EVALUATING USABILITY IN VIDEO CONFERENCING SERVICE IN METSO

Mia Suominen

Master´s Thesis
April 2013

Degree Programme in Information Technology
Technology, communication and transport

JYVÄSKYLÄN AMMATTIKORKEAKOULU
JAMK UNIVERSITY OF APPLIED SCIENCES

Author
SUOMINEN, Mia

Type of publication
Master's Thesis

Date
22042013

Pages
66

Language
English

Permission for web publication
(X)

Title
EVALUATING USABILITY IN VIDEO CONFERENCING SERVICE IN METSO

Degree Programme
Master's Degree Programme in Information Technology

Tutors
RANTONEN, Mika
HAUTAMÄKI, Jari

Assigned by
PURANEN, Terho

Abstract
The main target of the thesis was trying to evaluate the usability of video conferencing service in Metso. It was known that normally usability evaluations are conducted at the early phase of designing and developing a product or a user interface; however, in the thesis it was decided to implement usability evaluation methods for assessing usability of a ready product in use for several years.

Questionnaire was chosen as the usability evaluation method, as it enables reaching a large group of users easily. The System Usability Scale (SUS) questionnaire was selected because it is free, short and quick to perform, it is not technology dependent and has references in hundreds of publications. In addition to traditional 10 items of SUS, users were requested to evaluate the user-friendliness of the system with an adjective rating scale. The respondents were also asked to give voluntary free comments regarding the service.

The analysis of the responses provided a SUS score result 67. Being a numeric value it does not provide much information on its own. There were no previous scores available and therefore it was not possible to compare against previous values. If compared to an overall SUS average of 68 it can be noted the usability of video conferencing in Metso is slightly below average. If compared to benchmark the usability level is way below average. Adjective rating scale provided an average result 4.76 which can be interpreted as OK. Totally 35 respondents gave comments about the video conferencing service in general.

The SUS score could have been expected to be higher as the end users were familiar with the use of video conferencing devices. The received SUS score is not very informative as such and does not provide solutions to improve the usability therefore turned out the feedback given by end users was more useful when thinking about concrete actions for improving usability level of video conferencing in Metso.

Keywords
usability, video conferencing, SUS, System Usability Scale

Miscellaneous

JYVÄSKYLÄN AMMATTIKORKEAKOULU
JAMK UNIVERSITY OF APPLIED SCIENCES

OPINNÄYTETYÖN
KUVAILULEHTI

Tekijä SUOMINEN, Mia	Julkaisun laji Opinnäytetyö	Päivämäärä 22.04.2013
	Sivumäärä 66	Julkaisun kieli Englanti
		Verkkojulkaisulupa myönnetty (X)

Työn nimi
EVALUATING USABILITY IN VIDEO CONFERENCING SERVICE IN METSO

Koulutusohjelma
Informaatioteknologian koulutusohjelma

Työn ohjaajat
RANTONEN, Mika
HAUTAMÄKI, Jari

Toimeksiantaja
PURANEN, Terho

Tiivistelmä

Työn tavoitteena oli yrittää arvioida Metson videoneuvottelupalvelun käytettävyyttä. Normaalisti käytettävyyttä tutkitaan ja arvioidaan tuotteen tai käyttöliittymän suunnittelu- ja kehitysvaiheessa, mutta tämän työn tarkoituksena oli soveltaa käytettävyyden arviointimenetelmiä jo olemassaolevan ja pitempään käytössä olleen tuotteen käytettävyyden arvioimiseen.

Käytettävyyden arviointimenetelmäksi valittiin kysely, koska sen avulla on mahdollista tavoittaa suuri käyttäjämäärä vaivattomasti. Kyselyksi valittiin SUS (System Usability Scale), koska se on ilmainen, lyhyt, nopea toteuttaa, teknologiariippumaton ja siihen viitataan useissa sadoissa julkaisuissa. SUSin sisältämän kymmenen vakiokysymyksen lisäksi kyselyssä pyydettiin käyttäjiä arvioimaan palvelun käyttäjäystävällisyyttä adjektiiviarvosteluasteikolla. Käyttäjille annettiin myös mahdollisuus antaa vapata palauteta videoneuvottelupalvelusta.

Vastaukset analysoitiin ja SUS arvoksi saatiin 67. Sellaisenaan tämä tulos ei kerro juuri mitään videoneuvottelupalveluiden käytettävyydestä. Tulosta ei voinut verrata aikaisempiin arvoihin, koska niitä ei ollut. Jos tulosta vertaa yleiseen SUS keskiarvoon 68, voidaan todeta käytettävyyden olevan hieman keskimääräistä huonompi. Verrokkiryhmiin verrattuna käytettävyys on selkeästi keskimääräistä huonompi. Kysymys, jossa käyttäjiä pyydettiin kuvaamaan käyttäjäystävällisyyttä adjektiivilla, tuotti vastaukseksi adjektiivin OK (numeerinen keskiarvo 4.76). Yhteensä 35 vastaajaa antoi vapaata palautetta videoneuvottelupalvelusta.

Koska kysely tehtiin käyttäjille, jotka olivat tutustuneet palvelun käyttöön aikaisemmin, olisi SUS arvon odottanut olevan korkeampi. Saatu arvo ei itsessään ole kovin informatiivinen eikä tarjoa keinoja käytettävyyden parantamiseen. Vapaat kommentit palvelusta olivatkin parasta antia ajatellen konkreettisia toimenpiteitä käytettävyyden parantamiseksi.

Avainsanat (asiasanat)
käytettävyys, videoneuvottelu, SUS, System Usability Scale

Muut tiedot

Contents

FIGURES

TABLES

ABBREVIATIONS

AVL	Adaptive video layering
B2B	Business to business
B2C	Business to consumer
CRM	Customer relationship management
CSUQ	Computer system usability questionnaire
fps	Frames per seconds
HCI	Human-computer interaction
HW	Hardware
IP	Internet protocol
ISO	International Organization of Standardization
IT	Information technology
IVR	Interactive voice response systems
kbit/s	Kilobit per second
LAN	Local area network
LTE	long-term evolution, marketed as 4G LTE, is a standard for wireless communication of high-speed data for mobile phones and data terminals.
Mbit/s	Megabit per second
MCU	Multipoint control unit
MPLS	Multiprotocol label switching
PSSUQ	Post study system usability questionnaire
QUIS	Questionnaire for user interaction satisfaction
SUMI	Software usability measurement inventory
SUS	System usability scale
SVC	Scalable video coding
SW	Software

VGA	Video Graphics Array
VPN	Virtual private network
QoS	Quality of service
WAN	Wide area network
3G	short for third Generation, the third generation of mobile telecommunications technology.
4G	short for fourth generation of mobile phone mobile communication technology standards.

1 INTRODUCTION

Video conferencing seems to be gaining more ground, as companies are interested in reducing their carbon footprint as well as cutting travelling costs. Video conferencing technology has also improved tremendously over the years and is now providing companies cost effective way to communicate.

Metso, which is a global company supplying sustainable technology and services for mining, construction, power generation, automation, recycling and the pulp and paper industries, has used video conferencing services for almost three years now. There are over 100 video room systems installed globally and the amount is increasing, as Metso has about 30,000 employees in more than 50 countries. The quality of the video and audio has been really good and therefore video conferencing has become a popular tool in internal communication between different locations globally.

Benefits of the video conferencing solution can be various. Polycom, which is known for its video solutions, has listed their opinion of top five benefits of video conferencing. According to Polycom, a large percentage of routine or regular business trips can be eliminated by communicating over video. This will soon show as reduced travel costs. Polycom also sees that the use of video conferencing increases productivity across dispersed workforces and teams. This is justified by the fact that a large amount of communication is actually based on non-verbal visual cues and by using video people will most likely stay more focused, as they can be seen and heard - and all this will finally result in increased productivity. One of the benefits according to Polycom is also improved hiring and retention of top talent as organizations with video conferencing systems can reduce expenses and time by bringing candidates into the nearest facility and allowing interviews to be conducted both in person and over video. They also suggest video communication impacts employee retention just as positively as there will be improved cooperation by allowing remote employees to become faster closer with other team members or helping employees retain

work/life balance by reducing travelling so they can spend more time with their families. Polycom also states that video conferencing offers multiple paths for creating and maintaining competitive advantage as teams can share knowledge more widely. One of the top benefits is naturally supporting environmental initiatives. By communicating over video, organizations can also substantially reduce their carbon footprint and help ensure a basis for regulatory compliance. (Polycom Fact Sheet: The Top Five Benefits of Video Conferencing, 2010.)

As the author of this thesis read these Polycom's views about the benefits of video conferencing she started to wonder if these statements are all true. Are video devices used and utilized as well as they could, as presumed by Polycom? Surely productivity is not increased, if it takes 10-15 min before a successful video meeting can be established as users find devices hard to use? Are video conferencing devices actually easy to use, what is their usability like?

Before one can answer those questions one has to consider what usability is, how it can be evaluated and what kind of methods there are. Usability as a concept seems more to be about designing usable user interfaces and www-pages. However, Kuutti (Kuutti, 2003, 13) defines usability as a feature which describes how fluently users can achieve their goal when using the functions of a product. It is also said that bad usability of applications can even cause threats to business strategy. If a system is not learnable and it is difficult to adopt this can in worst case prevent or slow down products and services becoming general. (Wiio, 2004, 38) These definitions got the author interested in, what the usability is like regarding video conferencing service in Metso.

The author works for Metso Shared Services, and to be more specific, for Metso IT. Metso IT is an internal organization providing common information technology (IT) infrastructure and application services for all Metso's businesses. Video conferencing is one the many IT services provided by Metso IT. The author's current responsibility is to manage the video videoconferencing service and continuously improve the service together with the service provider. Experiences about the service and its functionality all in all over the past years have been rather good; however, still in

some cases there are complaints how difficult it is to start a video meeting and how challenging it is to use the devices, which inspired the author to find out if there was a way to discover how end users experience the usability in the video conferencing system in Metso.

The main purpose of this study was to find out what is usability and if there is a way to measure or evaluate the current usability level of video conferencing service in Metso. If video conferencing devices are easy to learn and use should this not be seen in end users' opinion about usability as a good score? As far as the author has understood, usability is not normally evaluated like this, with a product already in use and with end users being familiar using the product. Usability is - and of course should be - normally taken into consideration when designing and developing a product; usability tests are performed to see what could be done better for example with the user interface. Could usability tests or questionnaires, however, be used from the end user point of view as well instead of a tool meant only for developers? The goal was to find out whether there is a quick and easy way to determine the usability in video conferencing service. If this could be done, what it would inform on and is there a way to utilize these results to improve the overall usability level? Perhaps training affect on the opinion of usability – if users are trained better, do they feel usability is also improved? If the current level of usability can be evaluated, is it worth while doing it again, just to follow the results on a regular basis?

As the author is the service manager of the video conferencing service, the aim is to do best so that the service is easy to use and end-users will find it usable – they are able to achieve their goals when using the video conferencing service. With this thesis effort was made to find if the current level of usability can be easily evaluated and even better, the situation improved. If this study will produce improvement ideas to user interface, technology provider is certainly happy to hear the suggestions and perhaps it could consider taking some of them into account when planning the next version of the software. After all, it is the best possible feedback: coming from real end users who are really using their product in real cases in daily work – not in some simulated test situation in usability laboratories.

2 USABILITY

2.1 Definition of Usability

When talking about the definition of usability, Jakob Nielsen is perhaps the most quoted author. Nielsen sees usability as one attribute of system acceptability. System acceptability on the other hand is basically the question of whether the system is good enough to satisfy needs and requirements of the users and other potential stakeholders. (Nielsen, 1993, 24.)

Figure 1 illustrates Nielsen's (1993, 25) model of the attributes of overall system acceptability more closely. System acceptability consists of social and practical acceptability. One attribute of practical acceptability is usefulness, which according to Nielsen is the issue of whether system can be used to reach some desired goal. Usefulness can be divided into two categories; utility and usability. Utility defines whether the system is capable of performing what it is supposed to do and usability answers the question how well users can use the functionality. (Nielsen, 1993, 24-25.)

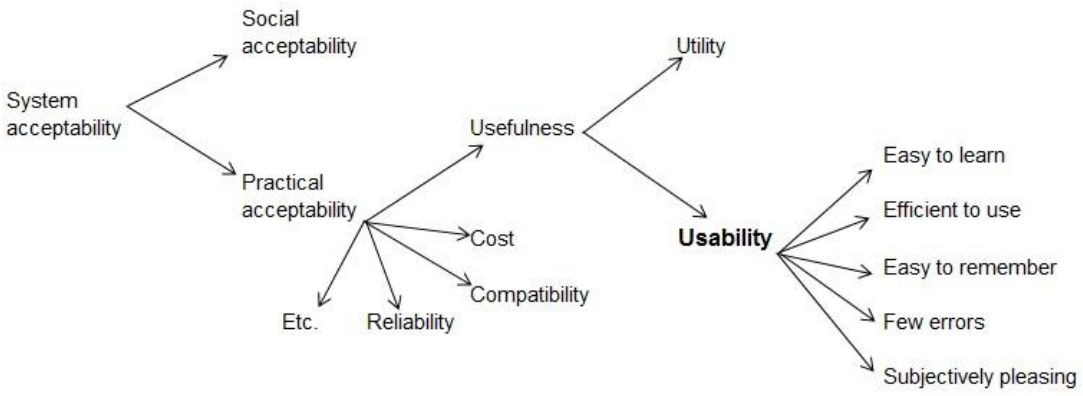


FIGURE 1. A Model of the attributes of system acceptability (Nielsen, 1993, 25).

As Figure 1 illustrates, there are five attributes associated with usability: learnability, efficiency, memorability, errors and satisfaction (Nielsen, 1993, 26). According to

Nielsen by defining usability in terms of these more measurable components, it is possible to approach, improve and evaluate usability in a more systematic way. Therefore we will have a closer look of these five attributes. (Nielsen, 1993, 26.)

- Learnability is perhaps the most fundamental usability attribute as it is quite obvious that systems should be easy to learn in order for user to start working fast with the system.
- Efficiency to use. System should be so efficient to use, so that once user has learned the system, a high level of productivity is possible.
- Memorability. Systems should be so easy to use that casual user remembers how to use it after some period of not having used it, without having to learn it all over again. By casual users Nielsen means people who are using a system occasionally rather than frequently like expert users.
- Few errors. The system should not have catastrophic errors, on the contrary it should have such a low error rate, so that users would not perform that many errors when using the system – and if errors are made users would easily recover from them. Nielsen defines error as any action which does not accomplish the desired goal.
- Subjective satisfaction. This attribute refers to how pleasant it is to use the system – users should like using it.

The International Organization of Standardization (ISO) has defined usability in their standard 9241-11. According to this standard, usability is defined as “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”. (SFS-EN ISO 9241-11, 1998, 2).

To be able to define or measure usability it is necessary to indentify goals, which users are meant to achieve, and divide effectiveness, efficiency and satisfaction and the components of the context of use into sub-components which contain measurable and verifiable attributes. The usability framework according to SFS-EN ISO 9241-11 is presented in Figure 2:

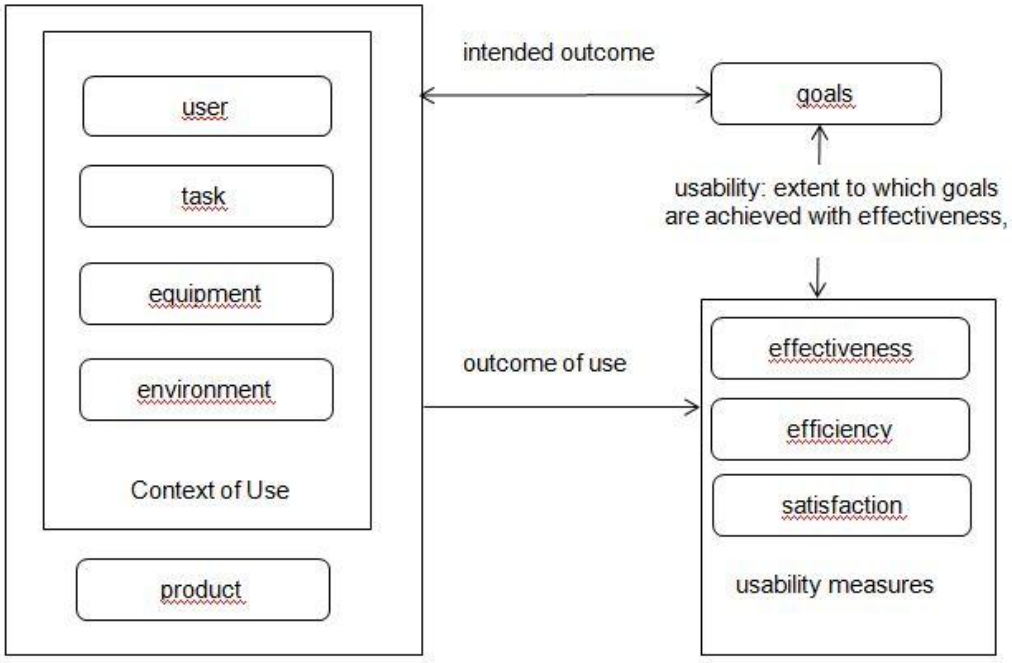


FIGURE 2. Usability framework (SFS-EN ISO 9241-11, 1998, 3).

The goals of using a product should be defined. Goals can be divided into sub goals which specify components of an overall goal and the criteria which would satisfy that goal. Then the context of use is described. This contains describing the users, tasks, equipment and environment. One has to describe the characteristics of the users, which can be for example experience, skills, knowledge, education and training. The description of tasks contains such activities that need to be taken in order to achieve a goal. Features potentially influencing the usability should be described. When evaluating usability, a set of key tasks will typically be chosen to represent the significant aspects of the overall task. Equipment characteristics should be described. This can be done for example by listing attributes or performance characteristics of the hardware, software and other materials. Environment characteristics could include describing things like the physical environment (meaning like workplace, furniture), the ambient environment (like temperature, humidity and further) and the social and cultural environment (issues like work practices, organizational structure and attitudes). (SFS-EN ISO 9241-11, 1998, 4.)

For measuring the usability ISO suggests to provide at least one measure for each for effectiveness, efficiency and satisfaction, and if it is not possible to gain objective measures, subjective measures based on the user's perception can provide an indication of effectiveness and efficiency. (SFS-EN ISO 9241-11, 1998, 5.)

However, according to Faulkner effectiveness in this ISO standard definition simply means that a user is able to perform the intended task – time is not taken into consideration or the ease of use. With efficiency time is an essential factor. The faster a task can be performed with a system, the more efficient it is. Faulkner states ISO makes no mention of learnability here. ISO also refers user satisfaction with the system which, according to Faulkner, can be defined how acceptable the system is from user's point of view, do they feel comfortable when operating the system and whether they prefer one system over another (Faulkner, 2000, 7-8.)

2.2 Usability and Human-Computer Interaction

When talking about usability of different applications the term human-computer interaction (HCI) is often used beside usability. HCI according to Preece, Rogers, Sharp, Benyon, Holland and Carey (1994) is about "*designing computer systems that support people so that they can carry out their activities productively and safely"*.

The goals of human-computer interaction are defined to produce systems which are usable and safe to use and at the same time functional. According to Preece *et al.* (1994) this was summarized in Interacting with computers (1989) as *"to develop or improve the safety, utility, effectiveness, efficiency and usability of systems that include computers".* Preece *et al.* state that usability is a key concept in HCI and its main goal is to make systems easy to learn and use. (Preece *et al,* 1994, 14.)

So in order to be able to produce usable computer systems, HCI specialists aim at:

- Understanding the factors determining people operating computer technology effectively.
- Having that understanding they try to develop tools and techniques to help designers to produce systems suitable for people using them.
- Users should be able to achieve efficient, effective and safe interaction, when using these systems. (Preece *et al*, 1994, 15.)

Preece *et al.* share the opinion that HCI research and design are based on the belief that people using a computer system should come first. Very often people have to adjust themselves to the system –this should not be the case; the system should be designed to match the user requirements. (Preece *et al*, 1994, 15.)

According to Sinkkonen, Kuoppala, Parkkinen and Vastamäki (2006) usability and HCI are seen as exchangeable terms, even in IT related publications. However, in theory HCI does not consider the person as a part of an organization, as an actor with an independent will, whereas usability takes these aspects of HCI into consideration as well. (Sinkkonen *et al*, 2006, 11.)

2.3 What Makes Something Less Usable?

When thinking about usability one cannot avoid thinking what makes something less usable? Rubin and Chisnell (2008, 44) have listed five main reasons why products are so hard to use.

- Development focuses on the machine or system.
- Target audiences change and adapt.
- Designing usable products is difficult.
- Team specialists do not always work in integrated ways.
- Design and implementation do not always match.

According to Rubin and Chisnell when designing and developing a product one does not pay attention to the ultimate end user rather than focus on the machine or the system.

Reason two states that target audiences can change and adapt rapidly and development organizations have been slow in reacting to this evolution. Rubin and Chisnell state that original users of computer-based products were kind of "geeks" – loving technology, desired to tinker and possessing more knowledge of computers and mechanical devices; also the developers of these products shared the same characteristics which meant users and developers were kind of one and the same. Thus machine-oriented or system-oriented approach could easily have been seen as the development norm. Compared to nowadays where users have little technical knowledge, they do not want to tinkle newly purchased device and have different expectations of the designer. In fact, today's users are not comparable to the product designer in almost any attribute relevant to the design process. So if there is a great discrepancy between the user and designer companies will continue producing hard-to-use products. (Rubin & Chisnell, 2008, 46.)

Designing usable products is difficult, yet according to Rubin and Chisnell many organizations treat it as if it was just "common sense" and it is being trivialized. Rubin and Chisnell share the opinion that usability principles are not obvious and there is still a great need for education, assistance and a systematic approach in applying usability to the design process. (Rubin & Chisnell, 2008, 47.)

For product and system development organizations employ very specialized teams and approaches; however, somehow they manage to fail integrating them with each other. There is actually nothing wrong with this kind of specialization but it might cause difficulties when there is little integration of these specialized components/teams and poor communication between different development teams. If each development group functions independently the result can be seen in the final product – for example user documentation and help will be redundant with little cross-referencing. Rubin and Chisnell state that organizations unknowingly

worsen this lack of integration performing usability testing separately for each component. (Rubin & Chisnell, 2008, 48.)

The last reason on the list is how design and implementation do not always match. Rubin and Chisnell see design relating how the product communicates, whereas implementation refers to how it works. Previously this difference was rarely even acknowledged and designers were hired because of their technical expertise (programming) rather than for their design expertise. However, nowadays the challenge of technical implementation has decreased and the challenge of design has increased due to the need to reach broader, less sophisticated users and the rising expectations for ease of use. Therefore, the focus on required skills for developers has also changed toward design. (Rubin & Chisnell, 2008, 49.)

With this list of five reasons Rubin and Chisnell wanted to brush the surface of how and why unusable products and systems continue to exist. However, they wanted to emphasize that too much focus has been placed on the product itself and too little on the wanted effects the product needs to achieve. Somehow the user continues to receive too little consideration and attention in the heat of development process. (Rubin & Chisnell, 2008, 49.)

3 USABILITY EVALUATION METHODS

3.1 Heuristic Evaluation

Heuristic evaluation of usability is based on heuristics. Heuristics are lists of rules and guidelines for a usable user interface. There are a lot of gathered heuristics, some of them are more general and meant to be used with all kinds of user interfaces, some more narrow and suitable only in specified user interfaces. Especially the earlier

heuristics used to be rather wide containing as many as one thousand guidelines. However, these are not very practical when evaluating usability as people cannot remember nor evaluate this many rules regarding a product. (Kuutti, 2003, 47.)

According to Nielsen heuristic evaluation is carried out by having a look at the interface and trying to form an opinion what is good and what is bad about that interface. Ideally this evaluation would be performed according to certain rules and list but most likely people will perform evaluation on the basis of their own intuition and common sense. However, Nielsen describes heuristic evaluation as a systematic inspection of a user interface design for usability. The goal is to find the usability problems so that they can be fixed as a part of an iterative design process. This type of evaluation involves a small group of evaluators examining the interface and comparing its compliance with predefined usability principles (heuristics). (Nielsen, 1993, 155.)

The following usability principles listed by Nielsen was originally developed by Nielsen and Rolf Molich and it was designed for interface designers.

- Simple and natural dialogue. Dialogues should not contain information which is irrelevant or needed only every now and then.
- Speak the users' language. One should avoid using system-oriented terms but use words and terms familiar to the user.
- Minimize user memory load. Instructions should be visible and easily retrievable whenever possible.
- Consistency. Use consistent language so that user does not have to wonder whether different words or actions means the same.
- Feedback. System should give appropriate feedback within reasonable time about what is going on.
- Clearly marked exits. System should provide clearly marked exits for example situations where user has accidentally entered system functions and needs a fast exit out.
- Shortcuts. There should be shortcuts which accelerate the use of the system.

- Good error messages. Messages should be in plain language, precisely indicating the problem and suggest for a solution.
- Prevent errors. Even more recommended than good error messages is to prevent errors from happening with careful design.
- Help and documentation. It is good if system can be used without documentation but it may be necessary to provide some. Such information should be easy to find, focused on user's task, not to be too large and provide concrete steps on how to proceed. (Nielsen, 1993, 20.)

Heuristics can be used for evaluating a prototype as well as a product which is already in use. Of course evaluation produces more value if performed to a prototype because it is possible to notice usability issues in an early phase. Heuristic evaluation has also been used in iterative product development. In this case the tested usability issues will be fixed, tested again and this will be done as long as the product is stabilized. (Kuutti, 2003, 48.)

The output from this kind of heuristic evaluation is a list of usability issues with references to the usability principles that were violated. It should be noted that this evaluation type does not provide a systematic way to generate fixes to the problems found. (Nielsen, 1993, 159.)

In principle, according to Nielsen, individual evaluators can conduct an evaluation on their own; however, studies show that any single evaluator will miss most of the usability issues in an interface. It was noted the single evaluators found only 35% of the usability problems. Then again single evaluators usually pay attention to different topics so increasing the amount of evaluators and aggregating their results it is possible to reveal more usability issues. Nielsen recommends the use of five evaluators, as studies have revealed the proportion of found usability problems increases very rapidly when using more evaluators. Increasing the amount of evaluators from 5 to 10 does not increase the proportion of found usability problems as it does from 1 to 5. (Nielsen, 1993, 155-156.)

Evaluation will be performed in such a way that each evaluator studies and inspects the interface alone. During the evaluation session the evaluator goes through the interface several times, examines dialogue elements while comparing them with the heuristic list. In principle, evaluators can decide independently how to proceed with the evaluation. After evaluations are conducted the evaluators can communicate and have their findings aggregated. This is important in order to get independent and unbiased evaluations from each of the evaluators. Evaluation results can be either written down as a report or an observer will gather the comments from evaluators as they go through the interface. Written reports are normally more formal but they require extra work from both evaluators and evaluation managers. (Nielsen, 1993, 158.)

If one compares heuristic evaluation with traditional user tests two differences can be distinguished:

- Will the observer answer the questions from the evaluators?
- How much can observers give tips to evaluators on using the interface?

In traditional user testing observer does not answer questions or provide tips, unless it is absolutely needed. This is because in traditional user testing users should use the system to find answers to their questions rather than getting answers directly from an expert. Also user tests are meant for discovering the mistakes done by users. (Nielsen, 1993, 158.)

3.2 Usability Testing

Testing usability with real users is the most fundamental usability method and according to Nielsen can be seen in some way irreplaceable. User testing provides direct information about how people are using the system and what their concrete problems are. (Nielsen, 1993, 165.)

Kuutti states that user tests and heuristic evaluation are not competing methods, nor do they exclude one another. They are two different kinds of methods which reveal different kind of usability issues. In practice, more than one method is used in parallel to achieve better results. (Kuutti, 2003, 69.)

According to Nielsen in usability testing, as in all kind of testing, one needs to pay attention to reliability and validity. Reliability answers the question whether the same results would be received again if the test was repeated. Validity is about whether the result actually reflects the usability issues one is looking to test. Reliability is a problem in usability testing because there are huge individual differences between test users. Validity, on the other hand, requires methodological understanding of the test method used as well as common sense because typical validity problems involve using the wrong users or giving them wrong tasks. (Nielsen, 1993, 165 – 169.)

Usability testing can be divided into three larger phases according to Kuutti (2003, 70):

- Preparing the test.
- Conducting the test.
- Analyzing the test results.

Preparing the test is a very demanding process. One has to pick up test users, decide what areas one wants to emphasize and compile the test tasks. It is also good to check and prepare the devices being used in the test and perhaps perform a pilot test. (Kuutti, 2003, 74.) The usability test itself typically has four stages; preparation, introduction, the test itself and debriefing (Nielsen, 1993, 187). In preparation it is verified that a room is ready, materials are available, computers are in the start stage and further on. During the introduction test users are briefed of the purpose of the test, computer setup is introduced if necessary and test procedure is explained. During the test itself the experimenter of the test should not interact with the test users unless a user is clearly stuck and not happy with the situation. After the test users are debriefed and asked to fill in subjective satisfaction questionnaires.

(Nielsen, 1993, 187 -191.) During the usability test a huge amount of information is gathered. This information should be processed and transformed so that it is easy to analyze. Of course the main target is to find out if the test revealed any usability issues, what might have caused them and how they could be fixed. It is good to note that in most cases these tests generate more new questions rather than give answers. (Kuutti, 2003, 78 -80.)

3.3 Other Methods

In addition to heuristic evaluation and usability testing there are other usability methods which can be used to gather data. Nielsen (1993, 223) suggests at least the following methods:

- Observation.
- Questionnaires and interviews.
- Focus groups.
- Logging actual use.
- User feedback.

Observation is a very simple usability assessment method as it only involves the observer visiting users and observing them working with applications. Observers' goal is to intrude as little as possible and stay almost invisible so that users can perform their work normally with the system. It might be surprising to notice how users have found almost unexpected ways to use the system. (Nielsen, 1993, 207-208.)

Questionnaires and interviews are an excellent way to find out issues related to users' subjective satisfaction and possible anxieties, which are hard to measure objectively. This method is also great for finding out how users use systems and what features are like or disliked. However, questionnaires and interviews are considered

to be indirect methods as they study about users' opinions about the user interface rather than study the interface itself. (Nielsen, 1993, 209-210.)

As a method both are rather similar ones as both include asking users a set of questions and recording their answers. Questionnaires are printed on a paper or presented via computer and can be performed without anyone supervising the situation. Interviews, on the other hand, involve an interviewer, who will present the questions and also record the responses. Interviews can be more free-form than questionnaires which will make it more difficult to analyze the data quantitatively. Questionnaires are better if hard numbers are the main goal. (Nielsen, 1993, 209-210.)

Questionnaires are probably the only usability method, which enable such an extensive coverage as they could be distributed to the entire group of users. In practice, a target group is often limited to a randomly selected sample of users, depending how detailed data one is looking for. Questionnaires are usually administrated by mail according to Nielsen; however, nowadays e-mail and web questionnaires have replaced normal paper versions. Interviews can be done over the phone or personally, which gives the method quite high response rates. This type of method is recommended to situations where one does not know what one is actually looking for (Nielsen, 1993, 210 - 211.)

One thing in common with interviews and questionnaire is that you can not necessarily trust all the answers received from the users. In some cases where people find certain answers perhaps embarrassing or they think it might be considered socially unacceptable, people seem to answer as they think they are expected to answer. Thus, one should always consider the possibility that the situation is somewhat different from that indicated by the users in the case of such sensitive questions. (Nielsen, 1993, 212 - 213.)

Focus group is considered to be somewhat informal technique. It can be used to assess user needs and feelings both before the interface has been designed as well as after it has been used for a while. Basically, the focus group consists of a small group of users who discuss new concepts and recognize issues for a period of time. In each

group there is a moderator responsible for maintaining the focus on the issues of interest. The focus group should contain at least six participants in order to keep the conversation going. Also, it is recommended to run more than one group in order to get comprehensive results. (Nielsen, 1993, 214 - 215.)

Logging the actual use requires a computer to collect statistics about the use of the system. Normally this method is used after release; however, it can also be used during user testing to collect more detailed data. This is a very useful way to collect data because it shows how people perform their actual work and this method also allows data collection from a large number of users. However, logging user's system use might raise some privacy issues, which can be addressed by explaining how only summary statistics are being collected and individually users cannot be identified from the results. Logging is a very efficient way of gathering data compared to other usability methods as it is not interfering with the users in any way. (Nielsen, 1993, 216 - 220.)

User feedback can be considered as a major source of usability information. It also has advantages like showing users' immediate and pressing concerns, generating continuous feedback without any special effort of collecting it and showing quickly if users' needs, circumstances or opinions have changed. However, user feedback may not always represent the opinion of majority of users as the most dissatisfied ones give most feedback. There are several ways to collect user feedback – e-mails, bulletin boards, network newsgroups, software beta testing – but no matter how the feedback is collected it is important to make the users, who gave the feedback, feel their feedback is taken seriously. If this does not take place, users will soon end up giving feedback and this valuable source of information will be lost. (Nielsen, 1993, 220 - 222.)

3.4 Choosing Usability Method

Appendix 1 contains Nielsen's summary of these presented methods. According to Nielsen these methods are intended to supplement each other, since their advantages and disadvantages can partly make up for each other and because these methods address different parts of the usability lifecycle engineering. Therefore Nielsen highly recommends not relying on a single usability method to the exclusion of others. (Nielsen, 1993, 223-224.)

Choice of method may also be partly dependent on the number of users available for usability activities. If it is possible to reach a large amount of users one could perform questionnaires or systematic collecting of user feedback whereas heuristic evaluation should be considered if only very few users are available. Also, the experience of the usability staff available may also have an impact on choosing the method. For example a focus group moderator needs to be able to react to group dynamics in real time. (Nielsen, 1993, 224.)

4 QUESTIONNAIRES AS USABILITY EVALUATION METHOD

As described earlier questionnaires are an excellent way to find out how users use systems and what features they like or dislike. Questionnaires have turned out to be better if hard numbers are the main goal and they are probably the only usability method, which enables such an extensive coverage as they could be distributed to the entire group of users. So, what kind of questionnaires are there available for measuring usability?

It turns out there are several of them, some measuring the overall satisfaction to a system and some the noticed ease of use. Some of the most known questionnaires are introduced here.

The Questionnaire for User Interaction Satisfaction (QUIS) measures overall system satisfaction and nine specific interface factors (screen factors, terminology and system feedback, learning factors, system capabilities, technical manuals, on-line tutorials, multimedia, teleconferencing, and software installation). Each area measures the users' overall satisfaction with that facet of the interface, as well as the factors that make up that facet, on a 9-point scale. (Questionnaire for User Interaction Satisfaction, University of Maryland)

The Software Usability Measurement Inventory (SUMI) is a method of measuring software quality from the end user's point of view. It consists of 50 statements to which the user has to reply that they either Agree, Don't Know, or Disagree. SUMI is recommended to any organization who wishes to measure the detected quality of use of software. (SUMI Questionnaire homepage)

The System Usability Scale (SUS) is a simple ten-item scale giving a global view of subjective assessments of usability. Developed as a part of the usability engineering program at Digital Equipment Co. Ltd. SUS has proved to be a valuable evaluation tool which correlates well with other subjective measures of usability. (Brooke, 1996, 194.)

The Post Study System Usability Questionnaire (PSSUQ/CSUQ) is currently a 19-item questionnaire. Practically it is the same as the CSUQ (Computer System Usability Questionnaire), developed at IBM. They are both considered as overall satisfaction questionnaires. The PSSUQ questions are more suitable for a usability testing situation, and the CSUQ items are perhaps more appropriate for a field testing situation. Otherwise, the questionnaires are identical. (Lewis, 1993, 14-20.)

4.1 Comparing Questionnaires

Tom Tullis and Jacqueline Stetson of Fidelity Investments and Bentley College compared five questionnaires used for assessing website usability. In their study they compared SUS, Words (adapted from Microsoft's Product Reaction Cards), QUIS, CSUQ and their own questionnaire. The study was conducted with 123 participants and each of the participants performed two tasks in two websites (finance.Yahoo.com and kiplinger.com). This was to test the questionnaires´ ability to correctly identify which one of the two pages is more usable. (Tullis & Stetson, 2004, 1.)

Normally in usability tests a larger sample size is preferred to get more reliable results. Tullis and Stetson also wanted to find out whether any of the studied questionnaires would yield reliable results when the sample size is smaller than normally used in usability tests. They found out that one of the tests (SUS) increased its accuracy quicker than others. With sample size 6, all the questionnaires yield accuracy of no more than 30-40%. However, with SUS, sample size of 8 increased accuracy up to 75% while others remained in 40-55% range. It was also noted that most of the questionnaires seem to reach an asymptote when the sample size was 12. When going to sample size 14 the improvement was small in most cases. (Tullis & Stetson, 2004, 6.)

In their study Tullis and Stetson (2004) noticed that one of the simplest questionnaires (SUS) turned out to be one of those with the most reliable results across all sample sizes. According to them, from the studied questionnaires, SUS was the only one containing questions which address different aspects to the user's reactions to the website as a whole. Although, one has to keep in mind that due to the nature of the study, one should not draw too straightforward conclusions from the results, however, they still are very interesting indeed.

5 SUS – THE SYSTEM USABILITY SCALE

The System Usability Scale (SUS) was developed by John Brooke in 1986 as a part of usability engineering program in Digital Equipment Co. Ltd, Reading, UK. It has been referred as "quick and dirty" usability scale because it was developed to meet the needs of evaluating usability of systems within an industrial context. There was a need for cost-effective, practical, simple and fast way to evaluate usability and get an indication of the overall usability level compared to its competitors or previous versions of the software product. (Brooke, 1996, 190-194.) According to Jeff Sauro SUS is not dependent on technology and it has been tested not only with hardware and websites but also on consumer software, mobile phones and even with yellow-pages. Sauro also states that SUS has become an industry standard and it has references in over 600 publications. (Sauro, 2011, 10.)

SUS in short is a simple, ten-item scale which, according to its developer John Brooke, gives a global view of subjective assessments of usability. It consists of ten statements, which cover various aspects of system usability, such as complexity and the need for training and support. SUS questionnaire items are presented below: (Brooke, 1996, 192-193.)

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.

10. I needed to learn a lot of things before I could get going with this system.

However, according to studies about interpretation of SUS by non-native English speakers, it is suggested to replace word cumbersome to awkward in item 8 to avoid confusion. Also some studies suggest it might be better to use word "product" instead of "system" if it seems more appropriate. These minor changes did not lead on detectable differences on reliability. (Lewis & Sauro, 2009, 9.)

Statements are evaluated with five-level Likert items, as presented in Figure 3.

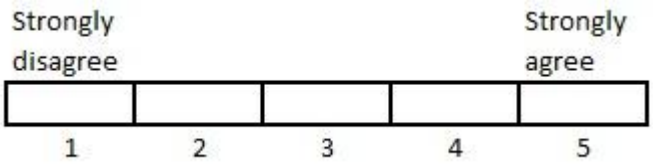


FIGURE 3. A five-level Likert Item.

As a result, SUS will produce a single number representing a composite measure of the overall usability of the studied system. The score is calculated by first summing the score contributions from each item. Each item's score contribution will range from 0 to 4. For odd items (1, 3, 5, 7, 9) one should subtract 1 from the user response. For even items (2, 4, 6, 8, 10) the contribution is 5 minus the user response. Then all these converted responses from one user are added up and this total sum is multiplied by 2.5. This way the overall value of system usability is obtained, as SUS scores have a range of 0 to 100. (Brooke, 1996, 194.) It is important to understand, John Brooke's original warning that "scores for individual items are not meaningful on their own" (Brooke, 1996, 194). However, lately there have been studies showing it would be possible, depending on the context, to examine the individual means and standard deviations of individual SUS items and compare them over time or to a benchmark.

SUS is generally used after the respondent has used the evaluated system but has not had any orientation or discussion has not taken place. They are asked to give

their immediate response to each item, rather than thinking about the statements too long. All items should be evaluated and if respondents do not have a clear opinion about some statement they should mark the centre point of the scale. (Brooke, 1996, 194.)

5.1 How to Interpret SUS Results?

As mentioned earlier, SUS produces a single number value which represents the overall usability of the studied system. This score can range from 0 to 100. According to Sauro (2011, 28), the best way to interpret a SUS score is to compare it to previous scores, benchmarks from the industry or to the overall SUS average value which is 68, according to Sauro's researches. Bangor, Kortum and Miller (2008, 576) on the other hand have performed over 2300 assessments and according to them an average SUS score would be 70.14. Bangor *et al*. claim that good systems get between 70-80 point and exceptional score 90 or more. If a product scores less that 70 it should be judged to be marginal at best. (Bangor *et al,* 2008, 592.)

As the SUS score ranges from 0 to 100 one might easily think this result can be seen as percentages. That is a common mistake; one should not call a scaled SUS score of 70 as "70 percent". It actually is technically correct that SUS score 70 represents 70% of the maximum score but calling it a percentage only confuses it with actual standardized scores. Because a score of 70 is so close to average score 68 (meaning it is around or at 50th percentile), calling 70 as 70% would suggest above average usability when it actually is a more likely average. (Sauro, 2011, 32.)

How can one tell what is a good score? If there are previous scores from the same system or similar ones, one can compare results to historical data. If there is no previous data to compare with, the score can be compared to SUS benchmarks. As mentioned earlier, the average score is 68. Anything above 68 can be considered as above average and below 68 naturally below average. If a score 76 or near to that is

reached that can be considered as a good SUS score, as it is then a higher score than 75% of all products tested. (Sauro, 2011, 32-33.)

Brooke (1996, 194) originally stated that scores for individual items are not meaningful on their own. However, Sauro (2011, 33) thinks that depending on the context it is possible to examine individual means and standard deviations of individual SUS items; however, in that case one has to be aware of the fact that there might be more errors in the measurements than at the aggregated level.

Despite the fact that SUS has been used so widely there seems to be very little guidance on how to interpret the score. The result, being a single number value, can raise questions how the numeric score translates into an absolute judgment of usability. Therefore Bangor, Kortum and Miller (2009) conducted a survey, where they added a seven-point adjective-anchored Likert scale as an eleventh question to nearly 1000 SUS surveys. By adding this adjective rating scale they hoped it would bring help to interpreting individual SUS scores and some aid in explaining the results to non-human factors professionals. The added eleventh question and its scale are presented in Figure 4.

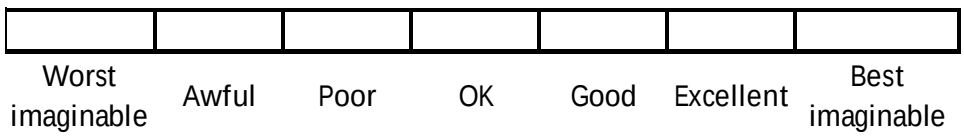


FIGURE 4. The adjective rating scale added to SUS.

Bangor *et al.* (2009, 119) found out indeed that as the adjective rating scale matches the SUS scale very closely it could be considered as a useful tool in providing a subjective label for an individual study's mean SUS score. It might also be very tempting to place entire SUS with this single item instrument, as it seems to correlate so well with the SUS score. However that is not recommended, as for example many studies have found out that multiple question surveys tend to yield more reliable results than single item surveys. Bangor *et al.* (2009, 120) also noticed that using OK

as option in the adjective rating scale might be too variable to use in this context and it might give the intended audience for SUS scores a mistaken impression that OK is satisfactory in some way, when it actually is not. OK can be connoted as satisfactory whereas the scores within OK range are telling that the perceived usability is clearly deficient.

5.2 Does SUS Measure Only Usability?

Originally SUS was designed to measure only usability. According to Sauro (2011, 85) it was long assumed all those ten questions of SUS questionnaire measure only usability and no other construct. However, in 2009 James R. Lewis and Jeff Sauro examined a set of SUS questionnaires and found in fact two detectable factors in SUS; usability and learnability (Lewis & Sauro, 2009, 5.)

According to Lewis and Sauro (2009, 5), eight items load on the usability factor and two items on the learnability factor. The two learnability terms are 4 ("I think that I would need the support of a technical person to be able to use this system") and 10 ("I needed to learn a lot of things before I could get going with this system"). These two gentlemen state that without any extra work SUS can provide not just the existing global score but also scores on two subscales; usability and learnability.

Sauro (2011, 86) provides the following rules to calculate scores for usability and learnability:

1. Start with scaling the scores the same way as with the regular SUS.
2. Learnability: total the scores for items 4 and 10 and multiply result by 12.5, which will scale the result from 0 to 100.
3. Usability: total scores for the rest eight items and multiply the result by 3.125 to scale the result from 0 to 100.

However, despite the fact they state these two factors can be detected from SUS they do not provide any practical ways to interpret the received scores.

5.3 Factors Affecting SUS Score

Naturally there are factors, which will have an impact on the result. According to Sauro (2011, 88), one of the most important is user experience – how much experience users have with the system being evaluated. One of the advantages of questionnaires like SUS are that you can compare very different types of systems with it – users will adjust their expectations of usability based on the context of use. However, it is not clear whether continued experience adjusts expectations and perceptions of usability more.

Therefore Sauro (2011, 88-91) performed a research to over 1100 SUS responses from 62 websites containing information about how many times the respondents had been to the site. He found out that those who had visited the website at least once or more gave 11% higher average SUS score than those who visited for the very first time. According to this research and conducting some more research to consumer software Sauro came to the conclusion that it is important to measure prior exposure to whatever is being measured. Also, it would be a good idea to report the difference between SUS scores for those using first time and repeating users. However, Sauro (2011, 93) states that while the experience matters, it explains less than 3% of the differences in the scores. It is more likely that differences in scores are attributable to actual perceived differences in usability.

Also the effect of age, gender and education on SUS scores has been researched. According to Sauro (2011, 91-92) as well as Bangor *et al.* they do not have a major impact on SUS scores.

5.4 Advantages of SUS

Jeff Sauro has analysed SUS a great deal and according to him, SUS is reliable and valid as well as comparable.

According to Sauro SUS is reliable, because it has shown to be more reliable and detect differences at smaller sample sizes than other, even commercial questionnaires. As sample size and reliability are unrelated, SUS can generate reliable results on a very small sample size. Validity, on the other hand, refers to how well something can measure what it is intended to measure. SUS has been shown to be effective on distinguishing unusable and usable systems from another, at least as well or even better than proprietary questionnaires. However, SUS was not meant to diagnose problems in usability. (Sauro, 2011.)

Another good feature about SUS is its free availability to be used as usability assessment tool. It has been used in many various research projects and industrial evaluations; the only requirement is that any published report should acknowledge the source of the measure. (Brooke, 1996, 194.)

6 VIDEO CONFERENCING SERVICE IN METSO

6.1 Video Conferencing Service

Metso started to utilize video conferencing service in April 2010 as an agreement was signed with a Finnish company called Videra. Metso did not want to invest in owning and maintaining the infrastructure, instead it was purchased as a service. Videra maintains all the core infra related to the service as well as is responsible for

delivering, installing and maintaining the video conferencing endpoints for Metso. Technology itself is provided by a US company called Vidyo.

6.2 Service Provider Videra

Videra is a Finnish company located in Oulu. Since 2010 Videra has been a part of Elisa Corporation, which is one of the leading producers of communication services in the Nordic countries. Videra is an independent subsidiary and is responsible for the visual communication solutions of the entire Elisa Group. (Videra homepages)

Videra has chosen its partners among the leading technology manufacturers in the market and it has not committed to using only the products of one manufacturer. The equipment manufacturers used by Videra include Polycom, Cisco/Tandberg and Vidyo. The manufacturer and the technology to be utilised is selected in a case-specific manner, taking the customer's needs into account. (Videra homepages)

In Metso's case Videra offered a technology solution based on Vidyo's technology, as it was cost effective but then again provided high quality even over internet.

6.3 Technology Provider Vidyo

Vidyo was established in 2005 in the USA. They have their headquarters in Hackensack, New Jersey. They are a privately held company employing over 150 people over the world. Their first product was launched 2008 and in October 2009 they were awarded a patent for their VidyoRouter™ architecture which delivers reliable, low latency, multipoint conferencing over any IP network including the Internet. Vidyo's product portfolio spans from VidyoMobile supporting tablets and

smart phones to laptops and desktops with VidyoDesktop to the VidyoRoom that encodes and decodes 720p and 1080p high definition (HD) quality video at up to 60 frames per second. (Vidyo Corporate overview)

Patented VidyoRouter™ architecture enables Vidyo's intelligent Adaptive Video Layering (AVL) technology. This AVL technology dynamically optimizes the video for each endpoint by leveraging H.264 Scalable Video Coding (SVC)-based compression technology and Vidyo's IP. This approach means costly hardware multipoint control units (MCU) are not needed but at the same time this technology offers error resiliency and low latency rate matching. Vidyo promises to provide and deliver high quality video over the Internet, LTE (long-term evolution), 3G and 4G networks. (Vidyo homepages)

As mentioned, AVL dynamically optimizes the video for each endpoint. During a video conference, Vidyo's core technology is monitoring the performance of the underlying network and the capabilities of each end-point device, and adapts video streams in real-time to optimize video communication. Video communications are dynamically layered into multiple resolutions, quality levels and bit rates. The overall result is error resiliency and natural HD quality video communications. Vidyo™ advertise themselves to be the provider of the first multi-point video conferencing solution delivering rate matching and continuous presence capabilities without additional video encoding and decoding. According to Vidyo this capability allows for less than half of the end-to-end latency of MCU-based solutions, which is crucial for a natural communication experience. (Vidyo homepages)

6.4 Video Conferencing Service Portfolio in Metso

From Vidyo's product portfolio Metso utilizes VidyoRoom as well as VidyoDesktop. VidyoMobile is also becoming more popular as tablets increase their popularity among users.

Videra provides Metso a standardized set of VidyoRoom product. This set is presented in Figure 5 and consists of:

- Two TV screens; one for sharing video stream (the images of the meeting participants) and the other for sharing presentation material during the meeting.
- Video codec with a remote control.
- HD camera.
- Audio devices (microphone and speaker).
- VGA cable (for plugging in to a laptop when sharing material).

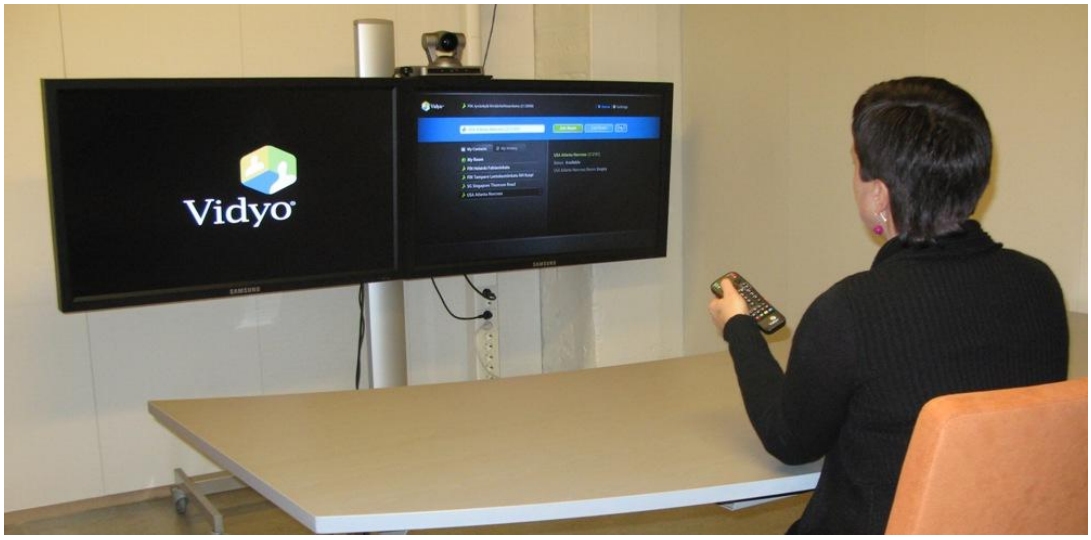


FIGURE 5. Standardized set of video conferencing devices in Metso

Screens can be either standing on a floor-stand (like in Figure 5) or they can be mounted on the wall. The screen size varies according to the size of the meeting room. Currently there are screens from 46” to 55” in use.

VidyoDesktop is a software client which enables having and joining video meetings from user’s own personal computer. VidyoMobile, on the other hand is a client to be installed to a mobile phone or a tablet. With these clients it is possible to join and have video meetings. However, in this thesis desktop or mobile clients are not included and their usability is excluded.

6.5 Video Conferencing Infrastructure

Metso has a closed, global corporate wide area network (WAN). Metso sites are connected to this corporate network either via Multiprotocol Label Switching (MPLS) connection or via LAN-to-LAN (Local Area Network) Virtual Private Network (VPN) connection. The capacity of these connections can vary, depending on the size of the site, from 512 kbit/s to something like 100 Mbit/s.

Due to adaptive video layering architecture VidyoRoom solutions do not require any dedicated data connections or Quality of Service (QoS) -definitions. For this reason, video meeting rooms can be situated in any Metso location where there is a connection to corporate network and enough free capacity. For example for HD-100 video codec Vidyo has stated that with minimum 1Mbit/s data connection transmit and receive resolutions will be HD 720p and frame rate 30 fps. The maximum data rates are for encoding 2 Mbit/s and decoding 4 Mbit/s. (VidyoRoom HD-100 datasheet, 2011.)

The infrastructure itself consists of VidyoRouters, VidyoPortal and VidyoGateway components. VidyoPortal and VidyoGateway are located in the service provider's network, from where the service is provided and maintained. VidyoRouters, on the other hand are physically located inside the corporate network but maintained by the service provider.

Currently the environment consists of over 100 meeting room solutions globally. Figure 6 illustrates how they are located globally around Metso.

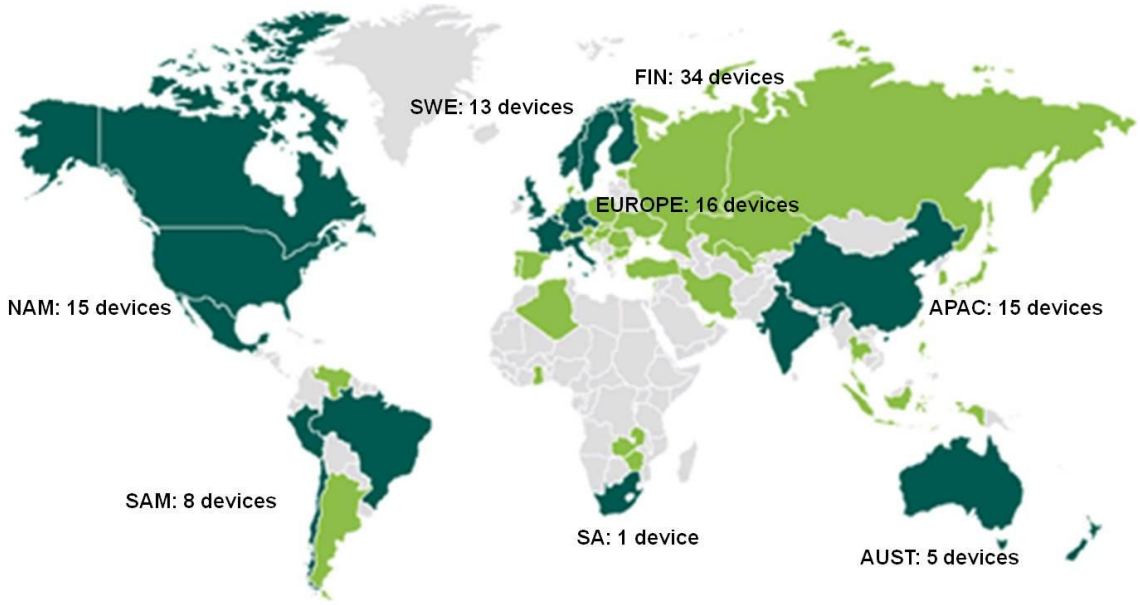


FIGURE 6. Installed video devices in Metso

6.6 Video Meeting Rooms

Every meeting room system, a set of devices, has been named according to an internal naming system. The name consists of country abbreviation, location city name and location street name. If the same location has several devices an additional explanation (E.g. meeting room name) has been added to the end of the name to separate the rooms.

Meeting rooms are listed in a directory, which is can be browsed from the user interface. It is possible to search for a meeting room by typing any part of the meeting room name to the search field. A list of suggested meeting rooms will appear on the screen as a user types in letters to the search field as illustrated in Figure 7.

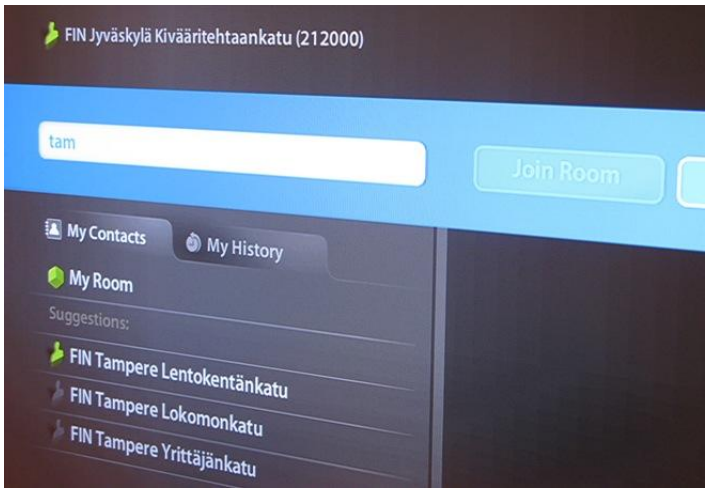


FIGURE 7. Searching a meeting room from the directory

6.7 Video Meeting Types

There are two types of video meetings:

- Point-to-point.
- Multipoint.

In a point-to-point meeting there are only two participants, two set of devices, joining the meeting. Point-to-point meeting is established when either of the participants calls the other one. This is like a phone call; one calls and the other one answers the call. No other participants can join or get invited to this meeting.

Multipoint meeting can contain two or more participants and it takes place in an agreed virtual meeting room. All participants join the agreed virtual meeting room in the agreed time. The amount of endpoints joining one multipoint meeting is currently limited to 20 but can be increased if necessary. However, when there are more than eight participants, the meeting is not as pleasant and easy to follow anymore, as the pictures showed on the screen start to change depending on who is talking. Figure 8 illustrates what a multipoint meeting with six participants looks like.

FIGURE 8. Multipoint meeting with Vidyo technology.

6.8 Video Meetings with Other Companies

Currently most of the video meetings held are purely Metso internal. However, it is possible to arrange video meetings with other companies. These meetings can be point-to-point or multipoint meetings like the internal ones.

Unfortunately arranging a video meeting with another company is not as easy as making a phone call with you mobile. There can be challenges to have the connection work, as companies have different kind of devices or have set up their environment in such a way that they do not allow video calls outside their own infrastructure.

Metso, together with the service provider, has released a set of instructions how to establish a video meeting with another company. If a meeting is not successfully established by Metso's own users according to instructions, then help from service provider is needed.

6.9 Using Video Conferencing Devices

As the user enters the meeting room and plans to have a video meeting, there are couple of things to check before a successful meeting can take place.

Video conferencing devices are normally powered on. Especially the video codec is instructed to be always powered on. This is because the service provider needs to access devices remotely and if they are powered off, remote access is not possible. However, in some cases the codec is powered off (for example a location suffers regular power cuts during nights) and the user has to power codec on before starting to use it. The screens are also normally always powered on – however, as big screens produce a great deal of heat, it is ok to shut them down, when they are not used. Audio devices should always be powered on and ready to use. However, in some cases users have turned them off or muted the device. Therefore it might be needed to power on / unmute audio device.

The system itself is used with one remote control. With the remote control users can

- Browse directory of meeting rooms.
- Search meeting rooms.
- Start and end a meeting.
- Control camera (pan, tilt and zoom).
- Control settings (for example volume settings and restart the system).

In Figure 9 Vidyo remote control is being presented.

FIGURE 9. Remote control for Vidyo video conferencing devices

During a meeting if any material is to be shared a VGA cable needs to be plugged in to a laptop.

So in a nutshell, when using video conferencing service a user has to be able to perform at least following actions:

- Turn on screens, video codec and audio device, if they are powered off.
- To be able to use the remote control in order to search the meeting room he needs to find and to establish the meeting.
- To be able to use remote control for controlling volume level and adjusting camera during the meeting.
- Share documents from his laptop using VGA cable.

6.10 Training and Instructions

After the devices are installed to a location the service provider provides training via video. These sessions are normally quite short, not more than 30 – 60 minutes.

During this training the basic functions are gone through. If necessary, the service provider will lay on more training sessions. However, it has been noticed not so many locations require additional training – whether this is due to the fact that use of these devices is considered rather easy and therefore training seems unnecessary or due to the fact that someone has used the devices before and will instruct others on how to use them.

Metso IT has produced some internal material on how to use the system. These instructions are available in the company intranet.

7 EVALUATING VIDEO CONFERENCING USABILITY IN METSO

7.1 Choosing Usability Evaluation Method

One of the main targets of this thesis was to find out if it would be possible to somehow evaluate the usability of current video conferencing service. Normally usability is evaluated when a product (for example a web page or a user interface to some product) is being developed, not that much when the products are actually already in production. It is very common to perform usability testing in the development phase to get information how people are using the system and what kind of problems they have.

However, I wanted to evaluate usability of a product which is in full production and very much used on a daily bases. Therefore I needed a method which would be fast to carry out, would not require setting up any separate testing sessions nor interviews and would produce a concrete result, so that perhaps in the future this procedure could be repeated and results compared.

Therefore I ended up to rejecting all other evaluation methods but questionnaires. Questionnaires are probably the only method, which can have such an extensive coverage. They can be easily distributed to a large group of users via e-mail or web. Questionnaires were considered to be an indirect method as they study a user's opinions about the studied target (for example user interface) that was exactly what I was aiming for.

My first thought was to create a questionnaire of my own but as I studied the subject I found out there are several existing usability evaluation questionnaires available. Therefore there was no sense to start to figure out questions on my own – why reinvent the wheel? I wanted to have a short and simple, yet reliable and valid, questionnaire as I knew end users would not be that anxious to reply, if the questionnaire even seemed long and time-consuming. After examining the options available I ended up choosing SUS. SUS was chosen to be the questionnaire used as it has proven to be an effective and reliable tool for measuring usability. It can be used with various products and services. It is short and therefore fast to implement. SUS would produce a concrete numeric value describing the usability of video conferencing service. However, I felt slightly insecure relying purely on SUS questionnaire and therefore I wanted to give end users also a possibility to give free comments about the service.

7.2 Conducting the Survey

Before sending the questionnaire to end users some original statements of SUS questionnaire were slightly adjusted, as suggested in some of the studies. Therefore the word cumbersome was changed into awkward in item 8. Instead, the word system was kept as it is and not changed to product. In this case it seemed more appropriate to keep it than change it.

SUS produces a single score result raising perhaps questions what it means to an absolute sense. As mentioned earlier, Bangor *et al.* (2009) introduced the possibility for adding an additional 11th question to the end containing adjective rating scale. This was to help interpreting the SUS score. This seemed a very good idea to carry out also in this study and therefore this 11th question was added and the respondents were asked to review the user-friendliness according to an adjective scale rate.

The respondents were also asked to give any free comments about the service if they wanted. This was just to make sure all possible feedback would be received now that end users were asked to give their opinions about the service usability. Appendix 2 presents the SUS questionnaire, additional questions and the cover letter sent to the respondents.

As mentioned earlier, SUS is normally performed to respondents after they have used the system being evaluated but have not had any orientation or instructions for using the system. In this case it was not possible to create this kind of situation and therefore respondents were selected from a database which contains all the bookings for video meetings. However, I tried to pick up respondents from such sites, which had recently received video conferencing devices. This way we could at least assume the respondents were not very experienced users.

The questionnaire was sent to 121 respondents on Thursday 24th of January 2013. They were asked to reply by Friday 8th of February 2013. The questionnaire was sent by e-mail, which contained a link to the web page where the questionnaire could be filled in. One reminder was sent on 4th of February, in order to make sure that as many as possible would answer the questionnaire. Of 121 respondents 66 replied and 55 chose not to, which lead to response percentage of 54.5%.

7.3 Scoring the SUS Items

Before getting the actual SUS score, responses needed to be processed according to a defined method. The received raw user responses range from 1 (Strongly disagree) to 5 (Strongly agree). First these raw SUS item responses should be converted like this:

- For odd items (1, 3, 5, 7, 9), 1 should be subtracted from the user response.
- For even items (2, 4, 6, 8, 10), subtract the user responses from 5.

This scales all the values to range from 0 to 4, with four being the most positive. After all the items are converted, responses from each user should be added up and multiplied with 2.5. In Table 1 this process is presented in a detailed level, with one respondent's responses.

TABLE 1. Example of scoring raw SUS items.

Question #	1	2	3	4	5	6	7	8	9	10
Raw item responses	5	1	4	1	4	1	3	1	5	1
Converted item responses	4	4	3	4	3	4	2	4	4	4
Sum of converted items	36									
Sum multiplied with 2,5	90									

To prevent mistakes from happening and to ensure faster and easier calculation, a SUS Excel calculator from Jeff Sauro was used. Responses were inserted to the calculator and it provided automatically a great amount of useful information.

First of all it was noticed that two respondents had not filled in all the answers. One response was missed two values and one response one item value. Sauro (2011, 24) suggests three different approaches for handling the situation, as it is not possible to leave the values simply blank because blank values would create an impossible SUS score due to the way SUS is scored. The first option to handle a missing value is that one could delete the whole SUS survey from that respondent, who has forgotten to answer all questions. This is perhaps the most objective way to handle the situation;

however, if the sample size is very small this could mean a significant loss of the data. The second option is to substitute the missing values. If only one value is missing, it could be reasonable to substitute it with neutral (3) response, although this might not be a fool-proof action either. Luckily, according to Sauro (2011, 24), SUS score will not be affected so dramatically regardless what response is inserted. The final approach would be changing the multiplier from 2.5 to another value to make sure that the scaled scores stay between 0 and 100. Sauro has implemented the third option (changing the multiplier) to his Excel calculator. This means that up to two missing values an updated SUS score will be provided – calculated with the changed multiplier.

If we use Jeff Sauro's Excel calculator and keep the changed multiplier for two responses we get the overall SUS score 67. Just out of curiosity, if those two responses are deleted from the results, SUS score remains the same. Excel calculator also measures internal reliability by Cronbach´s alpha, which in this case was 0.911. Values above 0.70 are considered to be good, values below 0.70 are poor and negative values are flagged as coding error. (Sauro, 2011, 18.)

7.4 Interpreting the SUS Result

So, now that we have received an overall scaled SUS score 67, what does this mean? How are we to interpret the result?

According to Sauro (2011, 28), the scaled score is best interpreted if compared to previous SUS scores, benchmarks from the industry or to the overall SUS average which is 68. As there are no previous SUS scores available it leaves us with two remaining options.

Received SUS score 67 can be compared to overall SUS average 68. This indicates that the according to users the overall usability of video conferencing devices is just

below the general average of 68. A good SUS score would be anything about a 76, which would mean it has a higher score than 75% of all products tested. (Sauro, 2011, 33.)

Comparing the received SUS score against benchmarks by interface type we can again use Sauro's (2001, 48) studies. He has generated a global benchmark for SUS combining data from three different datasets. Altogether these datasets contained 446 surveys with over 5000 individual SUS responses. The weighed mean from all three sources provide an average SUS score 68 with a standard deviation of 12.5. Then he created a summary table of benchmarks by interface type, which is presented in Table 2.

TABLE 2. Summary table of SUS scores by interface type (Sauro, 2011, 49).

	Mean	SD	N
Global	68,0	12,5	446
B2B	67,6	9,2	30
B2C	74,0	7,1	19
Web	67,0	13,4	174
Cell	64,7	9,8	20
HW	71,3	11,1	26
Internal SW	76,7	8,8	21
IVR	79,9	7,6	22
Web/IVR	59,2	5,5	4

Definitions to the benchmark sources are following:

Business to business (B2B) means enterprise software application such as accounting, customer relationship management (CRM) and order-management systems.

Business to consumer (B2C) is public-facing mass-market consumer software like office applications, graphics apps or personal finance software.

Web means public-facing large-scale websites (airlines, rental cars etc.) and intranets.

Cell stands for cell-phone equipment.

HW is hardware such as phones, modems and Ethernet cards.

Internal-SW (software) means internal-productivity software like customer service and network operations applications and most likely is having overlaps between the B2B and B2C groups.

IVR stands for interactive voice response systems (phone- and speech-based).

Web/IVR is a combination of web-based and interactive voice response systems.

In this research we could consider video conferencing service to be benchmarked against hardware, as the other options do not seem so appropriate. If we directly compare received result (SUS score 67) to the global mean score of hardware (71.3) we could say that the result is way below the average. However, Sauro (2001, 51) suggests to convert the received SUS score into a percentile rank with the help of a process calling standardizing or normalizing. To make it easier, he has added a tab to his calculation sheet, which will convert the score into percentile rank – which will then show directly, how usable the application or product is relative to other products.

The received SUS score (67) converted to percentile rank using Sauro's SUS calculator would be 34.6% - when selecting Hardware as benchmark. This can be seen in Figure 10.

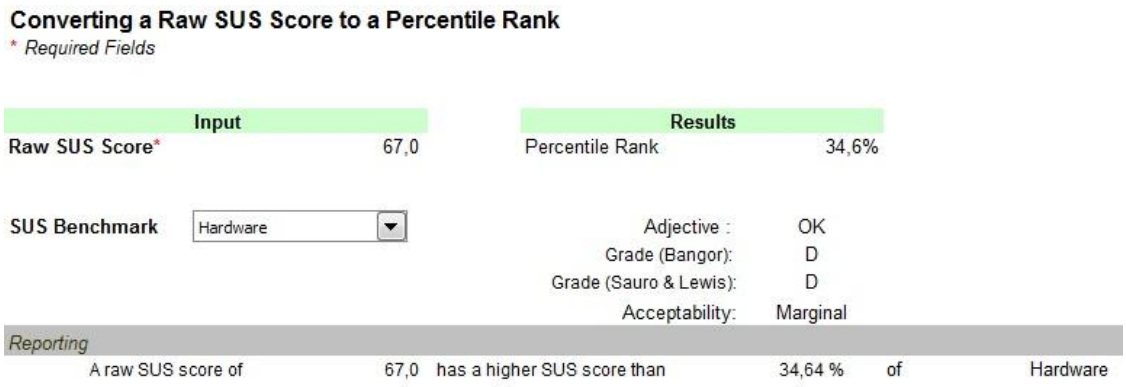


FIGURE 10. Converting SUS score to a percentile rank

As we can see, this SUS score of 67 for hardware would place it higher than only 34.6% of all hardware, meaning the perceived usability is way below average. Even if we compare it to all products, the percentile rank would be 46.9%, which is of course better than the value benchmarked against hardware; however, it is still below average.

7.5 Additional Adjective Scale

An additional eleventh question was added to the end of traditional SUS questionnaire. This question was added because Bangor *et al.* (2009) conducted a survey where they found that this adjective rating scale matches the SUS scale very closely and thus it could be considered as a useful tool in providing a subjective label for an individual study's mean SUS score. Therefore, out of interest, it was added to see how well it would match to this study.

In this eleventh question the respondents were simply asked to review the overall user-friendliness of this system with a seven-point, adjective-anchored Likert scale. This question is presented in Figure 11.

11. Overall, I would rate the user-friendliness of this product as:

Worst imaginable	Awful	Poor	OK	Good	Excellent	Best Imaginable

FIGURE 11. Eleventh question in the questionnaire.

When analyzing the responses, they were given in numeric values, 1 being worst imaginable and 7 best imaginable. All the respondents evaluated and replied to this question and the average was 4.79 – meaning OK as adjectively.

Bangor *et al.* (2009) have also studied and presented different ways to interpret SUS score by converting it into a grade or comparing it to a set of acceptability ranges. They presented this following Figure 12, which illustrates how SUS scores match with grades, adjectives or acceptability ranges.

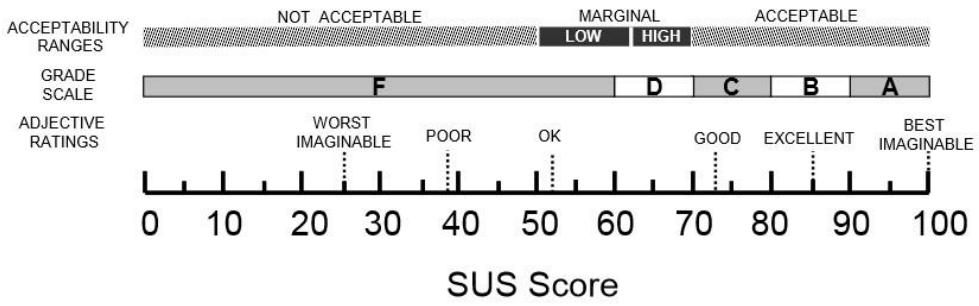


FIGURE 12. A comparison of the adjective ratings, acceptability scores and school grading scales, in relation to the average SUS score. (Bangor *et al*, 2009, 121.)

When comparing the received SUS score of 67 to adjective ratings, we can see the result is OK, rather close to good, but still below. The mean (4.79) calculated from the responses to eleventh question also supports this result. The school grade according to Bangor *et al.* would be D and the acceptance level is marginal.

All these adjective ratings, grades and acceptance levels are just another way to interpret the received SUS score and present the received result in a more understandable way compared to just a numeric value.

7.6 Feedback about the Service

Respondents were also given the possibility to give overall feedback in free form. Of the total 55 respondents 35 gave feedback. All responses are presented in appendix 3.

It was mentioned that the system is very good, very much used and saves plenty of money, because travelling is not needed. One respondent even referred the system as “a lifesaver”. There was also a respondent who referred current video conferencing system “works better than expected” and how previous video system was “too difficult to use”.

However, there were some development topics and feedback about things which would need improvement. The main topics mentioned are:

- Feedback about sharing data and presentations. It was mentioned that shared data updates slowly on the screen and is sometimes not so sharp. It is impossible to share videos via data sharing. Also some respondents hoped for interactivity for data sharing (for example other end could point out things from the presentation other end is sharing).
- Training and better instructions are needed; respondents reported they often struggle using the devices.
- Audio quality was mentioned to be weak.
- Remote control was mentioned to be difficult to use. Wireless keyboard was suggested to help the usage.

- Seems that when the devices work, people are happy but when an error occurs, help from IT is needed. Problem solving for normal end user is not that easy.
- Video meeting rooms seem to be very much utilized; there should be more rooms available.
- Picture freezes or lip sync is behind, trouble caused by network connections and delay.
- Video meetings with external partners and companies should be easy to establish and training should be offered on how to establish them.

There were also responses where it was obvious that respondents were simply not aware of how to perform certain available actions, like how to book several meeting rooms for your meeting, how to change shared material or how to establish a video meeting with external parties. These should be instructed better and more information distributed to the end users.

Some of the responses contained comments where more info would be nice to have. For example, one of the respondents claimed to have experienced "sudden software updates" in the middle meeting which seems very odd as that should never happen and no-one has reported anything like that before. Also, it would be interesting to have a talk with the respondent who replied that "Technology is somewhat archaic compared to modern day systems with better resolution, less lag, better presented material integration, etc."

7.7 Usability Evaluation Results in a Nutshell

The usability of video conferencing service in Metso was evaluated with SUS. As a result it produced a single numeric score of 67. If compared to overall SUS average of 68 we can see it is slightly below average. As there are no previous SUS scores available in Metso we cannot compare the results to that. It is also suggested to

compare the result to benchmarks. That tells the same story, usability is below average.

An additional 11th question was added to the end of the original SUS questionnaire. In this question respondents were asked to review the overall user-friendliness of the system with a seven-point adjective-anchored Likert scale. As the adjectives were given numeric values (1 being worst imaginable and 7 best imaginable) the average of all responses turned out to be 4.79 – meaning OK as adjectively.

More than half of the respondents gave overall feedback. There were many positive comments but also some very good improvement ideas and feedback how the service should be improved. It was definitely worth a while to ask for overall comments in free form.

8 CONCLUSION

Metso has used video conferencing for almost three years now. It is widely used and the personnel as users seem to be satisfied with it. At least that is the general impression; however, every now and then feedback is received how difficult it is to use the system, how for example Polycom devices are easier to use. Therefore I started to wonder if there is a way to find out or measure the level of usability in video conferencing in Metso. Would it be possible to show that the devices are actually not that usable or is this something related to lack of training or perhaps just dissatisfaction with the service in general?

I started to read material about usability and usability testing. I soon found out how usability testing with real users is the most fundamental usability method and it sounded very interesting and something I wanted to perform. As Nielsen stated (1993), testing usability with real users provides direct information on how people

use the system and what problems they might have. When I looked at Wikipedia, it states the following about usability testing:

> *" Simply gathering opinions on an object or document is market research or qualitative research rather than usability testing. Usability testing usually involves systematic observation under controlled conditions to determine how well people can use the product." (Usability testing, Wikipedia)*

So, in order for this thesis to be a proper usability testing study it would have required to set up sessions with end-users trying to use video-conferencing for the very first time, ask them to perform a set of pre-defined tasks and have them fill in questionnaires based on their experiences. That was definitely out of the question due to time and resources, no matter how interesting it could have been.

I had to find another way to evaluate usability in video conferencing service. Due to the fact that video conferencing service in Metso is spread globally, there was not much time or resources; I had to rule out methods like interviews, heuristic evaluation, observation and focus groups. I ended up choosing questionnaires, as they are perhaps the only method with which you can reach a large group of users easily, for example using e-mail.

First I thought I would create a questionnaire of my own. However, I started to think over as I studied the subject more and found out there were questionnaires available and ready to be used. Why would I invent questionnaire of my own, if there were options already available for me to choose? That is when I ended up choosing SUS, System Usability Scale. It seemed a perfect choice for my study: it was free, short, simple and quick to perform, is not technology dependent and has references in hundreds of publications.

I ended up sending the SUS questionnaire to 121 respondents. However, I was a bit suspicious relying purely on SUS, as the interpretation of the SUS score seemed a bit challenging according to some authors. Therefore I added an additional 11th question to the questionnaire, asking users to evaluate the user-friendliness of the system with an adjective rating scale. Respondents were also given the opportunity to give feedback in free form, if they wanted.

The questionnaire was sent to randomly selected end users – however, I tried to select users from such sites which only recently had their video conferencing devices installed and therefore one might suspect their level of experience is not yet so high. I was hoping for a high response rate as the questionnaire was short but to my surprise only 66 replied and 55 chose not.

When analyzing the results I had great help from Jeff Sauro´s material about SUS. He has even created an Excel calculator, which helped greatly and saved a lot of valuable time. As I had the questionnaire responses and analyzed them I had the final result in my hands – the measured result of usability in video conferencing service in Metso has the SUS score of 67.

SUS score as a numeric value did not provide much valuable information as such about the usability of video conferencing in Metso. The score turned out to be slightly less than average of 68, giving an adjective value of OK. One would have suspected the score to be higher, as end users were a bit more experienced perhaps than in cases, where SUS normally is performed. Maybe the selected end users were not that experienced after all and the score is somewhat comparable to a situation where users without experience try to use the system.

However, when working with engineers it feels good to have something concrete and measured to present as a result – a numeric value, which could be followed on a regular basis if necessary. Maybe if more training would be provided and after that the same questionnaire were to be conducted again, we would see improvement on the overall score – but on the other hand, would that be misinterpreting the result, as usability as such has not improved, only end-users are better trained and experienced and feel that devices are easier to use.

Perhaps the most useful information in this study was the voluntary feedback from the respondents. According to the feedback concrete actions can be defined to improve the video conferencing service level in Metso. Of course there were topics I knew beforehand people were not satisfied with, such as weak audio quality, lack of training and better instructions and the fact that video meetings with customers and partners should be easier to establish. These are the topics we have already been

working with to improve the current situation. To my surprise some new topics were also brought to my attention like how the use of remote can be difficult and how some people wished for interactivity to data sharing. These development ideas will be passed to Vidyo and hopefully they will consider implementing them in the future. Feedback from end users also revealed there is a need for informing more about available features, such as how to book meeting rooms or establish a meeting with external parties. More training and better instructions are clearly needed and wanted. This end user feedback was very useful and therefore it will be analyzed carefully and actions will proceed accordingly.

Normally usability tests are performed by the company developing the application or product. So was there any point of doing this, as this was not performed by Vidyo, technology provider developing the video conferencing devices. Most definitely Vidyo has used usability testing when developing the user interface for their video conferencing devices; however, perhaps this research can bring them some new information too as this is feedback from real users, really trying to use this equipment in their daily work.

One might also consider, what the point of conducting this research was as there were no previous SUS scores for comparing the received result. Now that the first SUS score is available, it would be possible to perform a new research after a while and see whether we see any improvement on the score, if for example some major user interface improvements are performed by Vidyo. Jeff Sauro also suggested comparing the received SUS score to benchmarks by interface type which he had created by combining data from several SUS studies. I compared video conferencing to hardware, as other options did not seem suitable. The result was less favorable than when compared to overall SUS average. However, I would not be too concerned about the result, as comparing video conferencing service usability to hardware usability does not quite seem the best option.

SUS as a tool for evaluating usability is good. It is short, containing only ten questions. Compared to other questionnaires containing much more questions, this is clearly an advantage, it is quick to do and rather easy to administer. SUS has turned

out to be reliable with smaller sample sizes compared to other questionnaires and it is a valid method, as it has been effectively shown to distinguish between usable and unusable systems. SUS is not technology dependent and can be used with websites as well as hardware. It is also a free tool, and therefore has been used and referred to in many publications. However, interpreting SUS scores can be challenging, as the score – being a numeric value – does not provide that much information as such. In order to be able to interpret the score, one should have previous scores available for comparison, compare the score to an overall average value 68 or compare the score with industry benchmarks. Also, some ways to interpret your score with grades and adjectives have been developed, which perhaps makes it easier to tell people what the result means.

One might question the fact that Jeff Sauro seems to be one of the few people who has studied SUS and its use. When I searched information about SUS his name was mentioned in most of the cases. I would have expected to find more material from other authors as well. So, for example is Sauro´s material for comparing the benchmarks comprehensive enough? It seems so but still I wonder why there are not that many other scientific researches about this matter, or perhaps I just did not come across to them.

All in all, trying to evaluate usability in video conferencing service was interesting and educational. Was it useful, I would have to answer yes and no. Some could say this study was an abuse of usability evaluation as it tried to perform usability evaluation on a product fully in use with end users who had been using the product for a while. But on the other hand – is that not usability on its best? People trying to get things done, trying to achieve their goals at work – why should we not study how they succeed in it? Some might say you do not need or should not use usability evaluation methods for that, however, why not cross some boundaries once in a while? From Metso´s point of view it might have been even more useful if a “home-made” questionnaire instead of SUS was used – perhaps it would have indicated more clearly how end users’ experience the usability in video conferencing service and what are the actions needed to improve that experience. In a nutshell, this study pointed out the usability level in video conferencing service is OK and acceptable, as

assumed; however, there are areas where development actions could be implemented.

REFERENCES

Bangor, A., Kortum, P. & Miller, J. 2008. An Empirical Evaluation of the System Usability Scale, International Journal of Human-Computer Interaction, Vol. 24, Issue 6, July 2008, pp. 574-594. Referred 3.2.2013. Http://www.jamk.fi/kirjasto, Nelli-portal, EBSCOhost Academic Search Elite

Bangor, A., Kortum, P. & Miller, J. 2009. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale, Journal of Usability studies, Vol. 4, Issue 3, May 2009, pp. 114-123. Referred 3.2.2013. Http://www.upassoc.org/upa_publications/jus/2009may/JUS_Bangor_May2009.pdf.

Brooke, J. 1996. SUS: A Quick and Dirty Usability Scale. In Usability Evaluation in Industry. Ed by. P.W. Jordan, B. Thomas, B.A. Weerdmeester & I.L. McClelland. London.Taylor & Francis.

Faulkner, X. 2000. Usability engineering. Basingstoke. Palgrave.

Kuutti, W. 2003. Käytettävyys, suunnittelu ja arviointi. Helsinki. Talentum.

Lewis, J. 1993. IBM Computer Usability Satisfaction Questionnaires: Psychometric Evaluation and Instructions for Use. Technical Report 54.786. Referred 3.2.2013. Http://drjim.0catch.com/usabqtr.pdf .

Lewis, J. & Sauro, J. 2009. The Factor Structure of the System Usability Scale. Published in HCD 09 Proceedings of the 1st International Conference on Human Centered Design: Held as Part of HCI International 2009. Referred 3.2.2013. Http://gate.ac.uk/sale/dd/statistics/Lewis_Sauro_HCII2009_SUS.pdf.

Nielsen, J. 1993. Usability engineering. San Diego. Academic Press, Inc.

Polycom Fact Sheet: The Top Five Benefits of Video Conferencing, 2010. Referred 8.1.2012.Http://www.polycom.com/global/documents/products/resources/video_education_center/top_benefits_of_video_conferencing.pdf.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. & Carey, T. 1994. Human-computer interaction. Harlow. Addison-Wesley.

Rubin, J. & Chisnell, D. 2008. Handbook of usability testing: how to plan, design, and conduct effective tests. JAMK Ebrary eBook Collection. Referred 15.3.2012. Http://site.ebrary.com.ezproxy.jamk.fi:2048/lib/jypoly/Doc?id=10232880.

Sauro, J. 2011. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Purchased and downloaded from http://www.measuringusability.com/.

Sauro, J. 2011. Measuring Usability with the System Usability Scale (SUS). Webpages about SUS. Referred 3.2.2013. Http://www.measuringusability.com/.

SFS-EN ISO 9241-11.1998. Ergonomic requirements for office work with visual display

terminals (VDTs) – Part 11: Guidance on usability. Helsinki: Finnish Standards Association. Referred 19.3.2013. Http://www.jamk.fi/kirjasto, Nelli-portal, SFS Online.

Sinkkonen, I., Kuoppala, H., Parkkinen, J. & Vastamäki, R. 2006. Psychology of usability. Edita Publishing Oy.

SUMI Questionnaire homepage. Referred 3.2.2013. Http://sumi.ucc.ie/.

Tullis, T. & Stetson, J. 2004. A Comparison of Questionnaires for Assessing Website Usability. UPA 2004 Presentation. Referred 2.2.2013. Http://home.comcast.net/~tomtullis/publications/UPA2004TullisStetson.pdf.

Usability testing. Wikipedia. Last modified 19.3.2013. Referred 30.3.2013. Http://en.wikipedia.org/wiki/Usability_testing.

Videra homepages. Referred 15.3.2012. Http://www.videra.com.

Vidyo Corporate overview. Referred 15.3.2012. PDF document in http://www.vidyo.com/documents/VidyoCorporateBackgrounder.pdf.

Vidyo homepages. Referred 15.3.2012. Http://www.vidyo.com.

VidyoRoom HD-100 datasheet. 2011. Referred 15.3.2012. Http://www.vidyo.com/documents/datasheets-brochures/VidyoRoomHD-100_DS_US.pdf.

Questionnaire for User Interaction Satisfaction (QUIS). Web pages about QUIS. University of Maryland. Referred 3.2.2013. Http://www.lap.umd.edu/quis/.

Wiio, A. 2004. Käyttäjäystävällisen sovelluksen suunnittelu. 2004. Edita Publishing Oy.

APPENDICES

Appendix 1. Usability methods according to Nielsen (1993, 223).

Method Name	Lifecycle Stage	Users Needed	Main Advantage	Main Disadvantage
Heuristic evaluation	Early design, "inner cycle" of iterative design	None	Finds individual usability problems. Can address expert user issues.	Does not involve real users, so does not find "surprises" relating to their needs.
Performance measures	Competitive analysis, final testing	At least 10	Hard numbers.Results easy to compare.	Does not find individual usability problems.
Thinking aloud	Iterative design, formative evaluation	between 3-5	Pinpoints the user misconceptions. Cheap test.	Unnatural for users. Hard for expert users to verbalize.
Observation	Task analysis, follow-up studies	3 or more	Ecological validity; reveals users' real tasks. Suggests functions and features.	Appointments hard to set up. No experimenter control.
Questionnaires	Task analysis, follow-up studies	At least 30	Finds subjective user preferences. Easy to repeat.	Pilot work needed (to prevent misunderstandings).
Interviews	Task analysis	5	Flexilble, in-depth attitude and experience probing.	Time consuming. Hard to analyze and compare.
Focus groups	Task analysis, user involvement	6-9 per group	Spontaneous reactions and group dynamics.	Hard to analyze. Low validity.
Logging actual use	Final testing, follow-up studies	At least 20	Finds highly used (or unused) features. Can run continuosly.	Analysis programs needed for huge mass of data. Violation of users' privacy.
User feedback	Follow-up studies	Hundreds	Tracks changes in user requirements and views.	Special organization needed to handle replies.

Appendix 2. Cover letter, SUS questionnaire and questions sent to respondents.

Subject of the mail: Please give your opinion on using video conferencing room system

Body of the email:

Hello,

Please find enclosed a link to a questionnaire concerning usability of video conferencing room system (Vidyo).

This questionnaire is a part of my Master's thesis. It contains only 11 short questions, so it will not take long of your time. I would also appreciate your free comments how you feel about using video conferencing room system overall (what is difficult, should there be more training etc).

Please click the link enclosed and the questionnaire will open to your browser. I am hoping to get your answers by Fri, 8th of February. If you have any questions about this questionnaire, please don't hesitate to contact me.

Your answers will be highly valued

Best Regards,

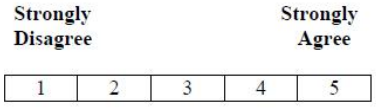
Mia Suominen

Service Delivery Manager, UCC

Metso IT

Link to the questionnaire Questionnaire Concerning Usability of Managed Video Conferencing Room System (Vidyo)

SUS Questionnaire with answering options



1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.

8. I found the system very awkward to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

11. Overall, I would rate the user-friendliness of this product as (Answer had to be chosen from following predefined options: Worst imaginable, Awful, Poor, OK, Good Excellent, Best imaginable)
12. Please feel free to give any comments on how you feel about using video conferencing system overall

Appendix 3. Received feedback about video conferencing service.

1	At the moment the video conference connection is the best communication channel over long distances e.g. to India, US, China, Brazil. There is still some problems to be solved: technical support to reconnect the unpluged cables and to restore muted speaker (hardware) etc. needed once in a while, video image gets frozen and shared presentation has huge delay too often.
2	I don't use it often enough. So when I need to use it, I struggle. I can fulfill most of my needs with Interwise.
3	This works much better than I expected. Earlier vidoe systems were too difficult to use. A lot of problems to connection working in China now. Normally we loose first 15 minutes with connecting promlems and they need every time technical help. Instructions to change presentation are not good. Normally they dont work. Best way is to unplug cable. Presentation screen updates slowly and we have to be slow changing pictures. Videos are impossible. I use this system several time each week.
4	When reserving the video conference equipment in Lotus Notes, it should automatically reserve the corresponding conference rooms.
5	User functionality is poor and we need a VC system that allows for video conferences with External Parties, such as suppliers, customers, etc.
6	It is hard to find the room to join. Sudden software updates are very bad during meetings. Still difficult to connect Metso partners and clients.
7	I use it a lot and I found it to be an excellent tool. Much better sound quality compared with the other systems we use. I use the system also on the ipad, and I would like to see that we can use it outside the Metso VPN. That is really the only extra thing I would need.
8	To use the system in out of company connections should be available (maybe it is?) and training for that arranged.
9	I usually use Vidyo for worldwide conference meeting, sometimes with 7 different locations. Main argue is we save a lot of flying ticket to set these meetings and we can absorb difference in time by selecting the correct hour suitable to every participant.
10	When it works it is perfect but the availability could be improved. It happens several times every month that there is some errors that needs attention from IT specialist. We use it a lot.
11	Have not been using it too many times (yet) but found out that the meeting rooms equipped with this kind of equipment are extremely popular. This is a sign of acceptance and that the organisation is finding the system good, functioning and time and cost saving.
12	Writing tool (remote-control) is not practical, system should be equipped with wireless keyboard. This is very good system and it save my time a lot. We need more video-conference rooms, sometimes it is very challenging to find video-room, especially when many location are involved to the same meeting.
13	Still too many times some participating location have problems when scheduled meetings (user of technology?). Microphones could be better. When "long" narrow room and only one microphone, it is still too hard to hear all participants
14	Incorporation of computer presented materials have severe lag making that portion unusable. System does not seem to be as seamless / well integrated as others. Technology is somewhat archaic compared to modern day systems with better resolution, less lag, better presented material integration, etc.
15	I think the new system will prove to be efficient once the users learn how to use all the functions in it. However, it would be practical if it would be possible to reserve more than one video room at the same time.

16 If you don't use the system often, you forget how to use all the functions. It would be quite helpful if there was a "quick tips" sheet in the video conference room with easy step by step instructions available. I find I have to get the IT dept involved 1/2 the time to assist at setup because something isn't functioning correctly, which is usually "user error". There is also a slight delay in communication back and forth but I guess that is to be expected. Overall, it works pretty good when a meeting is required and you don't want or need to travel for it.

17 Have had some problems with the connection (freezing) and especially with voice (some fault was found in the microphone - will be fixed). If many persons participate in a meeting the microphone loudness could be better.

18 Booking of the room needs to be confirmed by Assistant. It taking to long time and the prioritized is don't known. Also not always is given the information then the room is rebooked on somebody else. One Video room is not enough for our plant.

19 The concept is fantastic and its a very valuable tool, but additional training would be helpful. Our support person is difficult to get a hold of, so its tough to get answers sometimes when there are problems.

20 Very good system. extremely usefull to save traveling $$$.

21 After application is installed and all set up, the usability is very good. Presentations are sometimes little fuzzy, but the overall user experience is still a lot better than e.g. with Interwise.

22 System basically works OK, but requires maintenance/trouble shooting too often. Very useful tool in communication in big organization like Metso.

23 Miksiköhän tämäkin kysely on vain eglanniksi??? Ksymyksissä on niin hienoja sanoja että saa MOT:n kanssa selvittää että mitä kysytään! Itse video neuvottelu järjestelmä toimii kohtuudella. Suurin ongelma on heikko äänenlaatu josta on vaikea saada selvää. Neuvottelu- huoneet ovat aivan liian kaikuisia ja kaikki hälyäänet tulee lävitse. Hieman auttaa jos mikk&kaiutin paketin saa siirrettyä lähemmäs puhujia, yleensä ei kuitenkaan saa kun niissä on niin lyhyet piuhat. Toinen parannus olisi jos vastaanottajakin pystyisi näyttämään vaikka kursolla kohtia näytettävästä materiaalista. Nythän tämä on mahdollista vain esittäjälle. Tämä on varmaankin vaikea totetuttaa ohjelmaan.

24 The use is not problemous, the annoying part was the booking of the premises... (that has changed since then, but could be quite lean... On the other hand, if the purpose of the booking system is to keep the usage as low as possible, it's doing a great job :-)

25 Sharing the materials should be improved, including editing on-line

26 Overall it's not a complicated system, however, the navigation to select the video conference rooms is done through the remote control which isn't easy to use when you need to constantly type in the name of the conference room and a great improvement would be to have a wireless USB keyboard if possible.

27 Overseeing the technical disturbances i.e slow net speed (resulting in slow movement of image compared to voice speed) the system is very handy to avoid travels and save time and other resources.

28 The power buttons of the monitors should have been marked more clear that they are under the screen, not in the lower part of screen.

29 I have never used the video conference system nor has any upper management at my location offered training on how to use. I believe if given the opportunity for training with video conferencing system I could use it easily.

30 We need more video meeting rooms!!! The hardware (software (don't know which) used here needs to be upgraded. Material transffered from the computer to the big screen (and to all remote screen) is unsharp and updated too slowly. Othervise the system is a lifesaver :-)

31 Establishing the connection was hard. I was finally able to do it by using the name of the meeting room in Brazil. Maybe that should be instructed. Currently instruction advices to use name of the users.

32	I only think that people that are not so familiar with IT stuff might get problems only when issues occur. In standard use the system is intuitive and easy to use.
33	The only problem I see is, we cannot connect this system to other systems e.g. customer systems. Makes work more difficult than necessary.
34	really .....s.....l .....o.....w.....ly
35	No comments. It is very good.