FOI, Swedish Defence Research Agency, is a mainly assignment-funded agency under the Ministry of Defence. The core activities are research, method and technology development, as well as studies conducted in the interests of Swedish defence and the safety and security of society. The organisation employs approximately 1000 personnel of whom about 800 are scientists. This makes FOI Sweden's largest research institute. FOI gives its customers access to leading-edge expertise in a large number of fields such as security policy studies, defence and security related analyses, the assessment of various types of threat, systems for control and management of crises, protection against and management of hazardous substances, IT security and the potential offered by new sensors.

FOI
Defence Research Agency
Information Systems
SE-164 90 Stockholm

Phone: +46 8 555 030 00
Fax: +46 8 555 031 00

www.foi.se

Brain Budget

Evaluation of Human Machine Interaction in System Development for High Risk and Task Critical Environments

HANS JANDER, JONATHAN BORGVALL, MARTIN CASTOR

FOI-R--3272--SE User report
ISSN 1650-1942 September 2011

Information Systems

Hans Jander, Jonathan Borgvall, Martin Castor

Brain Budget

Evaluation of Human Machine Interaction in System Development for High Risk and Task Critical Environments

Titel	Hjärnbudget – Utvärdering av Människa Maskin Interaction i Systemutveckling i Högrisk och Uppdragskritiska Miljöer
Title	Brain Budget - Evaluation of Human Machine Interaction in System Development for High Risk and Task Critical Environments
Rapportnr/Report no	FOI-R--3272--SE
Rapporttyp/ Report Type	Användarrapport/ User report
Sidor/Pages	40 p
Månad/Month	September/September
Utgivningsår/Year	2011
ISSN	ISSN 1650-1942
Kund/Customer	VINNOVA
Projektnr/Project no	B55026
Godkänd av/Approved by	Jan Svetoft Frisén

FOI, Totalförsvarets Forskningsinstitut	FOI, Swedish Defence Research Agency
Avdelningen för Informationssystem	Information Systems Division
164 90 Stockholm	SE-164 90 Stockholm

Sammanfattning

Denna rapport beskriver arbetspaket 1 och 2 i projektet Hjärnbudget. Projektet är finansierat av VINNOVA (Verket för Innovationssystem), inom ramen för det Nationella Flygtekniska Forskningsprogrammet (NFFP) Det finns totalt sex arbetspaket. Projektet är ett samarbete mellan Saab Aeronautics, Totalförsvarets Forskningsinstitut och Stockholms universitet.

Projektet fokuserar på innovativ och effektiv systemutveckling för militära flygplan inom området Människa Maskin Interaktion (MMI).

Det övergripande målet för projektet Hjärnbudget är att utveckla kvantitativa och kvalitativa MMI-mått och ta fram metoder för att kunna utvärdera pilotens MMI i cockpit för att stödja arbetet med kravanalyser, kravspecifikationer och kravstängning.

Första delen i rapporten (arbetspaket 1) beskriver en behovsanalys som genomfördes i syfte att beskriva förutsättningar och behov från både pilotens (slutanvändarens) men även från systemutvecklarens sida inom systemutvecklingen. Resultatet visar på en samstämmig bild av hur MMI-aspekter adresseras idag inom systemutvecklingen. Idag finns det en brist av systematiskt återkommande MMI-utvärderingar och tester under systemutvecklingen. Det finns många relaterade orsaker till detta och dessa orsaker behöver tas hänsyn till för att kunna integrera MMI-utvärderingar i systemutvecklingen. Det är tre huvudsakliga aspekter att ta hänsyn till: *Organisatoriska aspekter; Metodologiska aspekter;* och *Tekniska aspekter*.

Andra delen i rapporten (arbetspaket 2) ger en överblick över ”state of the art” av MMI-utvärderingsmetoder inom systemutveckling. Potentiella kriterier att mäta under utvärdering och testning identifieras och redovisas. Baserat på resultaten från litteraturgenomgången (arbetspaket 2) och behovsanalysen (arbetspaket 1) föreslås en metodologisk ansats (MMI-bedömningsverktygslåda) för hur MMI-utvärdering skulle kunna genomföras under systemutveckling inom domänen.

Nyckel ord: Människa Maskin Interaktion, Människa Dator Interaktion, Människa Maskin Gränssnitt, Människa System Interaktion, Männniska System Integration, Användbarhet, Användbarhetsmetoder, Mental Arbetsbelastning, Användarcentrerad Design, System utveckling, Human Factors, MMI-utvärdering, MSI-utvärdering

Summary

This report describes the Work Packages 1 and 2 of the Brain Budget project. The project is sponsored by VINNOVA (Swedish Governmental Agency for Innovation Systems), within the National Aviation Engineering Program (NFFP). There are six Work Packages in total. The project is a joint effort between Saab Aeronautics, Swedish Defence Research Agency (FOI), and Stockholm University.

The project focuses in innovative and effective system development for military aircraft in the area of Human Machine Interaction (HMI).

The overall objective of the Brain Budget project is to develop quantitative and qualitative Human Machine Interaction (HMI) measures in order to assess the pilot's cockpit HMI to support the establishment of requirement analyses, specification, and closure.

The first part in the report (Work Package 1) describes a needs analysis that was conducted in order to describe conditions, prerequisites, and needs from both the pilot (end-user) and system developer perspectives of system development. The result gives a concurrent view of how HMI aspects are addressed today. There is a lack of systematic recurring HMI-evaluations and testing during system development. Several interrelated issues are the reason for this situation and this needs to be further considered during system development and design concerning the incorporation of HMI evaluation methods. There are three major aspects that need to be considered: *Organisational aspects; Methodological aspects; and Technical aspects.* The Brain Budget project will only focus on the methodological aspects.

The second part in the report (Work Package 2) provides an overview of "state of the art" of HMI evaluation methods used in system development. Potential HMI criteria to measure during evaluation and testing are also defined. Based on the result from the literature review (WP2) and the findings of the needs analysis (WP1), a methodological approach (HMI Assessment Tool Box) on how to conduct HMI evaluation during system development is proposed.

Keywords: Human Machine Interaction, Human Computer Interaction, Human Machine Interface, Human System Interaction, Human System Integration, Usability, Usability Methods, Mental Workload, User-Centred System Design, System Development, Human Factors, HMI Assessment

Table of contents

1 Introduction

This report describes work packages (WP) 1 and 2 of the Brain Budget project. The project is sponsored by VINNOVA, Swedish Governmental Agency for Innovation Systems, within the National Aviation Engineering Program (NFFP). There are six Work Packages in total.

The project focuses on innovative and effective system development for military aircraft in the area of Human Machine Interaction (HMI). The project is based on lessons learned from the Swedish Air Force Combat Simulation Centre (FLSC) and the Unit for Man-System Interaction (MSI) at the Swedish Defence Research Agency with regard to system evaluation techniques that can be tailored and applied to fit system development methodology. The project will use established methodology and develop agile methodology to enhance HMI evaluation in system development. The project is a joint effort between Swedish Defence Research Agency (FOI), Saab Aeronautics, and Stockholm University.

2 Objectives

Mission: The overall objective of the Brain Budget project is to develop quantitative and qualitative Human Machine Interaction (HMI) measures in order to assess the pilot-system performance (the interactive performance) to support the establishment of requirement analyses, specification, and closure.

Vision: The design solutions will allocate parts of an available so called Brain Budget. The exact components of this Brain Budget are an area of investigation for this project but the notion is the answer to the following question:

How can it be established that a specific design solution is sufficient according to HMI and operational criteria without exceeding the Brain Budget capacity?

Work Packages 1 and 2 have been conducted in parallel and iteratively, since the two packages overlap and interact to a great extent. According to User Centered System Development (UCSD), system development should be performed in an iterative manner to support the identification of challenges and problems, and to address them, earlier in the process.

2.1 Objectives Work Package 1 (WP1)

The objective of the 1st WP is to conduct a needs analysis and to describe conditions and prerequisites from both the pilot and system developer perspectives. In addition, organizational needs and conditions must be analyzed and addressed in order to ensure capturing the right context. Definitions of methodology evaluation criteria are also to be defined.

2.2 Objectives Work Package 2 (WP2)

The objective of the 2nd WP is to undertake a “state of the art” literature review regarding methodologies and techniques for HMI evaluation (verification and validation). Based on the result of the literature study and the result from WP1, WP2 is to present a methodological approach for HMI-evaluation that will feed into the following Work Packages.

2.3 Constraints

The work of the project will not directly generate any specific design solutions and is focusing on the methodology aspects of the HMI evaluation.

3 Method

The methods used in WP1 and WP2 are divided in two areas. The first consists of interviews and a workshop with the purpose of conducting the needs analysis and to verify identified HMI criteria. The second area is the literature review that investigates the relevant methodologies and techniques most suitable to apply when performing HMI evaluations in system development. Though the work of WP1 and WP2 has been carried out in parallel, information gathered in WP1 has been used as input to WP2 and vice versa. The utilization of these data collection methods will increase the applicability of the results for the intended context of use, both from a pilot and, system developer perspective as well as an organizational perspective.

4 Interviews & Workshop

To meet the objectives of WP1, (conduct needs analysis, establish and adapt to context) seven interviews were carried out with the purpose of eliciting needs of the stakeholders involved in system development with focus in HMI.

A workshop was also performed to validate the identified HMI-criteria from the literature study of WP2.

4.1 Interviews

To meet the demand in the assignment specification of WP1 of the Brain Budget project, seven key persons outside the project group were interviewed with the purpose of conducting a needs analysis. The criterions for selection were: A) directly involved in systems development of HMI for fighter aircraft, or B) working on systems evaluation of HMI for fighter aircrafts, or C), potential end users of the system (fighter pilots) or pilots with extensive experience, as pilots and in system development efforts.

Three Technical Managers at Saab Aeronautics were interviewed (qualified under criterion A). They all have responsibilities for HMI-requirements. Four pilots were identified under criterion C; one working as a fighter pilot today with five years of work experience, one working as a helicopter pilot with experience in system evaluation including HMI-evaluations (more than 20 years of domain experience), one working as a fighter pilot today with experience of requirement specifications and system development (more than 20 years of domain experience), and one retired fighter pilot with experience of requirement specifications and system development (more than 30 years of domain experience). Two of the pilots and all three of the technical managers were qualified under criterion B.

All the interviews except one were carried out face to face (F2F). All interviews followed a similar procedure and had a semi-structured approach using predefined questions *(see Appendix 1).* The interviews started with a short description of the Brain Budget project and the interviewee were than asked to make a short description of their current tasks and former experiences relevant to the project. They were asked to go through a list of predefined HMI evaluation criteria compiled from the literature study *(see Table 5 in chapter 5.4)* and prioritize the criteria according to importance in the context of HMI evaluations of fighter aircraft cockpits.

4.1.1 Results

The results from the interviews gives a concurrent view of how system development is carried out today, how HMI-aspects are addressed today, or rather how HMI-aspects are not addressed today.

There is no single cause to point out; it is a combination of several aspects that result in the lack of sufficient HMI-evaluations which in the end leads to unsatisfactory HMI-design and implementation:

- There is a lack of systematic recurring HMI-evaluations and tests during the design process.
- There is a need to implement HMI-evaluations earlier in the design process to get an indication of whether the current design is on the right track.
- There is also a need to plan for and implement HMI-evaluations in an iterative manner throughout the design process.
- There is a need to review the design process in order to incorporate HMI-evaluations.

- There is a need to standardize HMI-evaluations that later can be referred to, both within projects but also between different projects.
- Tests and evaluations should be performed with “man in the loop”, which is not always the case today. In the past mission simulations were performed, both functional evaluations and system evaluations. The pilots/test persons had some difficulties to make judgments and this led to “reversed line of argument”, that is, if no one complained during the evaluations and tests everything was assumed to be ”good enough”.
- There is a lack of diagnostic testing. The lack of HMI-evaluations and testing can lead to an inadequate HMI and insufficient pilot-system performance.
- There are also examples where no HMI evaluation has been performed during the design process and this has lead to problems in closing HMI requirements.
- There is a great need to define quantitative HMI requirements to meet the so called “good enough requirement”.
- There is a need to develop methodology to be able to evaluate HMI and gather evidence that HMI is good enough. Today there are two major issues to handle concerning HMI; the first is to identify and address errors in the implementation (components that does not provide desired functionality) and second to identify improprieties of the design. Today, much time is spent on correcting errors during implementation while little time is spent on identifying improprieties in the design. There is a need to work in a more proactive way than is the case today. A sub-optimal interface that not help the pilot to interact fast and correct will lead to delays in decision making which might lead to severe consequences in combat situations.
- There is a lack of resources for HMI evaluations, especially time available to perform evaluations is very limited.
- Methodical knowledge needs to be improved within the organization.
- There is a need to use technical equipment during the evaluations such as simulation and tracking functions.
- There are constraints for instance resulting in system developers at Saab are not allowed to interact directly with operational pilots at the squadrons. Saab has test pilots in the organization that support system evaluation efforts but they might have different perspectives and experiences as well as a lower operational currency compared to the pilots at the squadrons, which in essence are the targeted end users. This may lead to biased results.
- Due to time consuming development cycles there is also a risk that dysfunctional or sub-optimal HMI is not corrected due to lack of time. Today Saab conducts their own evaluations first, then FMV, and then the Operational Test and Evaluation Unit (OT&E) for JAS39 Gripen. It would be desirable that Saab, FMV, and OT&E could perform evaluations in cooperation to be able to shorten the development cycle which could lead to better possibilities of optimizing the design before is it too late.
- There is also a need to conduct more holistic system evaluations at the early stages of the design cycle. Today, evaluations are sometimes performed to evaluate different system functions. When a development project for a specific function is closed, there are no resources and no infra-structure for addressing any dysfunctions or sub-optimal solutions of that specific function if they are identified after closure.
- When it comes to designing HMI-concepts, the organization is very dependent on a few subject matter experts (SMEs) in-house which make the organization

vulnerable. Without accurate HMI-evaluations and testing there is always a risk that some HMI-design problems might slip through which may effect the end result negatively.

- The project management and program management need to understand how critical it is to incorporate HMI-evaluations in the design process.
- On a methodological level, there are also needs for defining measures of workload, time critical events, allocation of functions (buttons) etc. If this is not considered, there is a risk that more complexity is built in the system without consideration of the end user and his/her brain budget during operational performance.
- When testing new systems it is also important to consider that the heritage from the old system might affect pilot performance and opinions of the new system. It is important to give pilots sufficient, but not too much, time to train on the new system, to avoid biases in the results of the evaluation.
- When designing and testing new systems it is also important to consider that the design shall support different end users: both “best in class” and average performers. The use of “Personas” might be one way of addressing this during the early stages of the development process.
- There is also a need to use generic and standardized scenarios when performing testing and evaluation in system development to reflect the right context for the measurement.
- There is also a need to use dynamic evaluation and testing.

According to the HMI-criteria that were identified in the literature study (*see table 5 in chapter 5.4)* no additional criteria were identified during the interviews except for mental workload. Overall, the identified criteria were assumed by the interviewees to be relevant. The ranking of importance made by the interviewees were also concurrent and the criteria of greatest importance were referring to the ability to preventing and recovering errors, learnability, and memorability. However, the criteria referring to attitude, relevance, and efficiency were also considered important. There was also one comment of great importance which was to differentiate criteria by design goals and usability goals. For example; “Not feel frustrated from using the system” is a usability goal, while “Menus, symbols and text should be grouped in a logical way” is a design goal. This person preferred the notion of usability goals. In the list of HMI-criteria presented for the interviewee, design goals and usability goals were all considered to be HMI goals.

The interviews also revealed that ideas that refer to design solutions (which not is the main focus within this project but is worth mentioning) such as the potential use of different information presentation modalities. Use of 3D-Audio, Tactile, and Eye pointing techniques might be one way to preserve Brain Budget. This is for future HMI-evaluations to find out. Additional methodologies referring to design solutions that were suggested were to prioritize functions by importance (criticality, frequency of use, etc.) before implementation.

4.1.2 Conclusions

Conclusions of the data collection from the interviews are:

- There is a lack of systematic recurring HMI-evaluations and testing during the system development and design projects/phases. Several interrelated issues are the reason for this situation and these need to be further considered during system development and design concerning the incorporation of HMI-evaluation methods.
- There are three major aspects that have been identified as critical to enhance the consideration and use of HMI-methods for system development. The three major aspects are:
    - *Organizational aspects*
    - *Methodological aspects*
    - *Technical aspects*

The scope of this project mainly concern the methodological aspects but in order to understand the failure and success of system development and design, all aspects needs to be understood, considered, and managed. This is the rationale for including them here, and also the reason why they have been mentioned during the interviews. Table 1 presents an attempt to categorize the different aspects and the corresponding needs identified during the interviews.

Table 1.

Needs identified during interviews classified under different aspects (Organizational, Methodological ,and Technical) effecting HMI-evaluation in system development and design

Aspect effecting HMI in system development & design	Needs identified within each category
Organizational aspects	• Perform systematic and recurring iterations of HMI-evaluation throughout the development process to be able to make correction as early as possible • Plan and budget for HMI-evaluations and tests already in the project planning phase • Incorporate HMI-evaluations in the design process • Inform and establish "buy-in" from project management and line managers and project sponsors/stakeholders about the value of using HMI-evaluations • Focusing on identifying improvements in design rather than identifying and correcting errors during implementation • Establish more resources to HMI evaluations (money, time, and know how) • Increase the possibility to consult end users (in this case operational pilots) • Shorten development cycles to increase the possibility that no sub-projects are closed before the associated components/functionalities have undergone rigorous holistic evaluations.

	• Increase cooperation with different stakeholders (Saab, FMV, OT&E) in system development to decrease iteration cycles • Increase the HMI “know how” in the organization to decrease the risks associated with being dependent of a few subject matter experts (SMEs)
Methodological aspects	• Develop standardized HMI-evaluations & tests that later can be referred to, both within projects but also between different projects • Perform holistic system evaluations as far as possible (not only functional evaluations) • Perform evaluations with end-users • Use evaluations that can diagnose identified issues in design • Operationalize HMI criteria and requirements (Usability, Mental Workload, etc.) • More consideration with regard to the end-user to mitigate the increasing complexity of the interface • Heritage of old systems need to be considered though it might effect and bias performance and opinions of new systems • Different end-users must be considered when performing testing; both the “best in class” and the “average” – use of Personas in the early stages might support this process • Use standardized generic scenarios in test-simulations to capture the right context to the extent possible Use dynamic evaluation and testing
Technical aspects	• Use manned simulation when performing testing • Use video and sound recording to be able to replay test sessions and conduct debriefings and post-evaluation

According to the identified HMI-criteria there is also a need to perform workload measures during evaluations, with the intent of assessing the impact/effect of the design solutions. Some of the identified HMI-criteria were also considered more critical than other. The criteria of Learnability/Memorability and User Errors had a tendency to be considered more important than the others.

4.2 Workshop

After the literature review, when potential identified HMI-criteria were identified *(see table 5 in chapter 5.4)*, a workshop was conducted in an attempt to A) validate the suitability of identified criteria, and B) to categorize the different criteria in different categories. Three researchers in the field of human performance and HMI participated and another researcher in the field of HMI facilitated the workshop. The attendees in the workshop were provided with all the different criteria on separate cards and were asked to cluster them using card sorting into subgroups made up by themselves. The three participants all worked individually to make their own grouping and categorization.

4.2.1 Result

The results from the workshop showed acceptance for the identified HMI-criteria from the literature study. The result from the workshop revealed some difficulties when it comes to categorizing different criteria into subgroups. One of the reasons for that is caused by the fact that different criteria in many cases overlap with each other. Criteria are not just black or white, but rather of grey scales. Some were also considered to have the characteristics of design goals, which confirm the findings during the previously conducted interviews. Main categories identified from each participant are summarized in *Table 2* below. The original categorization made from the literature study is shown in the left column. For a more comprehensive description of the criteria of each category identified during the workshop, see *Appendix 2*.

Table 2

Main HMI categories identified during Workshop

Main HMI categories identified during Workshop			
Original categorization	**Participant 1**	**Participant 2**	**Participant 3**
Relevance	Human System Efficiency	Easy to use/learn	General Design Goals
Efficiency	Feedback	Feedback	Specifications in system & functions
Learnability/Memorability	Learnability/Memorability	Error Prevention/Management	User Experience
User Errors	User Errors	Effectiveness	General Design Goals – Avoiding Errors
Attitude	Utility Attributes		User Oriented Design Goals
	Utility Relevance		System Oriented Design Goals
			Process/Usage

The result of the classification made during the workshop indicates that some criteria can be used under several main HMI categories and that many of the criteria intercept witheach other. The result correspond to findings made in the literature study *(see chapter* 5).

5 Literature study

5.1 History

Within systems development there are many things to consider when developing effective, usable and safe systems to support decision making and human operations (Nielsen, 1993, Norman 1998, Oscarsson, 2002, Alfredson et. al 2004, Dumas & Salzman, 2006, Albert & Still, 2011). One thing of great importance is the Human Machine Interaction (HMI). The main point of interest for this project is how to evaluate and measure HMI-criteria in the context of a fighter pilot-system interaction during system development.

The area of HMI originates from the Human Factors and Ergonomics (HFE) research field which has its roots as far back as 1850. The field of HFE has evolved significantly during the last century, especially since the 1950ies. From the beginning the main focus was in adapting equipment, workplaces, and tasks to human capabilities and limitations. This is also referring to the “Physical fit” or generation 1 of HFE. Later on, as the field evolved, the HFE started focusing in harmonizing and integrating humans, technology and work to enable effective systems. This is also referring to the “Cognitive fit” or generation 2 of HFE. Boff (2006) also discusses generation 3 and 4 of HFE. Generation 3 is about amplifying human cognitive and physical capabilities to perform work in symbiotic coupling with technology. Generation 4 is to biologically modify physical and/or cognitive capabilities to maximize human effectiveness (Boff, 2006). There is still more work to be done within all generations described above. This project is moving somewhere around generation 2.

5.2 HMI Evaluation Methods and Techniques for Systems Development

There are plenty of methods and techniques to choose from in the literature when conducting system evaluations. As always, the result from the methods used are depending of how well the methods are applied in the specific context of use. Oskarsson (2002), Alfredson et al. (2004), Castor et al. (2003), and Dumas & Salzman (2006) all provide extensive overviews of methods and techniques used in HMI-evaluation.

5.2.1 Expert methods

Expert methods relate to evaluation methods that are carried out by usability experts without involvement of end users. One of the most common methods is *“Heuristic Evaluation”* (Nielsen, 1992, 1993) where the interfaces are evaluated according to ten principles. Some examples of principles are: Good error messages, Preventing errors, Speak the user’s language, Feedback, and Consistency. Two of the advantages of this method are that it can be performed early in the design process and does not require extensive resources. Another advantage of the method that is often mentioned is that it can be performed without users, though methods not involving users will effect the validity of the results in a negative way.

5.2.2 Observations

Observation methods can be performed/designed/used in different ways. The most common way to observe user interaction is by *note taking*. Other ways to perform observation is *video recording* and *logging* of for example key stroke interactions. *Eye tracking techniques* can also be used to study user eye point of gaze and are preferably used in combination with video recording. Most observation methods generate a lot of data and a considerable amount of time for analysis is needed. Observations must always involve users, which effects validity of the result positively.

5.2.3 Verbal protocols

Verbal protocols can be used to study interaction. The theory is that both perception and cognitive activities can be translated into verbal language. One disadvantage is that highly automated behaviour is hard to verbalise. Two verbal protocol methods are *"think aloud"* (during interaction) and *"retrospective methods"* that refers to verbal protocols performed after the interaction. Memory may affect the use of retrospective method in a negative way. Combination of video recording and retrospective methods can be used in combination to enhance memory for more valid results.

5.2.4 Subjective evaluations

Subjective evaluation methods can be used when collecting the user opinions of a system. User opinions are very crucial for the success of the design. There are two main methods to chose from or combine; *Interviews* and *Surveys*. Both the use of interviews and surveys demands a lot of preparation and administration before, during, and after conduction.

5.2.5 Simulations

The use of *simulation methods* can reach from evaluating simple prototypes to evaluation of full scale system simulations. Simulation methods are often combined with other evaluation methods, for example subjective evaluation methods, verbal protocols and observations. Simulation is not a method in itself, but rather a means to provide a meaningful and realistic setting for interactions.

5.2.6 Performance measures

Performance measures most often refer to quantitative measures of performance. To obtain performance measures the use of above described evaluation methods can be used. Example of qualitative performance measures are: What does the user prefer (e.g. Design one or two), Workload measures, speed and time, amount of errors done, etc. Performance measures can also be set up as comparison to predefined goals of for example acceptable usability measures.

5.2.7 Mental Workload Measures

Methods that measure *Mental Workload* can be used to study how different design solutions affect the mental effort of the user. Subjective measures can be used (for example Bedford Rating Scale, see *Appendix 3*) to rate MWL, but also psycho-physiological methods, i.e. methods that measure a physiological response or behaviour in order to make assumptions concerning the psychological state, such as measuring heart rate and eye movements.

5.2.8 Summary of HMI Evaluation Methods and Techniques for Systems Development

Identified methods are seldom used alone and are preferably used in combination. They all have their strengths and weaknesses *(see Table 3).* The left column in Table 3 shows different evaluation methods (described earlier in this chapter) and the upper row shows different quality criteria for different methods. The different quality criteria are explained and listed below:

- *Early or late in design process* – refers to when in the design process the method should be applied (early/late)
- *User involvement* – refers to if the method require user involvement or not (yes/no)
- *Diagnose* – refers to the ability to diagnose specific problems and/or improvement in design (high/medium/low)
- *Resources* – refers to the amount of resources needed (time, costs, personnel, and equipment) (high/medium/low)
- *Know how* – refers to the methodological knowledge needed in the evaluation team (high/medium/low)
- *Analysis* – refers to the amount of data gathered and the amount of resources (time) needed for analysis (high/medium/low)

Table 3

Evaluation methods and their different quality criteria

Quality criteria / Evaluation method	Early/late in design	User involvement	Diagnose	Resources	Know how	Analysis
Heuristic evaluation	Early &Late	No	Medium	Low	Medium	Medium
Note taking	Early &Late	Yes	Medium	Low	Medium	Medium
Video recording	Early & Late	Yes	Medium	Medium	Medium	High
Logging	Early & Late	Yes	Medium	High	Medium	High
Eye tracking	Early & Late	Yes	Medium	High	High	High
Think aloud	Early & Late	Yes	Medium	Low	Low	Medium
Retrospective	Early & Late	Yes	Low	Low	Low	Medium
Interviews	Early & Late	Yes	High	Low	Medium	Medium
Surveys	Early & Late	Yes	High	Low	Medium	Medium
Simulations	Early & Late	Yes	Medium	High	High	High
Mental Workload	Late	Yes	Low	Medium	Medium	High

It is sometimes hard to separate different methods from each other due to the fact that they are often used in combination. For example, it is pointless to just do a simulation without evaluating the result using other methods. The quality criteria of every method must be carefully considered before the choice of method to be able to use the most suitable one/ones to fit the purpose of a particular evaluation.

5.3 Usability Assessment Methods

Usability assessment methods evolved from traditional human factors and ergonomics methods beginning in the early 1980s (Dumas & Salzman, 2006). Usability assessment methods have much in common with many of the HMI assessment methods and can to some extent be considered as subset of HMI methods. The most commonly used definition of usability is *"the extent to which a product can be used by specified users to achieve specific goals with effectiveness, efficiency and satisfaction in a specified context of use"* (ISO DIS 9241-11, 1998). Quesenbery (2004) broadens the ISO standard adding engaging (e.g. how pleasant, satisfying, or interesting an interface is to use). There are many variations in usability methodologies but there are some fundamental characteristics (Quesenbery, 2004): A focus in understanding the entire context of use; Evaluation and iteration as part of the development process; An user-centred approach to design; Designing for a specific audience.

5.3.1 Usability testing

Usability testing is an empirical method for eliciting and finding strengths and weaknesses in the usability of a product or a system (Dumas & Salzman, 2006). Dumas & Salzman (2006) claim that valid usability tests have the following six characteristics:

1. The focus is on the usability of a product or a system
2. The participants are end users or potential end users of a product or a system
3. The participants perform tasks with the product or the system
4. The participants are usually asked to think aloud as they perform tasks or immediately afterward
5. The data are recorded and analyzed. Data typically include qualitative measures such user satisfaction ratings and quantitative measures such as task success and error rates
6. The result and recommendations for improvement are communicated to appropriate audience, such as user interface designers, product managers, and programmers.

5.3.2 When is usability testing used

Usability assessment should be applied early in the development cycle and preferably in an iterative manner throughout the development cycle (ISO 13407, 1998, ISO 9241-210, 2010, Dumas & Salzman, 2006). Problems identified late in the development cycle can be hard to fix because of budget and time constraints. The Return of Investment (ROI) of HMI evaluation is greater when performed early in the design process (http://www.usit.com, viewed 2011-08-11).

Usability testing is conducted throughout the product development to guide design. The tests are then focusing on usability strengths and weaknesses and how to improve the design. Tests can also be conducted near the end of product development and are then focusing on measuring product or system efficiency and if it can be used as planned.

Usability tests can be used for *Early concept testing*, *Diagnostic testing*, *and Benchmark & comparison testing.* For example when performing benchmark and comparison testing of two systems the main point of interest is not to find errors in design but rather to find which of the systems is the best. On the other hand when performing diagnostic testing, the focus is on finding errors to be able to correct them. This put different demands on the method used to perform testing.

5.3.3 Examples of Usability Assessment Methods

Different kinds of HMI methods and techniques can be used as means to evaluating usability. Heuristic evaluation is one method used, but as stated above it does not involve end users. Interviews, surveys, note taking, video recording, logging, eye tracking, think aloud, retrospective testing and simulations are all methods that involves users and is suitable to use in usability testing.

5.3.4 Usability Assessment Methods Considerations

When choosing usability method for evaluation there are always many considerations in order to choose the most suitable method or methods. Often the best way is to combine different methods to get the most value from an evaluation. Examples of considerations that need to be made (http://www.usabilitynet.org, viewed 2011-08-10, Alfredson et al. 2004, Dumas & Salzman, 2006) are summarized in *Table 4.*

Table 4

Consideration for the choice of method

Consideration for choice of method	Comments
Early/late in design process	Ability to perform evaluation early and/or late in design process
Access/No Access to users	Accessibility to perform evaluation with or without end users
Resources	The amount of time, equipment, and money available
Ability to diagnose	Importance of finding design errors and/or design improvements
Know how	Skills and experience of evaluation team
Diagnostic versus Benchmark & comparison testing	Ability to make diagnoses of systems or just find out the best out of two systems

5.4 Usability Criteria

One of the most critical parts when conducting a usability evaluation is how to choose the most suitable criteria to measure and evaluate. As stated in the ISO 13407 (1998), ISO 9241-210 (2010), Quesenbery (2004) defining the context of use is crucial. Usability criteria can to some extent be considered as general but the result of usability test must always be put in the context of use. For example the criteria of *preventing errors* obviously have a more crucial importance in the context of a fighter pilot than in a context of an administrator working with an Office program.

There exist a large number of possible measurement of usability criteria (Nielsen, 1993, Löwgren, 1993, Quesenbery, 2004, Vallstrand, 2009, Usabilitynet.org, viewed 2011-07-01, Shneiderman, 2009). Nielsen (1993) describes 10 interaction design heuristics and Shneiderman (2009) describe 8 golden rules for interaction design that can be used in evaluating design. An attempt is made to classify different usability criteria under the five main categories; *Efficiency, Relevance, Learnability & Memorability, User Errors, and Attitude in table 5.*

Table 5

Usability criteria classified under the five main categories; Efficiency, Relevance, Learnability & Memorability, User Errors, and Attitude

Usability criteria classified in usability categories				
Relevance	**Efficiency**	**Learnability & Memorability**	**User errors**	**Attitude**
Ability to complete task	Visibility of system status	Recognition rather than recall	Support undo redo action	Not feel frustrated when using the system
Minimalistic design	Logical grouping of menus, symbols & texts	Consistent use of words & symbols	Preventing errors	Pleasant to use
	Enough time to complete tasks	Easy to understand words and symbols	Supporting recovering errors	Feeling of achieving high task effectiveness
	Fast system	Information is logical	Error massages appear when action may lead to severe errors	Fulfils the needs
		Short time to understand how to solve task	Confirming choices	Not worry that things went wrong
		Easy to learn how to use	Carefully considered default actions & values	No bad features
			Clear information when a task has been completed	

Note that some measurement criteria might be relevant to several usability categories. Different usability criteria affect each other and correlation between various usability attributes depend on the domain, the users experience and the context of use (Frokjear et al., 2000). Also aspects concerning Maintenance & Support, Tasks, Safety aspects, Marketing, and Business Goals might effect correlation and importance of the different usability criteria (*see Figure 1*) where an attempt is made to visualise the concept. The concept is very complex and there are many aspects to consider when measuring HMI criteria.

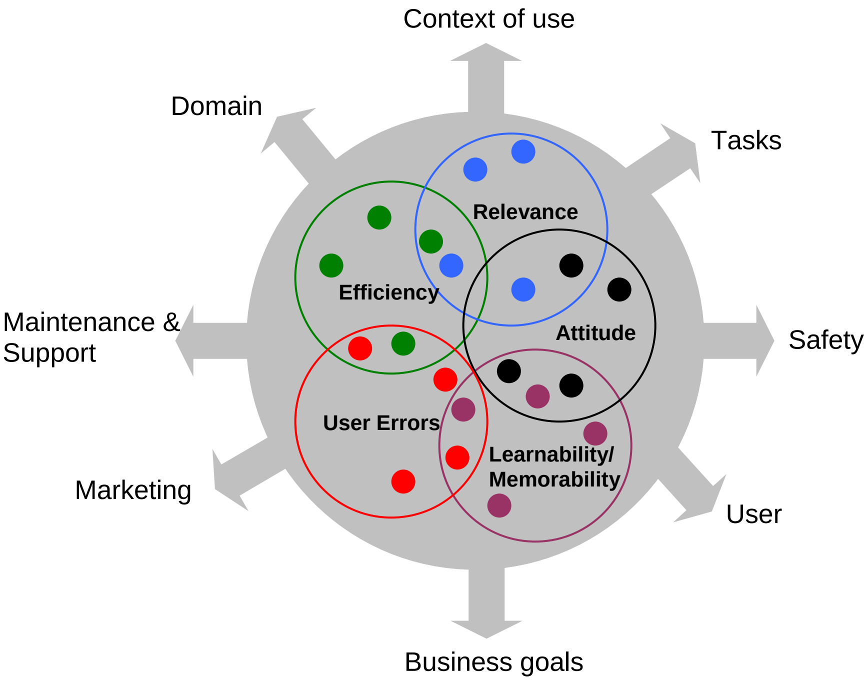


Figure 1

Visualisation of how different usability criteria intercept and effect other aspects.

Quesenbery (2004) uses a similar approach and speaks of the 5Es dimensions of usability (Effective, Efficient, Engaging, Error tolerant, Easy to learn). See *Table 6*.

Table 6

Quesenbery´s 5Es dimensions classified in user needs and possible design approaches.

Dimension	**User Needs**	**Possible Design Approaches**
Effective	Accuracy	• Provide feedback on all critical actions • Eliminate opportunities for error • Provide sufficient information for user decisions
Efficient	Operational Speed	• Design navigation for ideal and alternative workflows • Provide shortcuts • Use interaction styles and design widgets that support speed • Minimize extraneous elements on the screen
Engaging	To be drawn in	• Use clear language and terminology • Set a helpful tone, with level of conversation suitable for the user • Structure functions to match users tasks

Error tolerant	Validation and confirmation	• Transform "errors" into alternative paths • Use controls that aid in accurate selection • Be sure actions are easily reversible
Easy to learn	Just-in-time information	• Make the interface helpful with minimalist prompts and instruction provided where they are needed • Create "guided" interfaces for difficult or infrequent tasks

5.5 Weighting and Balancing Usability Criteria

There are obviously different ways of categorising usability criteria. Assume that we use the categorisation from *Table 5*. It would be convenient if each of the categories was equally important in all systems for every user in every context of use. However, this is not the case and there is a need to balance the importance of the categories between different systems for specific users in the specific context of use (Quesenbery, 2004).

An example is within the domain of fighter pilots where we can assume (according to performed interviews) that the criteria in the category User errors should be more heavily weighted (meaning higher importance in evaluation) than the criteria in the category Attitude. See example in *Figure 2*. Note that this is just an example about the concept of weighting and in depth analysis must performed according to the evaluation of specific system, task, context and user. Empirical data needs to be analysed within different contexts of use to get more valid results. For poorly performing systems, investigation what the circles should look like and testing could help identifying design problems. The size (relative proportion) of the different circles could either be defined as the level of importance for each of the measured criteria or defined as the level of criticality for each of the measured criteria.

We can also assume that the rating of importance of the different criteria will differ between experienced and non-experienced users, but also between end users and system developers. More research is needed to make any conclusions about the above hypothesis.

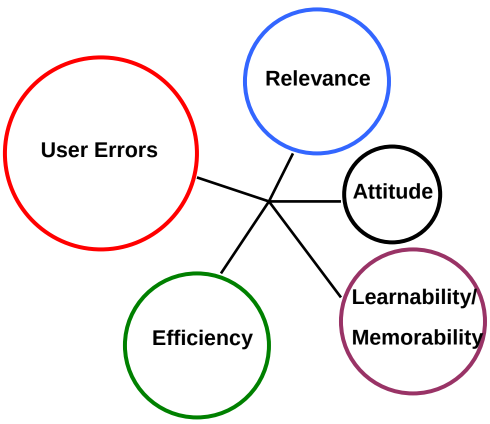


Figure 2

Example of the concept of weighting different usability criteria

5.6 Design Models

Traditionally within software system design different steps within the design processes has been executed in sequence. One commonly used design model is the “Waterfall model” (Preece, 1994), see *Figure 3*.

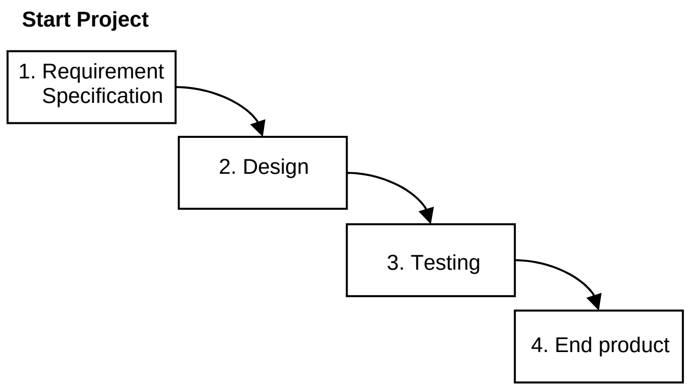


Figure 3

The Waterfall design model

There are two major drawbacks of using the Waterfall model. First it is almost impossible to understand all the users needs early in the design process, and secondly it is very hard to correct errors in design when testing and evaluation is performed late in the design process according to lack of time and money.

Due to the drawbacks of using the Waterfall model in system design a need to use a more iterative design model is evident. This corresponds to a more user-centred design approach called “User-Centered System Design lifecycle” described in ISO 13407 “Human-centred design process for interactive systems” (1999) and have been further developed in ISO 9241-210 “Ergonomics of human-system interaction” part 210 “Human-centred design for interactive systems” (2010). The model is described in *Figure 4* below.

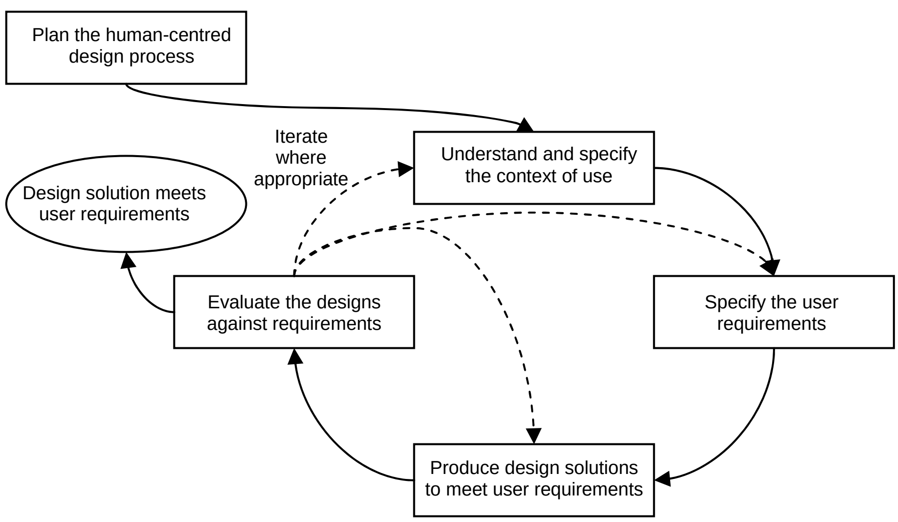


Figure 4

User-Centred Design lifecycle.

5.7 Usability Requirements versus Functional Requirements

To set up usability requirements when initiating a new development project we need to understand what we are trying to achieve, e.g. what do we mean when we are talking about usability in a specific context of use? This is not an easy question since usability is a large concept that consists of many different dimensions mentioned earlier in this report. It is not enough to say that we want a more user friendly system in the requirement specification.

Quesenbery (2004) gives the following description: Usability requirements answer questions like:

- How do users approach this work?
- How do users think about the tasks?
- How do users judge a successful experience?

Functional requirements answer the question:

- What does this system have to do?

To evaluate and answer the question about a system fulfilment of specified functional requirements the questions tend to be quite simple and binominal, meaning the answer is either yes or no, i.e., does the system work or not.

Because of the differences between usability and functional requirements some confusion can occur, especially when setting up usability requirements. Within system evaluation of usability requirements they often seem to be treated as functional requirements. However, the characteristics of usability requirements are not binominal but rather continuous. “What is good enough” usability requirements must be specified according to the context of use, task, and user. Specified usability requirements can then be evaluated using various types of usability methods.

5.8 Brain Budget

The term Brain Budget which also is the name of the project reported in this report has no clear definition in the literature. Here an attempt is made to investigate the term and give an explanation what we put in the term.

The human brain is assumed to have a limited capacity of information processing (Wickens, 1984), e.g. the brain has a limited capacity (budget) to process information. Different system design solutions are assumed to effect the brain budget differently (bad HMI-design put increasing load on the brain budget). Therefore, the aim in system design is to obtain design solutions supporting a sustainable resourcing of the brain budget. Design solutions should therefore carefully be evaluated with consideration to there effect on the brain budget. There are also other aspects such as for example experience, task complexity and hostile environments that affect the load on the brain budget.

The way we see it, there are also many similarities with the term brain budget and the mental workload (MWL) concept. Although the concept of MWL has been around for at least 40 years there is still no unified theory of mental workload (Castor, 2009). Hart & Staveland (1988) describe workload as *“…the perceived relationship between the amount of mental processing capability or resources, and the amount required by the task “*, O´Donnel & Eggemeier (1986, p. 42) describe workload as *“that portion of the operators limited capacity actually required to perform a particular task”*, and Hart & Wickens (1990, p. 258) describe mental workload as *“the effort invested by the human operator into task performance”*. The logic in the brain budget concept is that no task shall put more load on the brain budget than necessary. The goal when designing systems is to decrease and make free as much mental capacity/brain budget as possible to leave extra capacity for the operator to handle unexpected events.

The use of MWL measures in system evaluation can indicate the load of the brain budget affected by different design solutions. The MWL and the Brain Budget concepts are multi-faceted and are not just effected by design solutions interface, which must be considered in evaluation. Therefore a given design solution in an evaluation might just explain one dimension or maybe a couple of the different dimensions of the concept. Dimensions such as for example Experience, Knowledge, Skills, Task Demands, and Performance must be considered during systems evaluations.

In Castor et al. (2003) the multi-faceted concept of mental workload is further described. As Mental Workload is the theoretical concept most closely related to the Brain Budget concept some elaboration of similarities and differences are provided in *Figure 5*.

Both for Mental Workload and Brain Budget a number of factors and theoretical constructs are assumed to affect Mental Workload and Brain Budget. Both concepts are multi dimensional.

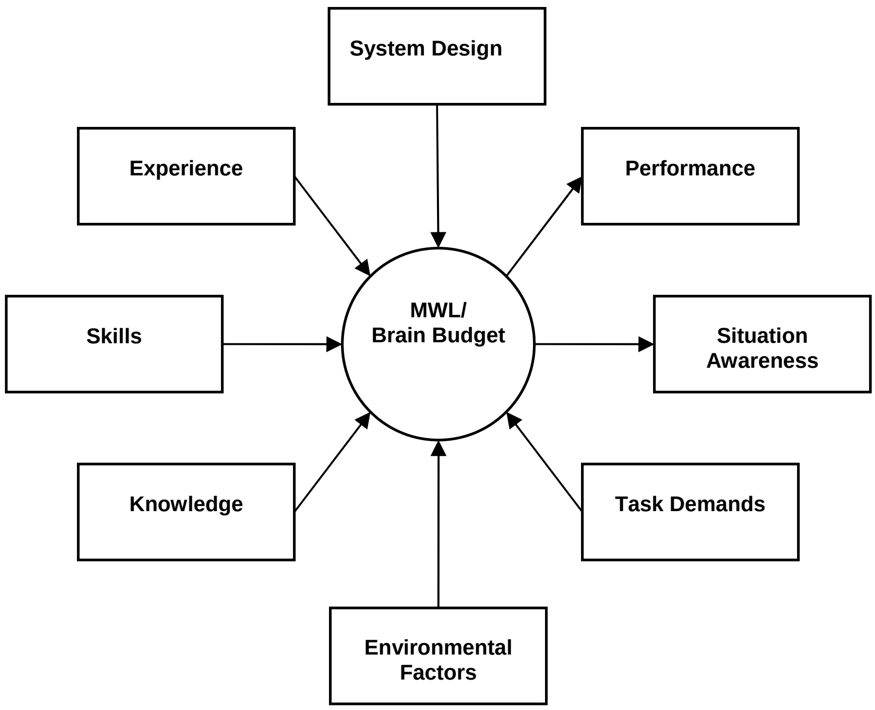


Figure 5

Visualisation of the similarities between Mental Workload and the Brain Budget concept.

Note again that the system design is just one among many other aspects that affect the brain budget and the brain budget and mental workload in turn affect aspects such as performance (overload decreases performance) and situation awareness (overload decreases situational awareness).

Figure 5 illustrate the complexity of the Brain Budget concept and show that there are no simple answers how to measure brain budget in an accurate way. Measuring and capturing a couple of the aspects that affect the brain budget will hopefully lead to a better understanding of the interrelations between the aspects which in the end will lead to better understanding in how to minimize the load of the operator in system design.

6 Results

Based from the results from the needs analysis (WP1) and the literature study (WP2) many aspects needs to be considered when performing HMI-evaluations during system design. One of the major aspects is the context of use of the system. All methods identified in this report have their strengths and weaknesses and the use of combinations of methods is preferable within HMI-evaluation in system development. Some of the needs that have been identified and needs to be covered are listed below.

- Ability to perform early and/or late in design
- Involve end-users (both the “best in class” and the “average”)
- Ability to make diagnoses
- Not be too complex
- Should not demand too much resources (money, time, and know how)
- Ability to perform within system simulations
- Consider learning curve for new system tested

In terms of performing HMI-evaluation it is also important to consider what we actually want to measure and evaluate. A list of identified and suggested criteria is presented in *Table 5 in chapter 5.4*. Those criteria should be complemented with the use of mental workload measures. The use of the identified criteria could also serve as input when setting up HMI-requirements in the requirement specification. More work needs to be done to evaluate they applicability of the identified criteria and this will be carried out later within the Brain Budget project.

As a result of the scope from the project plan, needs identified in the interviews, and identification of pros and cons made in the literature study of using different evaluation methods a new HMI methodology approach is suggested.

The methodology approach of HMI-evaluation is presented in the next chapter (chapter 7) that more thoroughly describes the suggested approach. The presented HMI Assessment Method Approach covers a wide range of methods described in *Table 3*. The HMI-methods used in the suggested approach are:

- Note taking
- Video recording
- Logging
- Think aloud
- Retrospection
- Interviews
- Surveys
- Simulations
- Mental Workload Measures (Bedford Rating Scale)

6.1 Constraints

According to the project specification the methodology approach only covers HMI-evaluation at a methodological level. Organizational aspects identified within WP1 and WP2 are at least as important as the methodological aspect and needs to be considered by the organization to make successful User-Centered System Design “happen” but are not further investigated in the scope of this project.

7 Methodology Approach "HMI Assessment Toolbox" (HAT)

The methodology approach could be described as a combination of dynamic measure of mental workload and HMI during simulated flight. The overall evaluation procedure is illustrated in *Figure 6.*

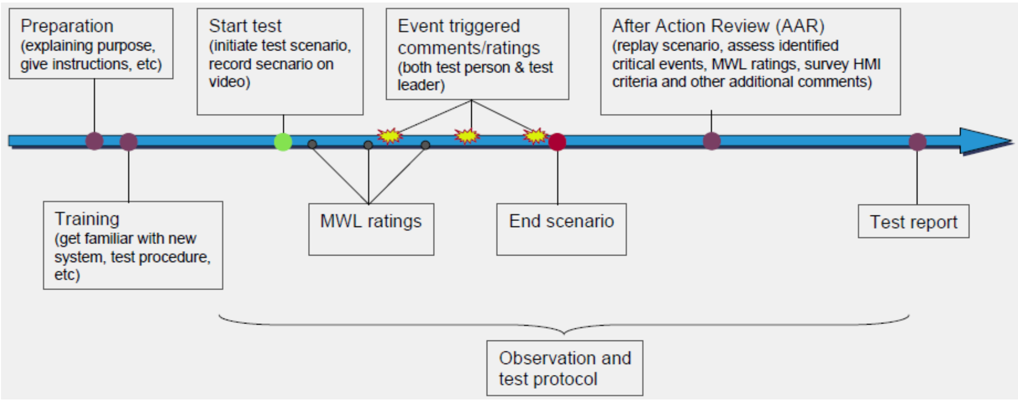


Figure 6

HMI evaluation method procedure.

7.1 Instructions HMI-assessment

- Use a generic (or design a new) relevant test scenario to form the platform for the evaluation
- Prepare logging equipment, surveys, interviews
- Prepare the test person by explaining purpose of the evaluation and give instructions
- The simulation is recorded on video
- The test person is asked to rate MWL during simulated flight at predefined occasions (scenario dependent) using the Bedford Scale
- The test person can at any time during simulated flight tag critical events "high lights" by talking aloud and these tags are noted by the test leader
- The test leader can also tag critical events
- After simulation an interview is performed and the test person gives the opportunity to make general and/or specific comments about the HMI
- The test person is then asked to answer the HMI assessment survey. The completion of the survey is facilitated by the test leader (see separate instruction for the HMI Assessment Survey)
- The simulated flight is then played back together with the test person and test leader
- The test leader facilitate the discussion based on tagged high lights, MWL ratings, HMI Assessment Survey and other comments gathered from the interview

7.2 Instructions HMI Assessment Survey

- The test person is guided by the test leader to answer each statement in the HMI Assessment Survey (*see Appendix 4*). Some statements might not be applicable.
- Each statement shall be rated on a six point scale (1-6) where 1 is “totally agree” and 6 is “totally disagree”. The value 3 shall be considered as “acceptable”
- After each statement the test person shall make comments and motivate the choice of value. The test leader can use the principle of asking the five why questions to penetrate the answer and if possible make a diagnose of potential issues
- The criticality of identified issues shall also be rated on a six point scale (1-6) where 1 is “not critical” and 6 is “very critical”
- The product of rated value indicate the issues should be prioritized (high values indicate needs for action → redesign)

7.3 Usability Assessment Matrix

A HMI Assessment Matrix (*see Appendix 5*) inspired from risk assessments techniques used within companies such as Vattenfall and SAS can be used in prioritizing issues according to the importance and criticality. The rated value of a specific HMI criterion is multiplied with the value of the rated value of criticality/importance. The multiplied value gives an indication if there is a need for action (re-design). The multiplied value should not be considered as absolute and additional comments made need to be considered.

7.4 Advantages of the HMI Assessment Tool Box

- Involves end-users
- Considers the context of use
- Proactive rather than reactive
- Can be used during different design phases
- Can be used during both function and system evaluation
- Can be used both in diagnostic testing and comparison & benchmark testing
- Combining quantitative and qualitative measures
- Can help prioritizing HMI problems
- Additional measures such as Eye-tracking measures, Key-stroke measures can be added to evaluation
- Considerations are taken to the importance and criticality of the different evaluated criteria

8 Appendices

8.1 Appendix 1. Interview Template

1. Describe your work tasks (what do you work with, which systems, contact persons, role in assignments?)
2. Do you have any experience in working with requirement specification and follow up of requirements (especially according to HMI)?
3. Are you or have you been involved in HMI-evaluation work?
4. How does the information flow look like between different actors within projects?
5. Do you see any needs how to ease and improve requirement specifications according to HMI?
6. Do you see any needs how to ease and improve HMI-evaluation work?
7. Do you have any suggestions or thoughts how you want persons who work with HMI-issues shall work to satisfy your needs according to manage HMI-issues?
8. What kinds of criteria do you think is applicable to use when measuring HMI?
9. Do you have any suggestions of studies that we should investigate or any other documentation, persons to talk to etc.?

8.2 Appendix 2. Result from Workshop

The tables describe the different categorisations created of the identified criteria from the literature made by the three workshop participants.

Participant 1:

Human System Efficiency
o Get clear information on when tasks have been full completed o Short time to understand how to sole tasks o Several tasks may be solved simultaneously o Ability to complete task o Symbols and buttons should be easy to understand o Enough time to complete tasks
Feedback
o The system status should be shown at all times and it is clear what is going on o Visibility of system status o Confirming choices o Error massages should appear when action may lead to severe error
Learnability/Memorability
o Easy to learn how to use o Recognition rather than recall o Information appear in a logical order o Menus, symbols and text should be grouped in a logical way
User Errors
o Preventing errors o Support undo redo action o Recovering errors o Consistency of words, symbols and standards o Carefully considered default action/values (not misleading the user)
Utility Attributes
o Pleasant to use o Not feel frustrated from using the system o Minimalist design o The system should perceived as fast o Feeling of achieving high task effectiveness o Not be worried that things went wrong using the system
Utility Relevance
o Feel that the system fulfil the needs o Number of good and bad features (explain and give examples) o Good or bad features (explain and give examples) o Number commands invoked/not invoked by users

Participant 2:

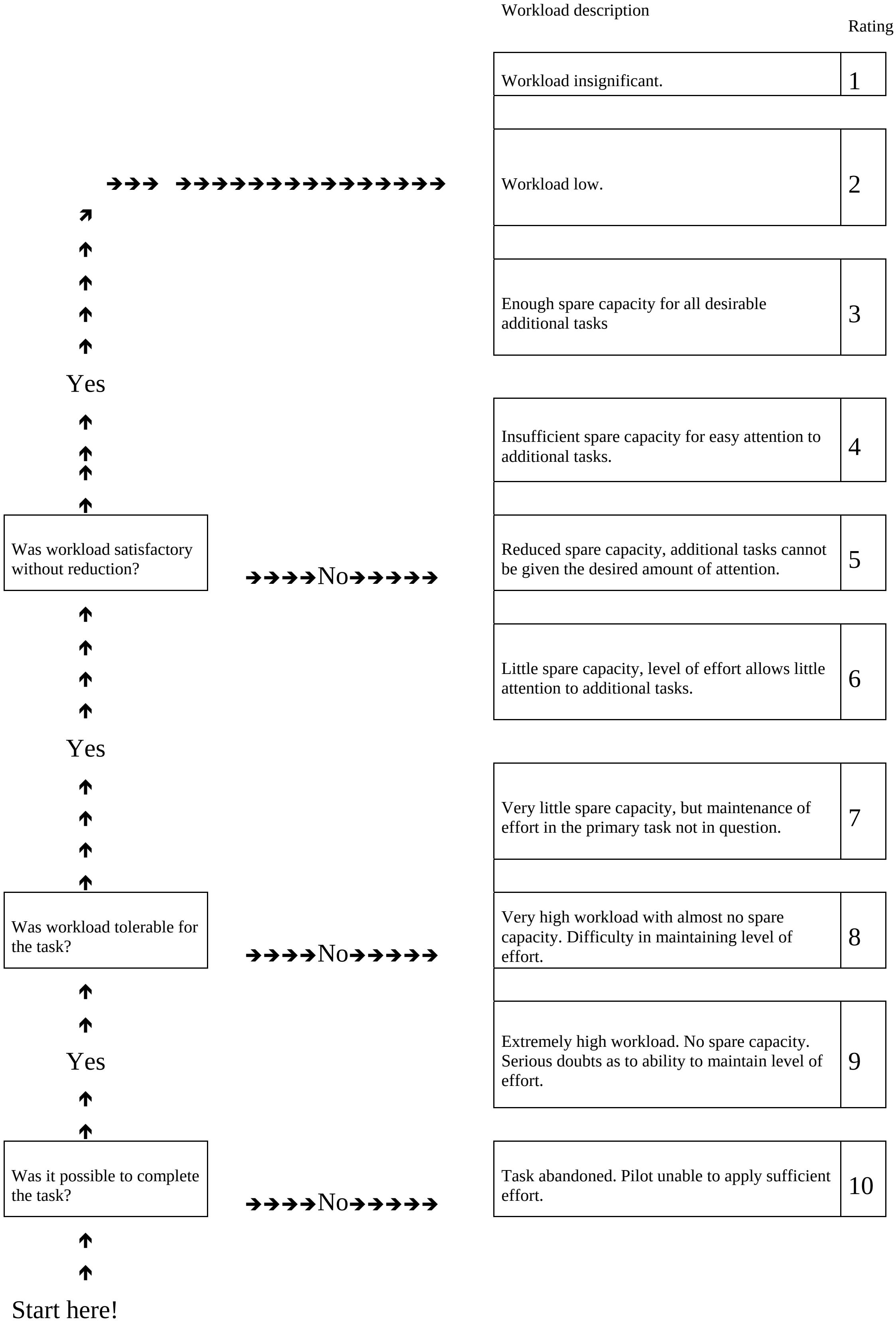
Easy to use/learn
o Information appear in a logical order o Consistency of words, symbols and standards o Menus, symbols and text should be grouped in a logical way o Number of commands invoked/ not invoked by users o Recognition rather than recall o Easy to learn how to use o Symbols and buttons should be easy to understand o Short time to understand how to solve tasks o Pleasant to use o Minimalist design
Feedback
o Visibility of system status o The system status should be shown at all times and it is clear what is going on o Get clear information on when tasks have been fully completed o Not feel frustrated from using the system
Error Prevention/Error Management

o Carefully considered default action/values (not misleading the user) o Error massages should appear when action may lead to severe errors o Preventing errors o Confirming choices o Recovering errors o Support undo redo action o Not be worried that things went wrong using the system
Effectiveness
o Enough time to complete tasks o The system should be received as fast o Feel that the system fulfil the needs o Ability to complete task o Feeling of achieving high task effectiveness o Several tasks may be solved simultaneously o Number of good and bad features (explain and give examples) o Good and bad features (explain and give examples)

Participant 3:

General Design Goals
o Minimalist design o Recognition rather than recall o Information appear in a logical order o The system should be received as fast
Specifications in System Design
o Support undo redo action o Visibility of system status o Ability to complete task o Confirming choices o Preventing errors o Recovering errors
User Experience
o Pleasant to use o Not feel frustrated from using the system o Feeling of achieving high task effectiveness o Feel that the system fulfil the needs o Not worried that things went wrong using the system o Get clear information on when tasks have been full completed
General Design Goals – Avoiding Errors
o Carefully considered default action/values (not misleading the user) o Number of good and bad features (explain and give examples) o Good or bad features (explain and give examples)
User Oriented Design Goals
o Enough time to complete tasks o Easy to learn how to use o Short time to understand how to solve tasks
System Oriented Design Goals
o Menus, symbols and text should be grouped in a logical way o Symbols and buttons should be easy to understand o Consistency of words, symbols and standards
Process/Usage
o The system status should be shown at all times and it is clear what is going on o Error massages should appear when action may lead to severe error

8.3 Appendix 3. Bedford Scale



8.4 Appendix 4. "HMI Assessment Survey"

HMI Criteria	Comments/Diagnose on a specific issue	Measured value (1-6) 3=acceptable	Weight/ Criticality	Total value
Relevance				
The system gives me the ability to complete task				
The system design is minimalistic				
Efficiency				
There is high visibility of system status				
Menus, symbols and text are grouped in a logical way				
There is enough time to complete tasks				
The system is perceived as fast				

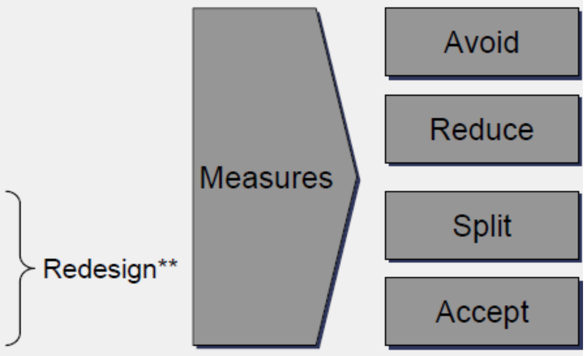
HMI Criteria	Comments/Diagnose on a specific issue	Measured value (1-6) 3=acceptable	Weight/ Criticality	Total value
Learnability & Memorability				
I use recognition rather then recall when interacting				
There is a consistent use of words and symbols				
Symbols and buttons are easy to understand				
Information appear in a logical order				
There is short time how to understand how to solve tasks				
The system is easy to learn how to use				

HMI Criteria	Comments/Diagnose on a specific issue	Measured value (1-6) 3=acceptable	Weight/ Criticality	Total value
User Errors				
The system support undo redo action				
The system preventing me from doing errors				
The system supporting me in recovering errors				
Error massages appear when action may lead to severe errors				
The system allow me to confirm choices				
There are carefully considered default actions and values				
I get clear information when a task has been fully completed				

HMI Criteria	Comments/Diagnose on a specific issue	Measured value (1-6) 3=acceptable	Weight/ Criticality	Total value
Attitude				
I do not feel frustrated from using the system				
The system is pleasant to use				
I have a feeling of achieving high task effectiveness when using the system				
I feel that the system fulfill the needs				
I do not worry that things went wrong when using the system				
There are many bad features in the system				

8.5 Appendix 5. "HMI Assessment Matrix"

V\C	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	4	6	8	10	12
3	3	6	9	12	15	18
4	4	8	12	16	20	24
5	5	10	15	20	25	30
6	6	12	18	24	30	36



Redesign**

* The logic of evaluation comes from risk assessments techniques (Enterprise Risk Management) used at SAS and Vattenfall

** The result (total HMI assessment value) must consider additional comments made by test persons

The V in the left column is the rated **v**alue of a specific HMI criteria and the C in the upper row is the rated **c**riticality of that specific criteria.

9 References

Albers, M. J., Still, B.(2011). Usability of complex information systems: Evaluation of user interaction. New York, NY: CRC Press, Taylor & Francis Group.

Alfredson, J., Oskarsson, P-A., Castor, M., & Svensson, J. (2004). Metodvalsverktyg: Ett hjälpmedel vid planering av MSI-utvärdering (FOI-R--1295--SE) [Instrument for choice of methods: A means of assistance in planning of man-system interaction evluation]. Linköping: Swedish Defence Research Agency.

Boff, K.R. (2006).Revolutions and shifting paradigms in human factors and ergonomics, Applied Ergonomics 37, pp 391-399.

Carlshamre, P. A. (2001). Usability perspective on requirements engineering. From methodology to product development, Doctorial Thesis No. 726, Linköping University, SE-581 83 Sweden.

Castor, M., Hanson, E., Svensson, E., Nählinder, S., LeBlaye, P., MacLeod, I., Wrighr, N., et al. (2003). GARTEUR Handbook of Mental Workload Measurement. Group of Aeronautical Research and Technology in Europe Technical Paper 145.

Castor, M. (2009). The use of structural equation modelling to describe the effect of operator functional state on air-to-air engagement outcomes, Doctorial Thesis No. 1251, Linköping University, SE-581 83 Sweden.

Dumas, Joseph S., Salzman, Marilyn C. (2006). Usability assessment methods. Reviews of Human Factors and Ergonomics, Volume 2, 2006, pp. 109-140.

Frokjaer, E., Hertzum, M., and Hornbaek, K. (2000). Measuring usability: Are effectiveness, efficiency, and satisfaction really correlated? Conference on Human Factors in Computing Systems (CHI-00) (N.Y., April 1-6 2000)

Hart, S. G., & Staveland, L. E. (1988). Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), Human Mental Workload. Amsterdam: Elsevier.

Hart, S.G., Wickens, C.D. (1990) Workload assessment and prediction, in H.R. Booher (Ed) MANPRINT. An approach to system integration (van Nostrand Reinhold, New York), 257-296.

ISO 9241-11:1998. (1998). Ergonomic requirements for office work with visual display terminals (VDTs), Part 11, Guidance on usability.

ISO TR 18529:2000. (2000). Ergonomics of human-system interaction: Human-centred lifecycle process descriptions.

ISO 9241-210:2010. (2010). Ergonomics of human-system interaction, Part 210, Human-centred design for interactive systems.

Löwgren, J. (1993). Human-computer interaction. Studentlitteratur, Lund, Sweden.

Myer, M.D. (2008). Qualitative research in business and management. London, Sage Publications.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. Proc. ACM CHI-92 (Monterey, CA, 3-7 May), pp. 373-380.

Nielsen, J. (1993). Usability Engineering. AP Professional, Cambridge, MA.

Norman, D.A. (1998). The Design of Everyday Things, MIT Press, London, England.

O´Donnell, R.D., Eggemeier, F.T. (1986). Workload assessment methodology, in K.R. Boff, L. Kaufman, J.P. Thomas (Eds). Handbook of perception and human performance. Vol. II. New York, NY: John Wiley & Sons.

Oskarsson, P-A. (2002). Översikt över metodik för MSI-utvärdering vid systemutveckling (FOI-R--0583--SE)[Overview of MSI evaluation methodology for systems development] Linköping: Swedish Defence Research Agency.

Preece, Sharp & Rogers. (2007). Interaction Design: Beyond human computer interaction, Second edition, John Wiley & Sons Ltd.

Quesenbery, W. (2004). Balancing the 5Es: Usability. Cutter IT Journal 17(2), p. 4-11.

Shneiderman, B. (2009). Designing the User Interface: Strategies for Effective Human Computer Interaction. New York. NY: Addison & Wesley.

Usabilitynet. (2011). http://www.usabilitynet.org/tools/methods.html, Viewed 2011-04-21).

Useit. (2011). http://www.useit.com/alertbox/roi-first-study.html, Viewed 2011-08-17

Vallstrand, J. (2009). Usability goals for administrative software. Master Thesis. Royal Institute of Technology, Stockholm, Sweden

Wickens, C.D. (1984). “Processing resources in attention”, in R. Parasuraman & D.R. Davies (Eds.), Varieties of attention, (pp. 63-102). New York: Academic Press.