research can be generalised [65 p. 43]. The proposed usability test plan in this research is composed so that it is adaptable for usability testing of other systems also and not just for NFC-ACT usability testing purposes. However, as the usability test plan has not been tested in practice yet (i.e. the actual testing is not yet done) its adaptability cannot be fully stated. When taking into account this limitation it can nevertheless be argued that the external quality of this research is good.

Reliability then covers the aspect on how well the research can be repeated. In essence good reliability requires good documentation so that other investigators can review the evidence directly also from its original source and not be limited to the data presented in the research. [65 p. 45, 119]

The documentation used as a basis for data collection is clearly marked in this research. Also the condensed results from the first demo session questionnaire are presented in 5.3 in order to increase reliability. In addition, the usability test plan and its different phases are reported in detail so that the steps can easily be followed when conducting the usability testing. Appendix 1 also presents some of the used questionnaires and material needed for conducting the testing.

Moreover, the reliability for the usability test plan is increased as the selected usability techniques (cooperative usability testing and the SUXES method) are of such nature that they support repeatability better than e.g. observation or normal think-aloud usability testing methods as they require the use of specific questionnaires and procedures. This way the results provided by the usability testing can better be compared against each other, for example if the test plan is used for testing different increments of an application, thus increasing the reliability of the test plan even more.

## 7.3. Test plan validity

A detailed plan for usability testing serves as a blueprint for the whole testing session by addressing how, when, why, who and what will be tested [13 p. 81]. In addition, the testing activities should cover all the defined usability goals. This section will go through the validity of the proposed test plan as a basis for a usability testing session.

According to [13 p. 83] a detailed usability test plan should include the following: purpose, problem statement/test objectives, user profile, method/test design, task list/scenarios, test environment, test monitor role, evaluation measures (data to be collected), and report contents.

The proposed plan described in this research begins by defining the test goals and the profile for needed test participants. It then continues by describing the used measures and the reasons on why they are chosen. The test plan provides also the number of needed test participants and theoretical background for the selected amount.

The test plan then gives a comprehensive description of all four usability testing sessions including measures to be taken, used testing methods, needed materials and test monitor role. Thus it can be argued the research paper provides enough evidence for the validity of the test plan as a basis for usability testing.

Evidence for usability test plan validity can also be derived from the usability goals. A valid usability test plan should cover all the usability goals defined for testing. The remain of this section will go through all usability goals defined in 6.2 in order to see if the test plan covers all of them.

The first usability goal focused on finding out how clear and easy to use the UI in NFC-ACT is. Its purpose was to find out whether the UI is clear and easy to use, whether the icons and terminology are clear and how well the UI is structured. All of these aspects are covered during test phases 2-4.

The second usability goal focused on learnability aspects of NFC-ACT and this aspect is taken into account in the second test phase, i.e. during the actual usability testing session in the SUXES-questionnaires as well as during the interpretation phases of cooperative usability testing. Also the diary phase should provide results regarding the learnability aspects of NFC-ACT.

The third and fourth usability goals then covered the usability issues when writing the tags and uploading the game to the phones. These will be gone through during the third phase of the testing, i.e. the diary phase.

The last usability goal was about usefulness of NFC-ACT. Usefulness is covered during test phases 2-4. During the second phase the SUXES questionnaire already takes into account also the usefulness aspect as well as the elaborated question set that will be gone through after the last interpretation phase. In addition, the discussion held in the fourth testing phase (based on the feedback from the third phase) covers usefulness aspects.

To summarise, the test plan also covers thoroughly all five usability goals defined for testing. Hence, the usability test plan is valid also from the usability goals viewpoint.

## 7.4. Future improvement ideas

Chapter 5.3.2 already described some improvement ideas for the actual NFC-ACT application but since the aim of this research was to develop a usability test plan this section will next go through some improvement ideas for the proposed test plan.

First, the plan needs to be tested by actually conducting the test sessions. By conducting the usability test according to the proposed test plan could reveal any potential weaknesses and needed modifications in the test plan.

One modification to the current test plan could be to replace the diary phase by another usability test phase. With this option, a second round of test users would be required to participate to the testing. First, the usability testing phase (including cooperative and SUXES usability testing) is conducted with the first group of test users and the software is then modified based on the findings. Then a second group of test

users conduct the same usability testing phase. This way the results from both usability testing sessions could be compared.

Another option would be to elaborate the diary phase to a more comprehensive beta testing phase where the software would be available for a larger number of test users who would then report their findings. In this option the usability testing session could serve as a preliminary basis for conducting the beta testing in two ways. First, the findings from the usability testing phase could be used as a benchmark for required usability level. The other option would be to modify the software based on the findings from the usability testing phase before conducting the actual beta testing.

## 8. CONCLUSION

The target of this study was to develop a comprehensive usability test plan for NFC-ACT seamless learning tool. The focus was to develop a plan that is aligned with the theory presented in the background section of this thesis. The original intention was also to conduct the usability testing according to the test plan but due to delays in the software implementation of NFC-ACT the testing was postponed.

This work began with a through literature review about usability and seamless learning research. Different concepts of usability testing and usability metrics were introduced and they then served as a background for the actual usability testing plan. Also the characteristics of NFC-ACT as a seamless learning tool were presented.

Another source when implementing the usability testing plan was the demo sessions held in spring 2013. The demo sessions already revealed some usability issues and pinpointed focus areas for the presented usability testing plan. Results from the demo sessions also gave positive feedback about the usefulness and usability of NFC-ACT and gave motivation for continuing the implementation of the tool.

Based on the feedback from the demo sessions the UI of NFC-ACT underwent major changes and due to the delayed implementation of these modifications it was not possible to conduct the actual usability testing. Also due to scheduling issues not enough test candidates were not found in time to participate in the testing. Thus the natural next step is to put the plan into use and carry out the testing.

Once in use, the presented test plan will surely provide abundant information regarding usability and usefulness of NFC-ACT. Moreover, the test plan is presented thoroughly and justification for the logic behind the plan is given. Thus the plan is easy to adapt for usability testing of other seamless learning tools as well as other systems.

## 9. REFERENCES

- [1] Sollervall H., Otero N., Milrad M., Johansson D. & Vogel B. (2012) Outdoor Activities for the Learning of Mathematics: Designing with Mobile Technologies for Transitions across Learning Contexts. In: Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on, pp. 33–40.
- [2] Wong L.H. & Looi C.K. (2011) What seams do we remove in mobile-assisted seamless learning? A critical review of the literature. Computers & Education 57, pp. 2364 2381.
- [3] Riekki J., Cortés M., Hytönen M., Sánchez I. & Kortesmäki R.L. (2013) Touching Nametags with NFC Phones: a Playful Approach to Learning to Read. In: Transactions on Edutainment X, pp. 228–242.
- [4] Diah N., Ismail M., Ahmad S. & Dahari M. (2010) Usability Testing for Educational Computer Game Using Observation Method. In: Information Retrieval Knowledge Management, (CAMP), 2010 International Conference on, pp. 157–161.
- [5] Sharp H., Rogers Y. & Preece J. (2007) Interaction design: beyond human-computer interaction. John Wiley & Sons, Ltd., Hoboken, NJ, USA, 2nd ed.
- [6] Nielsen J. (1993) Usability engineering. Academic Press, Boston, 1st ed.
- [7] Shackel B. (2009) Usability Context, framework, definition, design and evaluation. Interacting with Computers 21, pp. 339 346.
- [8] Nielsen J. (1994) Usability Inspection Methods / edited by Jakob Nielsen and Robert L. Mack. John Wiley & Sons, Inc., New York, NY, USA.
- [9] Dumas J.S. & Redish J.C. (1999) A Practical Guide to Usability Testing. Intellect Books, Exeter, UK, UK, 1st ed.
- [10] Bevan N. (2001) International Standards for HCI and Usability. International Journal of Human Computer Studies 55, pp. 533–552.
- [11] ISO 9241-11 (1998), Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on usability.
- [12] Ruthford M. (2002) Mix and Match Usability Methods: Picking the Pieces for our Project. In: Professional Communication Conference, 2002. IPCC 2002. Proceedings. IEEE International, pp. 343–351.
- [13] Rubin J. (1994) Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests. John Wiley & Sons, Inc., New York, NY, USA, 1st ed.
- [14] ISO 13407 (1999), Human-centred design processes for interactive systems.
- [15] Nielsen J. (2010), Interviewing Users (accessed 17 Nov 2014). URL: http://www.nngroup.com/articles/interviewing-users/.

- [16] Følstad A. & Hornbæk K. (2010) Work-domain Knowledge in Usability Evaluation: Experiences with Cooperative Usability Testing. Journal of Systems and Software 83, pp. 2019–2030.
- [17] Hollingsed T. & Novick D.G. (2007) Usability Inspection Methods After 15 Years of Research and Practice. In: Proceedings of the 25th Annual ACM International Conference on Design of Communication, SIGDOC '07, ACM, New York, NY, USA, pp. 249–255.
- [18] van den Haak M.J., de Jong M.D.T. & Schellens P.J. (2003) Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. Behaviour & Information Technology 22, pp. 339–351.
- [19] Frøkjær E. & Hornbæk K. (2005) Cooperative Usability Testing: Complementing Usability Tests with User-supported Interpretation Sessions. In: CHI '05 Extended Abstracts on Human Factors in Computing Systems, CHI EA '05, pp. 1383–1386.
- [20] Bevan N. & Azuma M. (1997) Quality in Use: Incorporating Human Factors into the Software Engineering Lifecycle. In: Software Engineering Standards Symposium and Forum, 1997. Emerging International Standards. ISESS 97., Third IEEE International, pp. 169–179.
- [21] Turunen M., Hakulinen J., Melto A., Heimonen T., Laivo T. & Hella J. (2009) SUXES user experience evaluation method for spoken and multimodal interaction. In: INTERSPEECH, ISCA, pp. 2567–2570.
- [22] Reichheld F. (2003) The One Number You Need to Grow. Harvard Business Review 81, pp. 46 54.
- [23] Kristensen K. & Eskildsen J. (2011) The Validity of the Net Promoter Score as a Business Performance Measure. In: Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE), 2011 International Conference on, pp. 970–974.
- [24] Ismail M., Diah N., Ahmad S., Kamal N. & Dahari M. (2011) Measuring Usability of Educational Computer Games Based on the User Success Rate. In: Humanities, Science Engineering Research (SHUSER), 2011 International Symposium on, pp. 56–60.
- [25] Kuniavsky M. (2003) Observing the User Experience: A Practitioner's Guide to User Research. Morgan Kaufmann Publishers, San Francisco, CA, USA.
- [26] Nielsen J. & Landauer T.K. (1993) A Mathematical Model of the Finding of Usability Problems. In: Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems, CHI '93, ACM, New York, NY, USA, pp. 206–213.
- [27] Coskun V., Ok K. & Ozdenizci B. (2012) Near Field Communication: From Theory to Practice. Wiley Publishing, 1st ed.

- [28] Riekki J., Sanchez I. & Pyykkonen M. (2012) NFC-Based User Interfaces. In: Near Field Communication (NFC), 2012 4th International Workshop on, pp. 3–9.
- [29] Lehpamer H. (2008) RFID Design Principles. Artech House, Inc., Norwood, MA, USA.
- [30] (2010), Enlighten NFC Tags for Smart Posters (accessed 17 Nov 2014). URL: http://www.smartposter.co.uk/enlighten/enlighten-nfc-tags.
- [31] Ok K., Coskun V., Aydin M. & Ozdenizci B. (2010) Current Benefits and Future Directions of NFC Services. In: Education and Management Technology (ICEMT), 2010 International Conference on, pp. 334–338.
- [32] Benyo B., Vilmos A., Kovacs K. & Kutor L. (2007) NFC Applications and Business Model of the Ecosystem. In: Mobile and Wireless Communications Summit, 2007. 16th IST, pp. 1–5.
- [33] NFC Forum (2011), Smart Posters How to use NFC tags and readers to create interactive experiences that benefit both consumers and businesses (accessed 17 Nov 2014). URL: http://67.222.41.204/wp-content/uploads/2013/12/NFC-Smart-Posters-White-Paper.pdf.
- [34] Alzahrani A., Alqhtani A., Elmiligi H., Gebali F. & Yasein M. (2013) NFC Security Analysis and Vulnerabilities in Healthcare Applications. In: Communications, Computers and Signal Processing (PACRIM), 2013 IEEE Pacific Rim Conference on, pp. 302–305.
- [35] Siira E. & Haikio J. (2007) Experiences from Near-Field Communication (NFC) in a Meal Service System. In: RFID Eurasia, 2007 1st Annual, pp. 1–6.
- [36] Häikiö J., Wallin A., Isomursu M., Ailisto H., Matinmikko T. & Huomo T. (2007) Touch-based User Interface for Elderly Users. In: Proceedings of the 9th International Conference on Human Computer Interaction with Mobile Devices and Services, MobileHCI '07, pp. 289–296.
- [37] Nordmark S. & Milrad M. (2012) Mobile Digital Storytelling for Promoting Creative Collaborative Learning. In: Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on, pp. 9–16.
- [38] Wong L.H. & Looi C.K. (2012) Enculturing Self-Directed Seamless Learners: Towards a Facilitated Seamless Learning Process Framework Mediated by Mobile Technology. In: Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on, pp. 1–8.
- [39] Alrasheedi M. & Capretz L. (2013) A Meta-analysis of Critical Success Factors Affecting Mobile Learning. In: Teaching, Assessment and Learning for Engineering (TALE), 2013 IEEE International Conference on, pp. 262–267.

- [40] Kloos C., Garcia R., Fernandez-Panadero C., Blanca Ibanez M., Munoz-Organero M. & Pardo A. (2012) m-learning willDisrupt Educational Practices. In: Global Engineering Education Conference (EDUCON), 2012 IEEE, pp. 1–6.
- [41] Picek R. & Grcic M. (2013) Evaluation of the Potential Use of m-learning in Higher Education. In: Information Technology Interfaces (ITI), Proceedings of the ITI 2013 35th International Conference on, pp. 63–68.
- [42] Pozgaj Z. & Vuksic V. (2013) Mobile Phone in the Classroom. In: Information Communication Technology Electronics Microelectronics (MIPRO), 2013 36th International Convention on, pp. 732–736.
- [43] Ye S.H. & Hung Y.C. (2010) The Study of Self-Seamless Teaching Strategy for Ubiquitous Learning Environments. In: Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), 2010 6th IEEE International Conference on, pp. 182–186.
- [44] Kim S. & Yoon Y. (2009) Multimedia collaborative adaptation middleware for personalization E-learning. In: Collaborative Technologies and Systems, 2009. CTS '09. International Symposium on, pp. 558–564.
- [45] Ervasti M., Isomursu M. & Kinnula M. (2009) Experiences from NFC Supported School Attendance Supervision for Children. In: Mobile Ubiquitous Computing, Systems, Services and Technologies, 2009. UBICOMM '09. Third International Conference on, pp. 22–30.
- [46] Benyo B., Sodor B., Doktor T. & Fordos G. (2012) Student attendance monitoring at the university using NFC. In: Wireless Telecommunications Symposium (WTS), 2012, pp. 1–5.
- [47] Bucicoiu M. & Tapus N. (2013) Easy Attendance: Location-based authentication for students integrated with Moodle. In: Roedunet International Conference (RoEduNet), 2013 11th, pp. 1–4.
- [48] Miraz G.M., Ruiz I.L. & Gómez-Nieto M.A. (2009) How NFC Can Be Used for the Compliance of European Higher Education Area Guidelines in European Universities. In: Proceedings of the 2009 First International Workshop on Near Field Communication, NFC '09, pp. 3–8.
- [49] Benyo B., Sodor B., Doktor T. & Fordos G. (2012) University life in contactless way - NFC use cases in academic environment. In: Intelligent Engineering Systems (INES), 2012 IEEE 16th International Conference on, pp. 511–514.
- [50] Borrego-Jaraba F., Ruiz I. & Gomez-Nieto M. (2012) NFC Solution for Access to Bibliographic Sources. In: Global Engineering Education Conference (EDUCON), 2012 IEEE, pp. 1–7.
- [51] Sánchez I., Cortés M., Riekki J. & Oja M. (2011) NFC-based Interactive Learning Environments for Children. In: Proceedings of the 10th International Conference on Interaction Design and Children, IDC '11, pp. 205–208.

- [52] Ivanov R. (2013) NFC-based Pervasive Learning Service for Children. In: Proceedings of the 14th International Conference on Computer Systems and Technologies, CompSysTech '13, pp. 329–336.
- [53] Garrido P., Miraz G., Ruiz I. & Gomez-Nieto M. (2011) Use of NFC-based Pervasive Games for Encouraging Learning and Student Motivation. In: Near Field Communication (NFC), 2011 3rd International Workshop on, pp. 32–37.
- [54] Zhang B. & Looi C.K. (2011) Developing a sustainable education innovation for seamless learning. Procedia Social and Behavioral Sciences 15, pp. 2148 2154. 3rd World Conference on Educational Sciences 2011.
- [55] Ryane I., Idrissi M.K. & Bennani S. (2011) A proposition of an authoring tool, for pedagogical scripting, adapted to teachers. In: International Journal of Engineering Science and Technology (IJEST), vol. 3, vol. 3, pp. 8226–8237.
- [56] Annan N., Ofori-Dwumfuo G. & Falch M. (2012) Authoring m-learning Content: A case study of using power point mobile enabled tools to create content for learning anywhere anytime. In: Education and e-Learning Innovations (ICEELI), 2012 International Conference on, pp. 1–5.
- [57] Nurjanah D., Davis H. & Tiropanis T. (2010) A Framework of Collaborative Adaptation Authoring. In: Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2010 6th International Conference on, pp. 1–6.
- [58] Abdullah N., Raja R.H., Kamaruddin A., Razak Z. & Yusoff M.Z.B.M. (2008) An Authoring Toolkit Design for Educational Game Content. In: Information Technology, 2008. ITSim 2008. International Symposium on, vol. 1, vol. 1, pp. 1–6.
- [59] He Z. & Ueno H. (2012) Designing m-learning System Based on WebELS System. In: Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, ICIMCS '12, pp. 115–118.
- [60] Pyykkönen M., Riekki J., Jurmu M. & Sanchéz Milara I. (2013) Activity Pad: Teaching Tool Combining Tangible Interaction and Affordance of Paper. In: Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces, ITS '13, pp. 135–144.
- [61] Wong L.H. & Looi C.K. (2010) Mobile-Assisted Vocabulary Learning in Real-Life Setting for Primary School Students: Two Case Studies. In: Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), 2010 6th IEEE International Conference on, pp. 88–95.
- [62] Ahmad M. (2012) A Generalized Framework to Support Field and In-class Collaborative Learning. In: Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on, pp. 333–336.

- [63] Marchiori E., Torrente J., del Blanco A., Martínez-Ortiz I. & Fernandez-Manjon B. (2010) Extending a Game Authoring Tool for Ubiquitous Education. In: Ubimedia Computing (U-Media), 2010 3rd IEEE International Conference on, pp. 171–178.
- [64] Uosaki N., Ogata H., Sugimoto T., Hou B. & Li M. (2012) How We Can Entwine In-class Vocabulary Learning with Out-class one in English Course for Japanese EFL Learners. In: Wireless, Mobile and Ubiquitous Technology in Education (WMUTE), 2012 IEEE Seventh International Conference on, pp. 102–106.
- [65] Yin R.K. (2009) Case Study Research: Design and Methods. SAGE Publications, Los Angeles, 4th ed.

# 10. APPENDICES

Appendix 1. Documents needed for usability testing phase 2

# Appendix 1. Documents needed for usability testing phase 2

- The pretest questionnaire ('NFC-ACT General info').
- The SUXES questionnaires (Expectations and experiences questionnaires).
- Description of scenarios ('NFC-ACT Usability Test Session'). Note! In the actual testing session each scenario should be presented on a separate page (denoted as page break in the document).
- Observation notes example template for the test monitor.

DATE:

NFC-ACT – General info:			
1. Please select which cate	gory below describes yo	ur age:	
[ ] 17 or younger	[ ] 30 – 39	[ ] 60 or older	
[ ] 18 – 20	[ ] 40 – 49		
[]21-29	[]50-59		
2. What is your gender:	[ ] female	[ ] male	
3. Which subjects do you t	each:		
4. How many years have y	ou been as a teacher:		
5. Which of the following 6	edition software tools ha	ve you used before:	
[ ] Microsoft Office	[ ] Visual Studio	[ ] Adobe Premiere	
[ ] Adobe Illustrator	[ ] Photoshop	[ ] MovieMaker	
[ ] Others, please specify:			
6. Which of the following s	statements describes you	the most:	
[ ] I prefer creating as mu	ch teaching material as p	ossible by myself.	
[ ] I prefer creating teach	ing material using conter	t that I can find from internet/other sources.	
[ ] I prefer creating teach	ng material based on the	e material given by publishers.	
[ ] I prefer to use materia	I created by others.		
7. Are you familiar with NF applications based on NFC		ation) as a technology or have you previously use	≥(

## **Expectations questionnaire**

Please select how important for using NFCACT do you see the following statements when 1 describes low importance and 7 describes high importance. Please mark two values for each field: the acceptable level and the desired level. The acceptable level means the lowest acceptable quality level and the desired level is the uppermost level. For example:

		1	2	3	4	5	6	7
1	This color of this font is easy to read.		Х				Х	

## where 2= acceptable and 6=desired level

		1	2	3	4	5	6	7
1	Using NFC-ACT is intuitive/straightforward.							
2	It is easy to learn how to use NFC-ACT.							
3	The user interface is clear and easy-to-follow.							
4	Information is presented in an organised way.							
5	Symbols and icons in the user interface are clear and self-							
	explanatory.							
6	Using NFC-ACT is quick.							
7	Using NFC-ACT is pleasant.							
8	NFC-ACT provides enough help while it is being used.							
9	NFC-ACT benefits my work.							

## **Experiences questionnaire**

Please select how much the following statements describe using NFCACT from a scale of 1 to 7 where 1=using NFCACT is not at all like this and 7=using NFCACT is exactly like this. Please select only one value for each statement.

		1	2	3	4	5	6	7
1	Using NFC-ACT is intuitive/straightforward.							
2	It is easy to learn how to use NFC-ACT.							
3	The user interface is clear and easy-to-follow.							
4	Information is presented in an organised way.							
5	Symbols and icons in the user interface are clear and self-							
	explanatory.							
6	Using NFC-ACT is quick.							
7	Using NFC-ACT is pleasant.							
8	NFC-ACT provides enough help while it is being used.							
9	NFC-ACT benefits my work.							

DATE:

## **NFC-ACT USABILITY TEST SESSION**

You've been previously demoed a NFC based anatomy game and you have now also used the game yourself. In this test you are asked to recreate the game and its different tasks using the NFC-ACT editor. In this test you are asked to create two tasks to the game: pää and jalka.

This test consists of 3 different scenarios. After each scenario please say to the test monitor(s) when you feel you are finished with the scenarios. Please do not proceed to next scenario until the test monitor says you can proceed.

#### THANK YOU FOR PARTICIPANTIG IN THIS TEST!

WE MAY NOW BEGIN, PLEASE TURN TO PAGE 2.

[page break]

#### **SCENARIO 1**

Using the NFC-ACT editor, please start creating a new anatomy game in practicing mode and stop when you feel you have reached a point where you would be starting to create different tasks to the game.

Please turn to page 3 when the test monitor advises this.

[page break]

#### **SCENARIO 2**

Using the NFC-ACT editor, please create the first task into the game. In this task the word for 'pää' in English needs to be found. Please stop when you feel you have finished creating the first task.

Please turn to page 4 when the test monitor advises this.

[page break]

#### **SCENARIO 3**

Using the NFC-ACT editor, please create the second task into the game. In this task the word for 'jalka' needs to be found. Please stop when you feel you are finished.

## **OBSERVATION NOTES**

# Task 1: Creating a game template

Success rate:	S	F	P	
Rationale:				 
Problems:				
Remember to provide and/or frustation), 3				