CHALMERS

<table>
<tr><td colspan="12">1. Definition of evaluation</td></tr>
<tr><td colspan="12">2. Human-machine system description</td></tr>
<tr><td colspan="3">a. User profiling</td><td colspan="3">b. Task analysis</td><td colspan="3">c. Context description</td><td colspan="3">d. Interaction description</td></tr>
<tr><td colspan="12">3. Work load analysis (by GTS)</td></tr>
<tr><td colspan="3">a. Task demands</td><td colspan="3">b. Automation levels</td><td colspan="3">c. Mental workload</td><td colspan="3">d. Physical workload</td></tr>
<tr><td colspan="12">4. Interaction analysis</td></tr>
<tr><td colspan="4">a. Usability problem analysis
(by ECW and alarm-ECW)</td><td colspan="4">b. Use error analysis
(by PUEA)</td><td colspan="4">c. Ergonomics error analysis
(By PEEA)</td></tr>
</table>

Predicting mismatches in user-artefact interaction

Development of an analytical methodology to support design work

LARS-OLA BLIGÅRD

Division Design & Human Factors
Department of Product and Production Development
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2012

CHALMERS LARS-OLA BLIGÅRD Predicting mismatches in user-artefact interaction 2012

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Predicting mismatches in user-artefact interaction

Development of an analytical methodology to support design work

LARS-OLA BLIGÅRD

CHALMERS

Division Design & Human Factors
Department of Product and Production Development
Chalmers University of Technology

Gothenburg, Sweden, 2012

Predicting mismatches in user-artefact interaction
Development of an analytical methodology to support design work

LARS-OLA BLIGÅRD
ISBN 978-91-7385-744-4

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 3425
ISSN 0346-718X

Division Design & Human Factors
Department of Product and Production Development
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
The structure of the methodology Combined Cognitive and Physical Evaluation (CCPE)

Printed by
Reproservice at Chalmers University of Technology
Göteborg, Sweden 2012

Abstract

One human characteristic is that we use tools in our daily life. In the beginning they consisted of stones and sticks but today our tools have been developed into complex machines of different kinds, from consumer products such as mobile phones to technically complex systems such as nuclear power plants. The basic idea for all these products, machines and systems is that they are developed for improved comfort and to simplify our lives. However, this is not always the result, since sometimes there are problems when humans handle machines: what is known as mismatches are found in the interaction between human and machine. These mismatches not only decrease the utility of the machine, i.e. the human being's ability to reap the benefits of the machine, there is also the possibility that the human and the environment may be negatively impacted and damaged through use errors during interaction. When designing user interfaces for machines, human abilities and limitations on interaction need to be taken into consideration. An important part of product development is to study and analyse presumptive mismatches in a given design to be able to counteract them during subsequent re-designs at different stages of the product development process.

The purpose of this work has been to provide improved support for developers in handling and preventing mismatches in interaction early in the product development process. The goal of the work was to use existing methods to develop an improved Human Factors Engineering (HFE) methodology for predicting, identifying and presenting presumptive mismatches in the interaction between user and artefact.

This thesis presents a methodology, Combined Cognitive and Physical Evaluation (CCPE), which with a proactive and analytical approach evaluates mismatches in the interaction between human and artefact. CCPE methodology is built on the Cognitive Walkthrough (CW) and Predictive Human Error Analysis (PHEA) methods. These methods have been further developed into four new methods: Enhanced Cognitive Walkthrough (ECW), Predictive Use Error Analysis (PUEA), Predictive Ergonomic Error Analysis (PUEA) and Generic Task Specification (GTS). Apart from changes that prevent identified weaknesses and deficiencies in the original methods, the most important aspect of CCPE methodology is that it deals with both cognitive and physical ergonomics together. The aim of CCPE is to predict presumptive mismatches in human machine interaction, such as physical and mental work load, use error, usability problems and ergonomic errors, by using a process that supports the evaluators' cognitive processes. The purpose of the examination of both physical and cognitive usability problems and use errors in this interaction is to achieve a more holistic overall assessment. Furthermore, this results in a more cost-effective evaluation than would be the case if separate evaluation cognitive and physical ergonomic methods were used. CCPE also has a deep theoretical base in both these areas.

CCPE is a task-based methodology that uses a structured and systematic question process to search for mismatches in every single step in the interaction, as well as on more overall system level. The methodology was developed during work in product development projects in industry and academia, where existing evaluation methods were judged as not providing sufficient information about interaction problems. The research was problem-driven and performed as action research. During and after the development, CCPE and its methods were used in a number of evaluations where the methodology predicted, identified and presented presumptive mismatches in a structured way. The strength of CCPE is that its development was iterative and grounded in reality as well as based on a solid theoretical foundation.

The major strength of CCPE is the structured and systematic search for mismatches and the integration of cognitive and physical factors. The main weakness of CCPE is that it is more cumbersome and complicated to learn and use than the original methods as well as compared to other individual HFE methods. However, CCPE generates a more comprehensive result, which is presented in clear overviews, than is the case with other methods. CCPE also contributes to consensus and knowledge transfer in the evaluation group in a product development project. To conclude, this thesis has resulted in a methodology for predicting, identifying and presenting presumptive mismatches in the interaction between human and artefact. However, further work is needed to evaluate the reliability of the methodology and to develop computer aids to simplify its usage.

Sammanfattning

Ett av våra kännetecken som människor är att vi använder oss av redskap i vårt dagliga liv. Från att det från början har varit stenar och pinnar har det idag utvecklats till mycket komplexa tekniska maskiner av olika slag, allt från konsumentprodukter som mobiltelefoner till komplexa system som kärnkraftverk. Gemensamt är att de är skapade med grundtanken att göra våra liv enklare och mer komfortabla. Dock är detta inte alltid fallet En orsak är att det uppstår brister i samspelet mellan människan och maskinen, så kallade missmatcher i interaktionen. Missmatcherna leder inte bara till att nyttan med maskinen inte kommer människan till godo, utan människan kan också skada sig själv och sin omgivning genom att göra fel under interaktionen med maskinen. Maskinernas användargränssnitt bör därför designas så att de beaktar människans förmågor och begränsningar vid handhavandet. En del viktig del i produktutvecklingen är att studera och analysera möjliga missmatcher i en given design för att sedan kunna motverka dessa under produktutvecklingsprocessen genom att ändra utformningen.

Syftet med det arbete som presenteras här har varit att ge förbättrat stöd för utvecklarna att hantera och förebygga missmatcher i interaktionen tidigt under produktutvecklingsprocessen. Målet med arbetet har varit att, utifrån befintliga metoder, utveckla en förbättrad Human Factors Engineering (HFE) metod för att förutsäga, identifiera och presentera presumtiva missmatcher i samspelet mellan människa och maskin.

Avhandlingen presenterar en metodik, Combinded Cognitive and Physical Evaluation (CCPE), som med ett proaktivt och analytiskt angreppssätt söker efter missmatch i interaktionen mellan människan och maskinen. CCPE-metodiken bygger på metoderna Cognitive Walkthrough (CW) och Predictive Human Error Analysis (PHEA) som har vidareutvecklats till fyra nya metoder: Enhanced Cognitive Walkthrough (ECW), Predictive Use Error Analysis (PUEA), Predictive Ergonomic Error Analysis (PEEA) och Generic Task Specification (GTS). Förutom förändringar som motverkar identifierade svagheter och brister i ursprungsmetoderna, är det speciella med CCPE metodiken att den behandlar både fysisk och kognitiv ergonomi tillsammans. CCPE söker efter potentiella missmatcher i människa-maskininteraktion, såsom hög fysisk och mental belastning, användningsfel, användarvänlighetsproblem och ergonomiska fel, genom en process som stöttar utvärderarnas kognitiva processer. Syftet med den gemensamma sökningen efter både fysiska och kognitiva problem och fel är att uppnå en mer holistisk helhetsbedömning, samt att göra utvärderingen mer kostnadseffektiv än när separata utvärderingsmetoder används för kognitiva respektive fysiska ergonomiska aspekter. CCPE har och också en gedigen koppling till teorin inom respektive område.

CCPE är en uppgiftsbaserad metodik som strukturerat och systematiskt genom en frågeprocess söker efter missmatchar i varje enskilt delsteg i interaktionen, men också på en mer överliggande systemnivå. Utvecklingen av metodiken har skett under arbete med produktutvecklingsprojekt inom industri och akademi, där existerande utvärderingsmetoder inte har bedömts vara tillräckliga för att få bra svar om interaktionsproblem. Forskningens angreppssätt har därför varit problemdrivet och genomförts med aktionsforskning. CCPE metodiken och dess ingående metoder har efter utveckling använts i ett flertal utvärderingar, både i industri och akademi, där metodiken på ett strukturerat sätt upptäckt, identifierat och presenterat presumtiva missmatcher. En stor styrka är just att utvecklingen av CCPE skett på ett iterativt och verklighetsförankrat sätt. Vidare vilar metodutveckling på en solid teoretisk grund.

Den största styrkan med CCPE metodiken är det systematiska och strukturerande sökandet efter missmatcher samt integrationen av kognitiva och fysiska faktorer. Den principiella svagheten med CCPE är att den är omständligare och mer komplicerad att lära och utföra än originalmetoderna och jämfört med enskilda andra HFE-metoder. Emellertid skapar CCPE ett mycket mer omfattande och enkelt överblickbart resultat än separata metoder. CCPE bidrar också till att skapa en grund för koncensus och kunskapsöverföring i utvärderingsgruppen i ett produktutvecklingsprojekt. Avhandlingsarbetet har alltså resulterat i en metodik för att upptäcka, identifiera och presentera presumtiva missmatcher i interaktionen mellan människa och maskin. Vidare arbete behöver göras för att utvärdera metodikens reliabilitet samt utveckla instruktioner och datorstöd för att förenkla användandet.

Acknowledgment

First I would like to thank my former master thesis partner Sofia Wass (M.Sc.). If she had not contacted me from Switzerland and insisted that I contact a small medical-technology company in Mölnlycke, it is quite certain that this work would never have been done. Our lives are guided by the choices that destiny brings our way through chance.

I want to thank my supervisor and examiner Professor Anna-Lisa Osvalder for having faith in my crazy ideas and giving me the opportunity to work at Chalmers in many interesting projects, take part in conferences and project meetings in various parts of Sweden and Europe, and participate in courses on ergonomics/human factors engineering for engineering students and companies. Anna-Lisa has supported my work and provided me with invaluable reflections and comments, besides spending considerable time checking and correcting everything.

Furthermore, I would like to thank my assistant supervisor Professor MariAnne Karlsson for many rewarding discussions about the content of this thesis and about the subject in general, as well as for her fruitful comments on the writing in the thesis. My gratitude goes to Lina Lundgren (Lic. Eng.) for her faith in the methodology when applying it in an extreme case. I am also grateful to my colleagues Jonas Andersson (Lic. Eng), Cecilia Berlin (PhD), Cecilia Österman (Lic. Eng.) and Helena Strömberg (M.Sc.) for patience throughout all my discussions. Your comments and views have been very valuable for my work.

I am also grateful to my former colleagues Anna Thunberg (Lic. Eng.) for invaluable comments and views, and Maria Eriksson (M.Sc.) for immense assistance in the application of the methodology in industrial contexts. I also thank Carolina Osvalder and Samantha Pukala for help with language issues in the thesis.

Warm thanks are extended to Breas Medical AB, Gambro Lundia AB and Maquet Critical Care AB for the opportunity to test and further develop the different methods in their ongoing product development projects These real-life cases have been crucial for the methodology's emergence.

Thanks are due to Professor Emeritus Roland Örtengren who made it possible for me to register as a graduate student in spite of the difficult circumstances at the time. Moreover, I want to thank all my colleagues at the Division of Design & Human Factors for all our discussions and good humour during our pleasant coffee breaks.

Finally I want to thank a special group of people who contributed enormously to my belief in my work; in other words all the students who in their course projects and master thesis work have applied the methodology in various types of applications. You are too many to name individually, but great thanks to you all!

Lars-Ola Bligård
Göteborg, September 2012

Preface

My own intention with this work is not primarily to demonstrate my progress as a doctoral student and researcher. It is to ensure that we human beings can take better advantage of the benefits that technology offers, at the same time as we can avoid the disadvantages of the technology. We possess technology in order to make our lives easier, but unfortunately many people experience it as an obstacle, and in some cases the technology also constitutes a danger to life and health.

A reflection that came to me early in my contact with the field of Human Factors Engineering was that there is vast knowledge about how technology should be designed for adaptation to humans. In spite of this, much of the technology developed today is inadequately adapted to people. What I saw was a need to bring the knowledge out to those who design technology in reality, i.e. engineers, and to provide them with tools and methods for creating more human-centric technology. Knowledge that simply lies in a heap and is not used does no good.

The task that I want to carry out with my work is to establish a link between the knowledge produced in research and the engineers who design technology – in other words, to create and improve the methods and tools which are needed for adapting technology more closely to humans. I have therefore worked half-time in industry for three and a half years as a usability engineer to gain personal experience of work with human factors in real development projects in the field of medical equipment, identifying the problems and possibilities that exist. The hope is that this experience has contributed to the refined methods in operation during actual development projects.

Much of the work that I have done in the field of Human Factors Engineering has been in the form of practical usage of various methods in different projects. I have chosen to orient this thesis toward the more theoretical level as a complement. The thesis should be seen as a theoretical framework that weaves together the practically developed methodology.

In my master thesis, my co-author and I wrote: *"We believe that this thesis work is a step in the right direction, in the attempt to develop the Human Factors Engineering process."* I hope that my licentiate thesis also was a step in that direction and that this final part in the thesis trilogy is a further step in the same direction and that it can enhance the detection and identification of miss-match in the interaction between human and machine in a proactive manner.

I have to say that I feel privileged to have had the possibility to conduct research on my own idea without the control of an assigner or financer. This thesis is built on what I started to develop during my master thesis work. It has been a challenge not to work in a traditional pre-defined Ph.D. project, but it has been very instructive and inspiring to conduct my research that way.

To conclude this preface I quote George Bernard Shaw, who wrote: "*The reasonable man adapts himself to the world. The unreasonable one persists in trying to adapt the world to himself. All progress, therefore, depends upon the unreasonable man.*" Hopefully I belong to the group of unreasonable humans and will thus have been able to advance progress toward a better world. Life is too short for us to walk in each other's footsteps.

Appended papers to thesis

Paper I: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2006) *Using Enhanced Cognitive Walkthrough as a Usability Evaluation Method for Medical Equipment.* International Ergonomics Association Conference in Maastricht, the Netherlands, July 10-14

Content: The paper describes the early version of ECW and the application on an insulin pump.

Distribution of work: The method was developed by Lars-Ola Bligård and Sofia Wass. The insulin pump was evaluated by Lars-Ola Bligård and the paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder with Lars-Ola Bligård as having main responsibility. Lars-Ola Bligård also presented the paper at the conference.

Paper II: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2012) *Enhanced Cognitive Walkthrough – Development of the Cognitive Walkthrough Method to Better Predict, Identify and Present Usability Problems.* Submitted to Advance in Human Computer Interactions.

Content: Description of the current version of the Enhanced Cognitive Walkthrough (ECW) method and the developments that led to the method.

Distribution of work: The method was further developed by Lars-Ola Bligård. The paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder with Lars-Ola Bligård as having main responsibility.

Paper III: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2012*) Predictive Use Error Analysis – Development of AEA, SHERPA and PHEA to Better Predict, Identify and Present Use Errors.* Submitted to Industrial Ergonomics.

Content: Description of the current version of the Predictive Use Error Analysis (PUEA) method and the developments that led to the method.

Distribution of work: The method was developed by Lars-Ola Bligård. The paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder.

Paper IV: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2007) *An Analytical Approach for Predicting and Identifying Use Error and Usability Problem.* Lecture Notes in Computer Science, 4799 pp. 427–440.

Content: The paper describes how ECW and PUEA can be used together as one method.

Distribution of work: The further method development was performed by Lars-Ola Bligård. The paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder with Lars-Ola Bligård as having main responsibility. Lars- Ola Bligård also presented the paper at the conference.

Paper V: Bligård, Lars-Ola; Strömberg, Helena and Karlsson, MariAnne I.C. (2012) *Developers as users: A user evaluation of two new theoretical methods for usability assessment.* Submitted to Journal of Usability Studies

Content: The paper describes an interview study performed on user of ECW and PUEA (both students and professionals)

Distribution of work: The planning of the study was performed by all the authors with Lars-Ola Bligård as main responsible. The interviews and primary analysis of the data were performed by Helena Strömberg. A second analysis was performed by Lars-Ola Bligård. The paper was written in collaboration between all three authors.

Paper VI: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2006) *Predictive Ergonomic Error Analysis – A Method to Detect Incorrect Ergonomic Actions.* The 38th Annual Congress of the Nordic Ergonomics Society Conference (NES), Hämeenlinna, Finland, Sept. 25-27

Content: Description of the early version of the develop method Predictive Ergonomic Error Analysis (PEEA).

Distribution of work: The method was developed by Lars-Ola Bligård. The paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder with Lars-Ola Bligård as main responsible. Lars-Ola Bligård also presented the paper at the conference.

Paper VII: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2008) *Generic Task Specification – A Framework for Describing Task Demands and Mental/Physical Workloads in a Human-Machine System.* 2nd International Applied Human Factors and Ergonomics 2008, Las Vegas

Content: Description of the early version of the develop method Generic Task Specification (GTS).

Distribution of work: The method was developed by Lars-Ola Bligård. The paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder with Lars-Ola Bligård as main responsible. Lars-Ola Bligård also presented the paper at the conference.

Paper VIII: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2012) *CCPE - Methodology for a combined evaluation of cognitive and physical ergonomics in the interaction between human and machine.* Factors and Ergonomics in Manufacturing & Service Industries, Volume 19, Issue 6

Content: The paper describes the developed methodology Combined Cognitive and Physical Evaluation (CCPE) including the current version of Generic Task Specification.

Distribution of work: The methodology was developed by Lars-Ola Bligård. The paper was written by Lars-Ola Bligård and Anna-Lisa Osvalder with Lars-Ola Bligård as main responsible.

Appended appendix to thesis

The appendices present studies performed to evaluate the CCPE methodology. Each study is reported in the form of an appendix to fit in the thesis and will later be published in an altered form.

Appendix A *PEEA: Evaluation of ergonomic error in the interaction with computer mice*

Content: The appendix describes the comparison of evaluation by Predictive Ergonomic Error Analysis with an evaluation by empirical test of computer mice.

Distribution of work: The planning of the study was performed by Lars-Ola Bligård. The evaluation with PEEA was performed by Lars-Ola Bligård and Anna-Lisa Osvalder and the empirical tests were lead by Lars-Ola Bligård. The analysis of the video recordings from the empirical tests was performed by Magnus Renström. The comparison of results from the empirical and theoretical evaluations was performed by Lars-Ola Bligård, who also together with Anna-Lisa Osvalder wrote the appendix.

Appendix B *PEEA: Evaluation of ergonomic error in the interaction with stable tools*

Content: The appendix describes the comparison of evaluation by Predictive Ergonomic Error Analysis with an evaluation by empirical test of stable tools.

Distribution of work: The planning of the study was performed by Lars-Ola Bligård. The evaluation with PEEA was performed by Lars-Ola Bligård and Anna-Lisa Osvalder and the empirical tests were lead by Lars-Ola Bligård. The analysis of the video recordings from the empirical tests was performed by Magnus Renström. The comparison of results from the empirical and theoretical evaluations was performed by Lars-Ola Bligård, who also together with Anna-Lisa Osvalder wrote the appendix.

Appendix C *CCPE: Evaluation of mismatches in the interaction with a vacuum cleaner*

Content: The appendix describes the comparison of evaluation by Combined Cognitive and Physical Evaluation with an evaluation by usability test of a vacuum cleaner.

Distribution of work: The planning of the study was performed by Lars-Ola Bligård. The evaluation with CCPE was performed by Lars-Ola Bligård and Anna-Lisa Osvalder and the usability test were lead by Lars-Ola Bligård. The analysis of the video recordings from the empirical tests was performed by Magnus Renström. The comparison of results from the empirical and theoretical evaluations was performed by Lars-Ola Bligård, who also together with Anna-Lisa Osvalder wrote the appendix.

Appendix D *CCPE: Evaluation of mismatches in the interaction with an office chair*

Content: The appendix describes the comparison of evaluation by Combined Cognitive and Physical Evaluation with an evaluation by usability test of an office chair.

Distribution of work: The planning of the study was performed by Lars-Ola Bligård. The evaluation with CCPE was performed by Lars-Ola Bligård and Anna-Lisa Osvalder and the usability test were lead by Lars-Ola Bligård. The analysis of the video recordings from the empirical tests was performed by Magnus Renström. The comparison of results from the empirical and theoretical evaluations was performed by Lars-Ola Bligård, who also together with Anna-Lisa Osvalder wrote the appendix.

Appendix E *PUEA: Evaluation of use error in the interaction when kitesurfing*

Content: The appendix describes the comparison of evaluation by Predictive Use Error Analysis with an observation study of kitesurfing.

Distribution of work: The planning of the study was performed by Lars-Ola Bligård. The evaluations with PUEA were performed by Lars-Ola Bligård and Lina Lundgren and as well as the observation study. The analysis of video recordings from observation study was performed by Lars-Ola Bligård and Lina Lundgren. The comparison of results from the observation study and theoretical evaluations was performed by Lars-Ola Bligård, who also together with Anna-Lisa Osvalder wrote the appendix.

Appendix F *Templates for ECW, PUEA, PEEA and Alarm-ECW*

Also relevant published papers but not appended to thesis

Paper IX: Bligård, Lars-Ola; Wass, Sofia; Liljegren, Erik and Osvalder, Anna-Lisa (2003) *Using a Human Factors Engineering Process to Develop New User Interfaces for Home Care Air-Flow Generators.* Proceeding of the 35th Annual Congress of the Nordic Ergonomics Society Conference (NES), Reykjavik, Island, Aug. 10-13

Paper X: Liljegren, Erik; Bligård, Lars-Ola and Osvalder, Anna-Lisa (2003) *Developing User-Friendly Interfaces for Medical Devices.* Proceedings of the 8th IFAC Symposium on Automated Systems Based on Human Skill, Göteborg Sweden. Sept. 22-24

Paper XI: Bligård Lars-Ola, Jönsson, Anna and Osvalder Anna-Lisa (2004) *Eliciting User Requirements and Evaluation of Usability for Insulin Pumps*. Proceeding of the 36th Annual Congress of the Nordic Ergonomics Society Conference (NES), Kolding, Denmark, Aug. 16-18

Paper XII: Bligård, Lars-Ola and Thunberg, Anna (2007) *An analytical usability method for alarm message evaluation: Alarm-ECW*. Proceedings of the 39th Nordic Ergonomics Society Conference, Oct 1-3 2007, Lysekil, Sweden

Paper XIII: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2009) *Methods for risk analysis of use of medical equipment.* 18th Society for Risk Analysis - Europe Annual Meeting, Karlstad, Sweden

Paper XIV: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2010) *Methodology for a combined evaluation of cognitive and physical ergonomic aspects of medical equipment.* 3rd Applied Human Factors and Ergonomics (AHFE) International Conference 2010, Miami

Paper XV: Bligård, Lars-Ola and Osvalder, Anna-Lisa (2010) *Predicting and Identifying Mismatches in the Human Machine Interaction Design - A Method Useful in the Product Development Process.* NordDesign 2010

Paper XVI: Lundgren, Lina; Bligård, Lars-Ola; Brorsson, Sofia and Osvalder, Anna-Lisa (2011) *Implementation of usability analysis to detect problems in the management of kitesurfing equipment*. Procedia Engineering, APCST: 5th Asia Pacific Conference on Sports Technology

Paper XVII: Löfqvist, Lotta; Babapour Chafi, Maral; Osvalder, Anna-Lisa; Bligård, Lars-Ola and Pinzke, Stefan (2012) *Ergonomic Evaluation of Long Shafted Tools Used in Horse Stables - The effects of shaft length variation and work technique on working posture.* Submitted to International Journal of Human Factors and Ergonomics

Table of contents

Appended appendices

Appendix A	PEEA: Evaluation of ergonomic error *in the* interaction with computer mice
Appendix B	PEEA: Evaluation of ergonomic error *in the* interaction with stable tools
Appendix C	CCPE: Evaluation of mismatches in the interaction with a vacuum cleaner
Appendix D	CCPE: Evaluation of mismatches in the interaction with an office chair
Appendix E	PUEA: Evaluation of use error in the interaction when kitesurfing
Appendix F	Templates for ECW, PUEA, PEEA and Alarm-ECW

Appended papers

Paper I	Using Enhanced Cognitive Walkthrough as a Usability Evaluation Method for Medical Equipment
Paper II	Enhanced Cognitive Walkthrough – Development of the Cognitive Walkthrough Method to Better Predict, Identify and Present Usability Problems
Paper III	Predictive Use Error Analysis – Development of AEA, SHERPA and PHEA to Better Predict, Identify and Present Use Errors
Paper IV	An Analytical Approach for Predicting and Identifying Use Error and Usability Problem
Paper V	Developers as users: A user evaluation of two new theoretical methods for usability assessment
Paper VI	Predictive Ergonomic Error Analysis – A Method to Detect Incorrect Ergonomic Actions
Paper VII	Generic Task Specification – A Framework for Describing Task Demands and Mental/Physical Workloads in a Human-Machine System
Paper VIII	CCPE - Methodology for a combined evaluation of cognitive and physical ergonomics in the interaction between human and machine

1 Setting up the stage

The chapter begins with a fictional story about the problems that can occur when there is a mismatch between human and machine, and the effects this may cause. This is followed by a story about what can happen when one method from the methodology presented in this thesis is used. The story takes place in a fictitious company developing home-care ventilators.

A few weeks back the company purchased new ergonomic chairs. Kent, who is the company ergonomist, is now walking around to see how the chairs work for the employees in the office. Kent comes to Johanna, who sits and works at her desk. Johanna has a somewhat strange sitting position, seen most clearly in the posture of her hips.

Kent: Hi Johanna! How is the new chair?
Johanna: Hi Kent, it's nice to sit in. I've adjusted it myself for maximum comfort.
Kent: Well, your hips are positioned a bit strangely – you shouldn't sit this way because it may produce harmful effects on your lower back in the long run. Better that you adjust the angle between the thighs and the back so you get a more correct posture.
Johanna: Oh, I didn't realise that. It just felt good to sit in this position so that's the way I set it. But thanks for pointing this out to me.

Kent: Now I see that you also have the seat in a somewhat strange position.
Johanna: Yes, I was looking for the setting, but I couldn't find it.
Kent: Use this knob to adjust it.
Johanna: Ah! There you go. Strange I didn't see it before. I'll adjust the seat immediately.

Johanna: Hold on … how do I do? I thought I had to turn the knob but nothing happens.
Kent: You need to push it. It's a button you have to press, not a knob to rotate. Test again.
Johanna: Now it works, but how was I expected to know that this is how it works?

Kent: OK, why not adjust the lumbar support as well?
Johanna: Lumbar support? What's that?
Kent: You pump up a small cushion in the backrest so it conforms to the shape of your spine. The handle is there.
Johanna: Oh, what a lot of features there are!
Kent: Here's the manual. I'll show you.
Johanna: Yes, this feels much better for my lower back.

Kent: Um, I checked a bit here in the manual and you can adjust the armrests also like this, setting them at an angle. I had no idea.
Johanna: Oh, I had no idea either. This chair has many secrets.

Johanna: I just thought of another thing. It's really inconvenient to access this adjustment when you sit down, but you have to sit in the chair to know what level is good.
Kent: No, it's not a good twist of your arm when you make that adjustment. Hope you do'nt have to do that often.

Johanna: I'll just make a little adjustment to the seat back again.
Johanna: Ouch!

Kent: What happened?
Johanna: I pinched my finger when I changed the setting. But it doesn't seem to have punctured my skin.
Kent: Good that it wasn't any worse. Dangerous stuff this with chair settings...

Johanna: Yes, but also all these complex settings, there are so many settings and they depend on each other. They say that you should change your working position frequently, but it's very time-consuming to do. Especially when you have to sit in odd postures with the risk of pinching your fingers when making adjustments…

Kent: But at least we've got a good working posture for you now.
Johanna: Yes, much better than I could have managed myself.
Kent: Hope everything is okay. I'll come by this way in the afternoon to see how it works.
Johanna: OK, see you!

While Kent walks away from Johanna he thinks: "How can a chair that is said to be ergonomic be so difficult to understand...? Should I have to work in awkward postures or hurt myself to get an ergonomic work position? ... Or is this chair truly ergonomic? Even if the seating position is correct? Could it be that the chair has become so complex with all the features and settings that nobody knows how to handle it? And then no-one can get a good posture... "Kent summarizes his reasoning: "Good ergonomics must also include the journey and not just the goal... true ergonomics needs to apply to both body and mind... and above all, body and mind together."

This story described a number of areas where a mismatch between human and machine can be expressed. In the section below the problems are presented in the order in which they appear. The list makes no claim to being comprehensive but is just one example of how the problems manifest themselves.

- The user does not know how the machine should be set
- The user cannot find the settings on the machine
- The user cannot manage the settings
- The user does not know which settings are available
- The expert does not know which settings are available
- Poor working posture during settings
- The user is injured when handling the machine
- Use of the machine is complex

The above narrative has accordingly described how a mismatch can occur. But how to detect mismatches in advance so they could be counteracted? The next narrative describes how a session with the goal to detect mismatches can occur. The new context is the development of medical equipment produced by above the company.

The door opened and four people entered the meeting room. They were the usability engineer, the system engineer, the product manager, and a person from quality/regulatory. The system engineer and product manager had previously worked in medical care, while the other two had different backgrounds. They all sat down at the round table and the usability engineer explained the occasion.

"Welcome to this session about risk analysis in the use of our new ventilator. Today's assignment assumes that the patient in the home environment is to add water to the machine's humidifier before the night's treatment. Our goal at the meeting will be to investigate whether the user's behaviour can create any risks and, if so, how we should handle them. By conducting this analysis so early in the machine's development we have an opportunity both to set requirements and to change the physical design, not just to write warnings in the manual."

The other participants open their papers and look up the HTA tree which describes how the user should add water to the humidifier. Analysis of the HTA was done previously in order to determine how the development project views the method of using the machine. It took some time, but now there is agreement on what the "correct" method is. The usability engineer continues.

"We begin the procedure at the function level, and I think it is appropriate that we begin with the question 'What happens if the user performs functions/tasks correctly but at the wrong time?' We have the industrial designer's detailed sketches before us, and we have this simple physical model. As you know, the correct sequence is, in short, to remove the humidifier and carry it to the tap, open the lid and add water up to the mark, close the lid and carry the humidifier back, then attach it to the machine."

"I can imagine that someone might try to add water while the machine is running," says the product manager quickly.

"Is it actually a realistic scenario that someone would do that?" remarks the system engineer. "The machine would then blow freely and start an alarm for low pressure. Besides, I would like to see a patient who is wearing the mask while at the same time trying to add water..."

"Well, as we all know, patients can do the most amazing things with their machines. I know what you've told us about what you've seen on your home visits," says the quality/regulatory person and turns to the product manager.

"Yes, that's true," says the product manager. "The users are very inventive... So I think it's quite a plausible error that we must analyse."

"I've written it up," says the usability engineer, who is taking minutes. "I interpret this as an error of type P2, 'Incorrect plan executed'."
Nobody objects, so the usability engineer proceeds.

"What would be the cause of someone doing this? Based on what the system engineer said earlier, I don't think it's something that one just happens to do, since it is rather difficult to do while wearing the mask. I think it's because the user doesn't know how to act, in other words a mistake."

"It probably can't be classified as a rule-based error, since as far as I know there are no ventilators that one fills up during an operation," says the product manager. "But the user might associate it with some other product in the home. A steam iron can be handled like that."

"I note that the cause is mainly a lack of knowledge on the part of the user, so I'll record it as a rule-based or knowledge-based mistake," says the usability engineer. "Now let's turn to the consequences. If I understand the design right, it means that when the humidifier disappears, the air passage is no longer complete and air leaks out, which makes the motor blow as much air as possible? Is that the case?"

There are no objections, so the usability engineer goes on. "And will the effect on the patient be that no treatment is possible then?"

"Well, not in reality," replies the quality/regulatory person. "It is the patient who removes the humidifier, so the patient is not sleeping then. Therefore the level of consequence must be the lowest, since the user experiences only discomfort."

"The next item of investigation is error detection," explains the usability engineer. "Will the user notice that he/she has made a mistake before any serious consequences occur?"

"Without a doubt," answers the system engineer. "The ventilator alarm begins immediately, so I find it hard to see that anyone would fail to understand that something is being done wrong. Any objections?"

Silence prevails, so the usability engineer records a number 5 for error detection and says: "The next item of investigation concerns recovery from error. Is it just a matter of putting the humidifier back in place for everything to work again?"

The system engineer nods, and the usability engineer notes this, saying: "Is there any protection in order to counteract the consequences?"

"The alarm is there to deliver a warning when no treatment is given, so it should function as protection in this case."

"Duly recorded," says the system engineer. "Does the present design include any measures to prevent it from being removed during operation?"

"Nope," says the system engineer. "We must discuss at the project meeting whether such a feature is needed. There is no major risk involved in this error. But it can always be written into the manual."

"Shall we write in the manual that one must not add water while the machine is running? This would be yet another warning," thinks the quality/regulatory person.

"It is enough to write in the instructions for the humidifier that the machine should be completely switched off when one adds water," interjects the product manager. "There is also an electrical safety aspect to all this. I am noting that it should be a requirement for the manual writers."

"Then let's continue with the analysis," says the usability engineer. "Does anyone have any further errors connected with 'What happens if the user performs functions/tasks correctly but at the wrong time?"

"A similar event that can happen, and which seems to me more probable than the preceding one, is that the patient adds water while the humidifier is still coupled to the machine," says the system engineer.

"With a water pitcher through the air outlet?" wonders the quality/regulatory person.

"That's a possibility I hadn't thought of. Actually, what I imagined was someone putting the entire machine under the tap. We are, of course, supposed to design a machine that weighs very little."

The risk analysis of use continues like this and, once finished, it leads to the dismissal of many risks, but also to many important requirements and proposals for design changes. In the two fictional stories above, the first exemplified how a mismatch between the human and the artefact may appear and the second showed how it can be done in advance by trying to identify this mismatch using the methodology presented in this thesis. The following parts of the thesis will now describe in more detail the purpose of this dissertation, the theoretical framework of the methodology and detailed presentation of the methodology, known as CCPE - Combined Cognitive and Physical Evaluation.

2 Introduction

The thesis describes the proposed methodology and method development. The development resulted in a new analytical methodology called Combined Cognitive and Physical Evaluation (CCPE). It encompasses four new methods: Generic Task Specification (GTS), Enhanced Cognitive Walkthrough (ECW), Predictive Use Error Analysis (PUEA) and Predictive Ergonomic Error Analysis (PEEA).

The thesis also discusses the theoretical and methodological framework in which the methods operate. Development of the methods and methodology has been conducted primarily within the domain medical technology but has also been applied in several other domains, i.e. the application areas go from consumer products, particularly those sold as ergonomic products, to more advanced technical products.

2.1 Background

We humans are not perfect beings. We often have the best of intentions, but sometimes it just goes wrong. This has given rise to a well-known Latin proverb ”*Errare humanum est*” (To Err is Human). The proverb is ascribed to the Roman senator Cicero in the century before Christ. Hence, it is nothing new that humans err, but technical development has made the potential consequences of human error more extensive. In the era of Cicero it often needed to be an error made by a commander to harm or kill a large amount of people in a very short period of time. Today many more people are in a position to make catastrophic errors, such as air traffic pilots, nuclear power plant operators and so on. Many major accidents have been ascribed to the so-called human factor. In the same way as the technology increases the effect of our good sides as human beings, it also increases the effect of our bad sides. One of the bad sides is the ability to err.

Today much effort is being invested in the creation of technical systems that are reliable and safe. If the human component of the systems is not taken into consideration, the systems cannot be completely safe. A lot of research has been conducted in this area and the results unambiguously show that if technology is adapted to human characteristics, abilities and limitations, the probability of human error decreases. This implies that many of the errors that humans make are due to the fact that technology is not adapted to humans and thus the humans are not, from the designers' point of view, to be blamed. An error that occurs while using a device is nowadays called 'use error' instead of 'human error' or 'user error.' The reason for this is to point out that use error can be the result of a mismatch between the user, the device, the task or the environment.

There are also many other factors than the actual design of the technology that affects whether a person is performing correct or incorrect actions. In addition to the individual characteristics of the human and the machine, the organization in which the human being works plays a major role. Much research has been conducted to understand these mechanisms, such as safety culture and resilience engineering. The latter tries to focus on the factors that make a system able to handle known and unknown events without the occurrence of incidents and accidents. Ultimately, however, it is always in the individual decisions and in individual actions that humans act correctly or incorrectly, which mean that it is of great interest to study and analyse human actions and use errors.

Areas that are in focus for use errors are those spheres of technology that cause spectacular accidents such as flight and nuclear disasters. However, there are other areas that every year also kill and harm many more people, and the main one is medical technology. Research has shown that bad design is the origin of many of the errors that occur during the use of medical devices. These use errors might result in a patient being harmed or even killed.

Having said that, a use error does not need to directly kill or harm in order for it to be worth studying. In many work tasks humans use their bodies in poor ergonomic positions and suffer as a result. This has prompted the definition of the area of physical ergonomics, which is defined by IEA as: "*Physical ergonomics is concerned with human anatomical, anthropometric, physiological and biomechanical characteristics as they relate to physical activity*." If the human does not work in good postures musculoskeletal disorders (MSDs) can arise, which can affect the body's muscles, joints, tendons, ligaments and nerves. These poor body postures can also be regarded as a use error.

Furthermore, you do not need to physically hurt a person to make use error relevant. If you accidentally select the wrong recipients for a SMS message that too can have consequences. Designing a good artefact is of interest to try to reduce use errors and help the human perform the task correct.

But even if handled correctly there could be problems in the interaction between human and technical components. This interaction can be extensive and ineffective. One result may be that not all functions of the machine can be used and/or that the user may be negatively affected emotionally. The latter can affect the user’s ability to perform the work. Even these problems can be said to be a mismatch between human and machine.

The main approach to counteract mismatch is to adapt the design of the machine to the human, the task and the environment. However, in order to know if the design is good and without mismatch, it needs to be evaluated. For the developers to be able to counteract mismatches. The mismatches need to be identified and the causes made visible; if you do not know the error, you cannot attend to it. A classic way of doing this is accident reporting and investigation. However, this approach has the disadvantage that it is reactive and that something must happen before the mismatch is detected, and in many products there is no mechanism for reporting of mismatches.

Another way to detect errors is to perform testing with real devices or high-fidelity prototypes (usability testing) to discover possible mismatches. The problem with this approach is that it only discovers part of all possible mismatches, and an actual device is needed for the evaluation.

To counteract mismatches a more proactive and analytical approach is needed in order to identify potential use errors, investigating and attending them before any real accident occurs. Today such methods exist in the field of human factors engineering but further development is needed to better adapt them to the development processes. There is also a need to combine evaluation of cognitive and physical ergonomics to undertake a comprehensive analysis of human-machine interaction.

2.2 Problem description

The research problem addressed in this thesis is how the interaction between user and artefact can be better analysed in a systematic and structured way so as to be able to detect and identify mismatches in the interaction, i.e. usability problems, use errors and ergonomic errors. The central question has been how the methods can be improved in these areas. This thesis provides examples of a useful methodology for this.

2.3 Purpose and goal

The work presented in this thesis focuses on analytical investigation of mismatches in the interaction between user and artefact. The main idea for the work is that a holistic view is necessary in order to understand mismatches, particularly for physical and cognitive actions.

The purpose of the work is to provide improved support for developers in handling and preventing these mismatches early in the product development process.

The goal of the work was to develop an improved Human Factors Engineering methodology for predicting, identifying and presenting presumptive mismatches in the interaction between user and artefact, based on existing methods.

- Prediction – investigating when, where and how presumptive mismatches exist
- Identification – determining the type and properties of the predicted mismatches
- Presentation – describing the identified mismatches in a manner that facilitates counteractive measures.

2.4 Delimitations

The methodology and method development described in this thesis evolved in evaluations performed in actual projects in industry and academic. The purpose of these specific projects has been to improve the design of various items of equipment, not to improve methods. This has entailed two main delimitations for the work.

The first is that the choice of methods which have been refined is not based on any search for an optimal selection of methods, i.e. there has not been any mapping and evaluation of existing methods. Instead, the choice has been made on the basis of requirements concerning the methods used in actual product development projects and the selection of well-known methods which were regarded as suiting the evaluations in the respective projects.

The second delimitation is that no empirical validation has been conducted in order to demonstrate that the resulting methods are better than the original methods. The full empirical validity of the methods will therefore not be treated in the thesis.

2.5 Reading guidelines and outline of thesis

This thesis was written to give an overall picture of the background to the proposed methodology as well as a description of that methodology, which means that some parts may be perceived as repetition of what is described in the accompanying articles. Often, however, there are minor adjustments in the final presentation of the CCPE methodology presented in the thesis with regard to such parts as presented in the articles. Instead, to account for the differences, I have chosen to present the final methodology in its entirety in the thesis, which makes it easier for readers to obtain an accurate picture of the methodology and its parts. My choice for presenting the research results means that this thesis does not follow the usual structure of a thesis with appended articles. My structure is more like a technical report. The format of my thesis makes it easier for a reader who wants to apply the methodology in real world cases to absorb and use the information.

The following are the chapters in the thesis:

3 Description of procedure
This chapter describes the general approach used in the methodology and method development, and how this development was carried out. The chapter presents the various parts of the method development and their interrelationships.

4 Theoretical framework
This chapter describes the theoretical framework supporting the development of the methodology. The chapter consists of four parts each with a different focus: Human, artefact and activity, Engineering and research areas, Mismatch in interaction and Interaction evaluation. All these parts end with a summary in the form of a requirement specification of what the theory means for the development of the methodology. Compilation of the requirements occurred in parallel with development of the methodology, even though the requirements are presented earlier in this case.

5 Results
In this chapter the developed methodology is presented. It is followed by a description of the development of the methods and the methodology. The chapter ends by showing how the methodology was used in different projects.

6 Assessment
The chapter consists of two parts. In the first part the developed methodology is assessed by verification, validation and reflection. This is followed by a review of the methodology's relation to other methods and areas.

7 Discussion
The chapter begins with a discussion of the thesis work in relation to important factors relating to the development of new products and technological systems. Thereafter follows a discussion of the approach for method development and research.

8 Conclusions
The chapter describes the general conclusions and possible future work.

2.6 Abbreviations

Listed below are abbreviations used in the thesis.

AEA	Action Error Analysis	(method)
ACTA	Applied Cognitive Task Analysis	(method)
CCPE	Combined Cognitive and Physical Evaluation	(methodology)
CW	Cognitive Walkthrough	(method)
ECW	Enhanced Cognitive Walkthrough	(method)
EEMUA	Engineering Equipment and Materials Users Association	(standardisation organ)
EMG	Electromyography	(method)
ETA	Event Tree Analysis	(method)
FDA	US Food and Drug Administration	(agency)
FMEA	Failure Mode and Effect Analysis	(method)
FTA	Fault Tree Analysis	(method)
GEMS	Generic Error Modelling System	(theory)
GTS	Generic Task Specification	(method)
HAZOP	Hazard And Operability Study	(method)
HTA	Hierarchical Task Analysis	(method)
HE	Heuristic Evaluation	(method)
HEART	Human Error Assessment and Reduction Technique	(method)
HEI	Human Error Identification	(methodology)
HEP	Human Error Probabilities	(methodology)
HFE	Human Factors Engineering	(field of research)
HFS	Human Factors Science	(field of research)
HFI	Human Factors Integration	(field of research)
HRA	Human Reliability Assessment	(field of research)
ISO	International Organization for Standardization	(standardisation organ)
IEC	International Electrotechnical Commission	(standardisation organ)
JHEDI	Justification of Human Error Data Information	(method)
LA	Link Analysis	(method)
MSD	Musculoskeletal disorders	(illness)
Nasa-TLX	Nasa - Task Load Index	(method)
NUREG	US Nuclear Regulatory Commission Regulation	(agency)
PEEA	Predictive Ergonomic Error Analysis	(method)
PUEA	Predictive Use Error Analysis	(method)
PHEA	Predictive Human Error Analysis	(method)
PSF	Performance Shaping Factors	(theory)
PUEA	Predictive User Error Analysis	(method)
RULA	Rapid Upper Limb Assessment	(method)
REBA	Rapid Entire Body Assessment	(method)
SHERPA	Systematic Human Error Reduction Prediction Approach	(method)
SWAT	Subjective Workload Assessment Technique	method)
THERP	Technique for Human Error Reduction	(method)
UT	Usability Test	(method)
UTP	User-Technical Process	(method)

3 Description of procedure

This chapter describes the general approach used in the methodology and method development, and how this development was carried out. The chapter presents the various parts of the method development and their interrelationships.

3.1 Research approach

The approach for the research in this thesis can be described at different abstraction levels, as problem-driven research with similarities with actions research and the hypothetico-deductive model.

3.1.1 Problem-driven research

All research builds upon previous work and the scientific progress spotlighted four components - theory, data, problems and methodology (Learner and Phillips, 1993). As the research described in this thesis has been based on problems experienced when using existing methods in real-world evaluations, these method-related problems have been the main driving force in my work. The approach in this thesis is therefore problem-driven research. The research has not been theory-driven, i.e. starting from an existing theory and then testing or extending until a result is reached. Neither has the research has been method-driven. Although methods were in focus in this research, the research actually focused on use of the methods and not with the methods themselves.

Problem-driven research generally has two aims, firstly to solve the current problem, and secondly to use the lessons learned to deepen science (Learner and Phillips, 1993). For this thesis, it means that the initial focus was on solving the problems and then on contributing to knowledge-building with the help of the solution. Due to this, it has not been practical to use genomic research questions to guide this work. Instead, the purpose and aim have been the basis for driving the work forward. Similarly, the problems that have driven this research have shifted during the process; solving one problem has uncovered the next problem and thus further method improvements have been possible.

Because of this approach, the research has been problem-driven, i.e. the research has not been based on theory to identify where potential improvements and solutions could be found. There has been no study of 'state-of-the-art' solutions in the course of this research; instead, problem identification has been done entirely with existing methods. Of course, theories of methods played a major role in research into growth, but not in a systematic way that would be the case if the research has been theory-driven. The theory that was used in this research is described in Chapter 4, and even if it comes before the methodology of Section 5.1, the emergence of the theoretical framework has gone hand in hand with the method development. This process is described below.

3.1.2 Action research

The general approach for method development which is described in this thesis has close similarities with Action Research (Reason and Bradbury, 2001). Action Research is based on a combination of action and research. The term 'action' indicates that something is performed or tested, while 'research' means that systematic work and a relationship with theory yield new knowledge (Rönnerman, 2004). The essential idea of action research is described by

Reason and Bradbury (2001, p 2) as follows: *"So action research is about working towards practical outcomes, and also about creating new forms of understanding, since action without reflections and understanding, just as theory without action, is meaningless"*. Action research is consequently a way of uniting theory with practice.

Action research is a 'bottom-up' process where the persons who carry out the action are themselves agents of change (Rönnerman, 2004). The process is described by Dick (2003): *"Action research is a flexible spiral process which allows action (change, improvement) and research (understanding, knowledge) to be achieved at the same time. The understanding allows more informed change and at the same time is informed by that change.*" This spiral process is shown in Figure 3.1. The action researcher performs an action in order to reach a goal. Thereafter the result of the action is compared with the goal, and proposals for change are introduced so as to get closer to the goal in the next turn of the spiral. These steps are repeated and the result is examined critically, rolling the process further as it gradually converges on the goal.

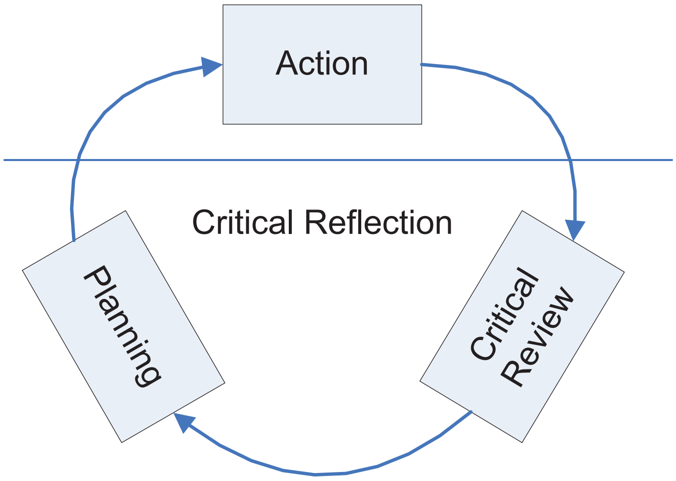


Figure 3.1 The action research spiral (after Dick, 2003)

3.1.3 Hypothetico-deductive

From a more general science perspective, the research can be described as a hypothetical-deductive method (Hansson, 2011, Sohlberg and Sohlberg, 2009). This is due to the fact that the proposed methodology is based on reasoning and does not follow from empirical studies. However, the developed methodology has also been tested empirically. According to Birkler (2008) the hypothetical-deductive method is composed of five stages: hypothesis, deduction, empirical consequence, induction and conclusion. First, hypotheses are presented and through deduction empirical implications of the hypothesis are derived. The consequence is then tested empirically and via induction, conclusions are drawn that confirm or negate the hypothesis, as shown in Figure 3.2.

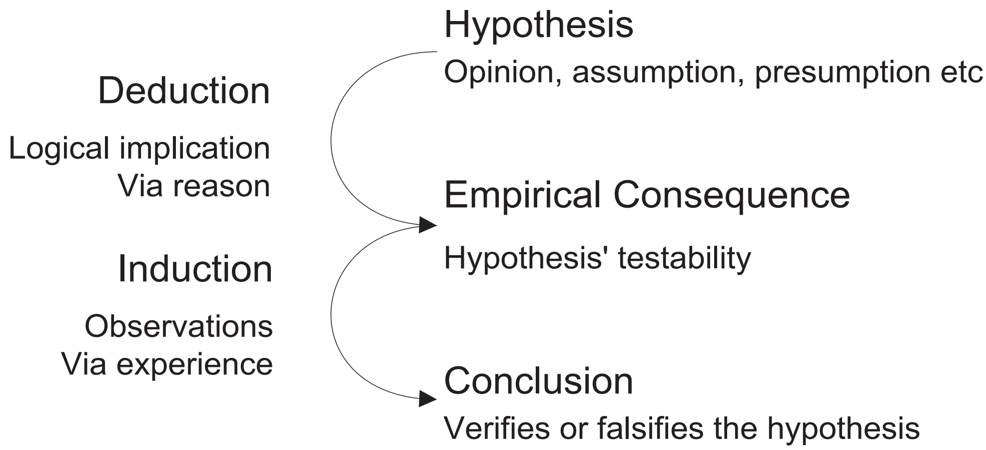


Figure 3.2 Hypothetico-deductive method (after Birkler, 2008)

When figure 3.2 is applied to the development of the methodology presented in this thesis, the process is similar to the hypothetical-deductive method as follows:

- There is an opportunity to develop a better methodology (Hypothesis)
- Using common sense and reasoning develop a better methodology (Deduction)
- The methodology is used in projects (Empirical Impact)
- Evaluation of the methodology based on empirical results (Induction)
- Assessment of the methodology (Conclusion)

Matching between the method development and the hypothetical-deductive method is not entirely consistent, but sufficiently similar to be useful as a framework and provide an important implication. The theory about the hypothetical-deductive method indicates that it is difficult to determine whether a hypothesis is true without the hypothesis becoming more or less likely based on the conclusion (Hansson, 2011). It is therefore impossible to completely prove that the hypothesis is correct, but the assessment of the methodology has to be based on whether the results of empirical use have made it more likely for the methodology to work better. The thesis then continues to describe in greater detail the implementation of the research process and the development of the methods and methodology.

3.2 Research process

The research presented in this thesis was conducted during application projects concerned with the evaluation and design of man-machine systems in various domains, primarily the medical equipment domain. In these projects, different HFE methods have been employed to evaluate and redesign the design of the devices. The objective of the projects was not methodology development, but where a need emerged for methods and method development, this has been undertaken as part of the project. The developed methods then evolved into a methodology.

3.2.1 Development methods

The research process mainly consists of method development, and a number of activities took place during the method development. The development of the methods can be described as a spiral process, Figure 3.3. The spiral process contains three steps: (1) Methods were used in projects. (2) During and after the methods' application in development work, certain problems and deficiencies were found (reflection on the method development). (3) These deficiencies led to proposals for changes/improvements. The method development then began a new cyclical process where (1) the method was used but now in its altered form, (2) the modified method was subsequently evaluated, and (3) new proposals for additional changes were made.

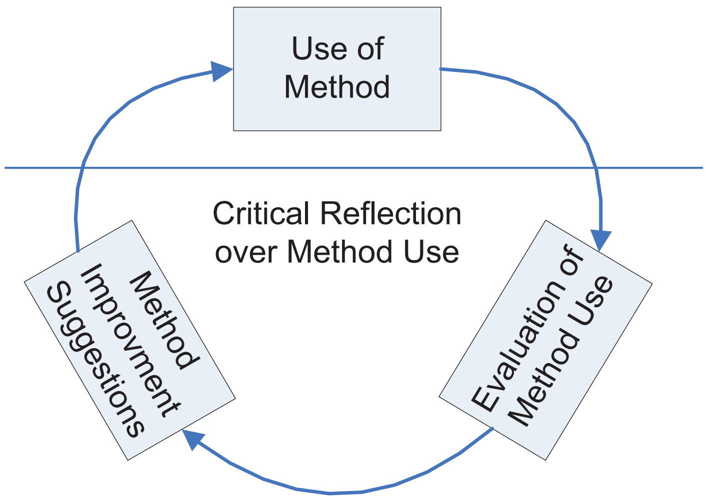


Figure 3.3 The general process of method development that has been used in this thesis. Adapted from Dick (2003)

During the cyclical process aims and prerequisites for the refined method were also established, requirements were noted and methods for further development chosen. The process took place both within and between the projects in the method development, i.e. the methods have been modified both before and during the projects. Thus, the development has primarily been a 'bottom-up' process, in which the deficiencies detected by the methods during use have served as a basis for the improvements. The main source of the deficiencies in the methods has been the results from other Human Factors Engineering methods – such as heuristic evaluation, usability tests, and interviews with and observations of users.

The spiral processes shown have been governed partly by the established requirements on the methods, and partly by the deficiencies found in the methods. Often the requirements and the deficiencies emerge simultaneously.

Continuously throughout the method development, there has been a critical reflection on whether the spiral process leads towards the aim of the work – a method for predicting identifying and presenting presumed use errors and usability problems, primarily within medical equipment. Further, all activities in the method development have taken place during the previously described projects, and this will be more exhaustively described in chapter 10.

This process was then repeated for each of the developed methods. The development of the method is more or less like the work in a human factors engineering process (Andersson et al., 2011).

3.2.2 Development methodology

When the development process started, there was no plan to develop a methodology, nor any plan to develop more methods. The methods previously described have instead emerged from the needs of the various projects. The individual methods and the year of their first releases are listed below.

- Enhanced Cognitive Walkthrough (ECW) 2002
- Alarm - Enhanced Cognitive Walkthrough 2003
- Predictive User Error Analysis (PUEA) 2003
- Predictive Ergonomic Error Analysis (PEEA) 2004
- Generic Task Specification (GTS) 2005

It may seem strange that a methodology shows up by itself, but all the methods are based on the same systematic and structured approach (as presented in the next theory chapter). Although they have been produced as separate units, this approach has always been the basis, even if it has not been fully pronounced. The methods have thus built on each other but they are designed to take into account different aspects, and to work together. Finally, the methods have much in common and are so interwoven that the boundaries between them have started to blur. There is a methodical and systematic structure that can be tailored in many ways on the basis of what should be studied. Thus, a methodology has gradually emerged with the development of more methods as well as through further development of already developed methods.

3.2.3 Evaluation of methods and methodology

In the development of the methods, their evaluation is a central part of the work and this is clearly shown in the spiral of action research (Figures 3.1 and 3.3). The methods have also been tested in many different projects and applications, as explained in Chapter 5.3. This has also been a form of evaluation of the methodology and methods.

There has also been a more formal effort to evaluate the methods and methodology and the results of the evaluation are presented in Chapter 6.2. The evaluation was performed as follows:

- PEEA was tested against empirical studies in two cases (Appendix A and B)
- The entire methodology was tested in three cases against empirical studies (Appendix C and D)
- PUEA was tested against empirical studies in one case (Appendix E)
- Evaluation of ECW and PUEA undertaken to interview users in industry and students (Paper V)

- Collection of ECW theoretical comparison to other versions of cognitive walkthrough (Mahatody et al., 2010)
- External test and evaluation of ECW and PUEA with other evaluation methods (Tancredi et al., 2012)
- Statements from two experienced users, and by the author, about the advantages and disadvantages of the methodology

3.2.4 Presentation of methods and methodology

During the course of development, the methods and methodology were also presented at several scientific conferences (Table 3.1). The aim has been to disseminate the methods and methodology to the scientific community and also to receive comments that can help improve the methods and methodology.

Table 3.1 Presentation of the CCPE at scientific conferences

Conference	Year	Method	Paper
Nordic Ergonomic Society	2003	ECW	Paper IX
IFAC Symposium on Automated Systems Based on Human Skill	2003	ECW	Paper X
Nordic Ergonomic Society	2004	ECW	Paper XI
International Ergonomics Association	2006	ECW	Paper I
Nordic Ergonomic Society	2006	PEEA	Paper VI
Nordic Ergonomic Society	2007	Alarm-ECW	Paper XII
Applied Human Factors and Ergonomics	2008	GTS	Paper VII
Society for Risk Analysis - Europe Annual Meeting	2009	ECW, PUEA	Paper XIII
Applied Human Factors and Ergonomics	2010	CCPE	Paper XIV
NordDesign	2010	ECW, PUEA	Paper XV
Asia Pacific Conference on Sports Technology	2011	ECW, PUEA	Paper XVI

4 Theoretical framework

This chapter describes the theoretical framework supporting the development of the methodology. The chapter consists of four parts each with a different focus: (4.1) Human, artefact and activity, (4.2) Engineering and research areas, (4.3) Mismatch in interaction and (4.4) Interaction evaluation. All these parts end with a summary in the form of a requirement specification of what the theory means for the development of the methodology. Compilation of the requirements occurred in parallel with development of the methodology, even though the requirements are presented earlier in this case.

4.1 Human, artefact and activity

The first part of the theoretical framework discusses the interaction between human and machine. The framework begins with systems theory, since it is the overall model for methodology development, and will then gradually proceed to explain how interaction between human and machine is described theoretically.

4.1.1 System theory

Systems theory is an umbrella term for the theories used to describe how parts together form a system with different characteristics than those of the individual parts alone (Skyttner, 2005, Flood and Carson, 1993). Systems theory supposes that it is impossible to understand a whole by breaking it down into smaller parts and then studying the smaller parts separately. You cannot for instance explain how a human being works by simply studying the cells. Systems theory thus focuses on arrangements and the relationship between the relevant parts that unite them into one whole. This approach originates from biology, but is now widely used within the natural sciences, engineering, psychology and social sciences. The systems theory approach allows it to be used to describe and understand the complexity of systems.

In systems theory it is the system itself that is the central concept. A system consists of several communicating elements of an organized whole. This organized whole can be tangible, such as a machine, or abstract, such as rules. An element can equally be something physical, social, or abstract. The communication between elements may consist of the transfer of matter, information or energy / power. Each system has also a system boundary that defines what belongs to the system and what does not. The boundary can be physical such as for a machine or an animal, social as in a herd, or abstract as in a set of rules. A system itself may be part of a larger system and likewise, an element can be a system in itself and comprise other elements. The purpose of each system is to process energy, information or matter into a result to be used within the system, outside the system (the environment), or both.

What defines a system? To be able to say that a system exists, the following characteristics usually need to exist for the system (Flood and Carson, 1993, Skyttner, 2005):

- The whole is greater than the sum of the parts
- The whole defines the nature of the parts
- The parts cannot be understood by looking at the big picture
- The parts are dynamically related and dependent

A fundamental concept within systems theory is holism, i.e. that the whole is greater than the sum of its component parts. This means that a system as a whole behaves differently than the elements of the system and that the elements individually cannot do what the system can. If this holism does not exist, there is no system.

An element is considered to belong to a system if it is within the system boundaries, and the element has a relationship (communicates) with other elements inside the system boundary. What is important for systems theory is therefore to describe this specific flow of energy, matter and information within and across the system boundary, as well as to describe the various elements' relationships and how they affect each other.

4.1.2 Activity theory

One systems theory which is suitable as a starting point for human and machine interaction is activity theory. Activity theory tries to explain how individuals interact with their environment and with artefacts. The basic idea for the use of activity theory is that we cannot study machines as individual subjects, instead we must study how they meditate use.

There are many variations of activity theory since it first occurred in Russia in the early 1900s, for example Engeström (1987). The theory described here is based on Karlsson's (1996) description and use of activity theory. Activity theory has extensive theory formation and the section below describes a suitable sample.

Five basic terms within the theory are activity, object, subject, mediator and context. An activity is defined as a human process directed towards an object. The object describes the goal, problem, motive, and so on that the activity intends to influence. The subject is the person performing the activity and the mediator is the tool (abstract or tangible) with which the activity is performed. The context is the situation and circumstance within which the activity occurs. To understand the activity the relationship between object, subject, mediator and context needs to be understood. The relationship is usually visualized by a triangle (Figure 4.1)

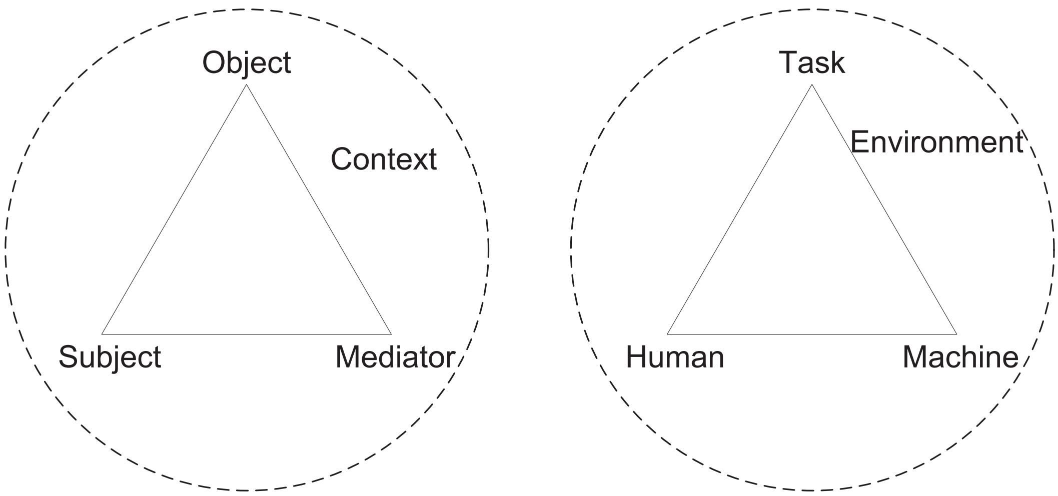


Figure 4.1 The relationship between the elements in the activity theory described by triangles

When activity theory is applied to human-machine interaction the actual use becomes the activity. The human is the subject and the object is the task to be performed to achieve the goals. The machine is the mediator that the human uses to perform the task and the context is the environment in which the use occurs. Activity theory therefore shows that the relationship between human, machine, task and environment is the central subject to study. The theory also shows that the machine is there because the human cannot perform the task without a mediator, i.e. the human cannot handle it on its own.

4.1.3 Human-machine system

Activity theory provides a good foundation but a more detailed system model is needed to describe the interaction between human and machine, i.e. a human-machine system. A human-machine system consists of humans and machines that interact in a specific environment to perform specified tasks to achieve the system goals. The goal for each human-machine system is to always in some way transform information, energy / power and / or matter. To achieve this goal, human-machine systems often need to perform one or more tasks in the most optimal manner possible.

The tasks are performed in interaction between the human and the machine, where the human and the machine are each individual systems, but together they create a larger system with other characteristics than the parts possess separately. The actual interplay between the human and the machine is described as interaction.

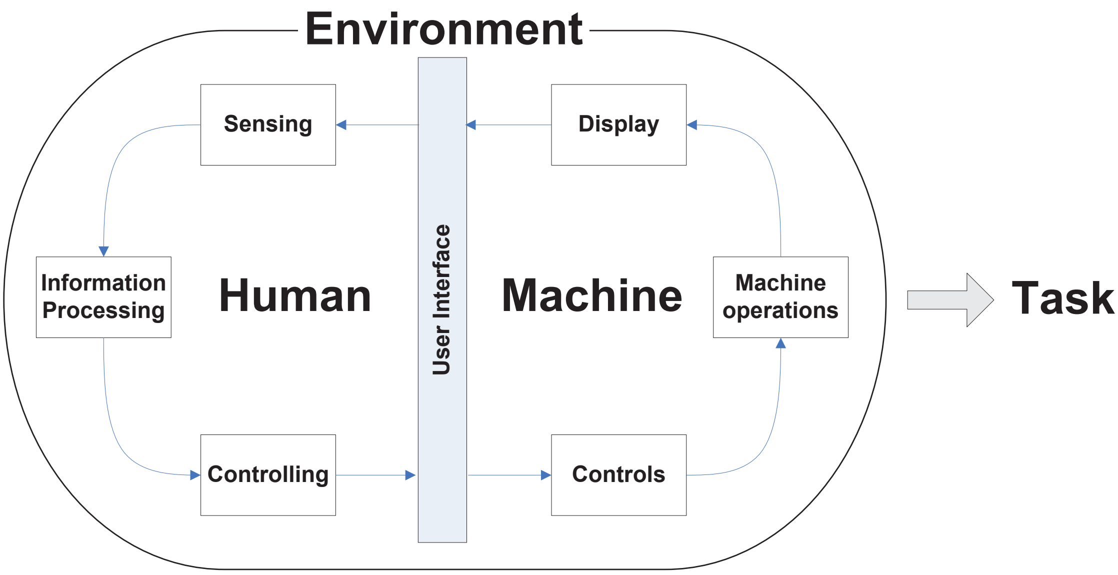


Figure 4.2 Model of a human-machine system from Chapanis (1965)

The exchange of information occurs at the interface between the human and the machine, in which the human directs or controls the machine, which is called the user interface. Introducing the user interface in the human-machine system will produce a model (Figure 4.2) of a cyclical process in which information is exchanged between the human and the machine under the influence of the context (user environment). The machine displays the information that the human intercepts through perception. The information is then further processed and the human decides on and performs an action. The action affects the machine's processes, which in turn changes the information available to the human and so on in the cyclical process.

4.1.4 The Interaction

The direct interplay between the human and the machine is thus described as an interaction and progress according to the cyclical process described by the model for human-machine system. At the interface between man and machine, the user interface, the exchange of information, takes place. However, the description of the interaction needs to be developed and become more detailed to be more useful. A more detailed way to describe the interaction is with Norman's "Seven Stages of Action" (Norman, 2002). According to Norman, interaction can be divided into seven steps.

1. Forming the goal
2. An intention to act so as to achieve the goal
3. The actual sequence of actions that we plan to undertake
4. The physical execution of that action sequence
5. Perceiving the state of the world
6. Interpreting the perceptions according to our expectations
7. Evaluation of the interpretations with what we expected to happen

The cyclical process of the human-machine system clearly shows how the human and the machine work together to accomplish the task. The description of the interaction is at a detailed level where the whole is divided into individual decisions and actions. However, there are also human mental activities at other levels, from culture to operation.

It is important to achieve good interplay between the human and the machine so that the system goals can be achieved and be useful. This means that the machine must be adequate and provide the right functionality. However, this can primarily be regarded as a relationship between the human and the machine. Some of the terms used to describe interaction are:

- Affordance
- Usefulness and usability
- Accident, risk and safety

Affordance

To make an interaction even possible, the user must physically be able to interpret information and perform the action. This quality can be described by affordance. The basis of affordance was designed by Gibson (1979, p 127) who writes: *"The affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment…"*

Affordance is thereby a possible action relative to an individual, regardless of whether the individual is aware of the action or not. If affordance does not exist, there is no possibility for the individual to perform the action. For interaction, affordance shows that the machine needs to be designed according to the human being's physical abilities. To be able to carry out the actions we have senses that allow us to take in information, cognition to process the information and finally the musculoskeletal system in order to act. What is needed is knowledge about human anthropometry and physiology as well as senses and perception. As a result of affordance the user also needs to be able to perform the action without being harmed.

Usefulness and usability

Since affordance concerns the user's knowledge about performing the action, the next step deals with whether the user understands the action and if the action is relevant. The terms usefulness and usability are employed for this.

According to Grudin (1992), usefulness is a measure of how well a technical system can achieve a desired goal. Usefulness can then be divided into two aspects: utility and usability. Utility depends on whether the functionality of the technical system can perform what is required, while usability depends on how well the user can use that functionality. In the case of a drilling machine, utility refers to the drill’s capacity to make holes, whereas usability describes how well the user can handle the drill while it bores holes.

In medical equipment for example, it is less complicated to employ Grudin’s (1992) distinction between usability and utility. Utility for medical equipment becomes the medical function of the equipment, such as how well a blood-pressure gauge measures blood pressure, or how well a scalpel cuts. Usability becomes how well the medical personnel can use the equipment’s medical functionality, such as whether a nurse can understand what the blood-pressure gauge displays or whether the scalpel fits in a surgeon’s hand. Usefulness responds to the overall question of how well the medical personnel can treat patients with the medical equipment. Usability in medical equipment, therefore, does not concern the medical aspect.

So, the term usability describes how well the human-machine interaction works. The way this term should be defined more precisely has, however, long been debated. A review and description of this debate is given by Liu (2004) among others.

A refinement of the concept of usability was offered by Jakob Nielsen (1993). Nielsen follows Grudin’s (1992) distinction that functionality (utility) is not a part of usability. Moreover, Nielsen maintains that usability has multiple components and is associated with five usability attributes:

- *Learnability:* The system should be easy to learn so that the user can rapidly start getting some work done with the system
- *Efficiency:* The system should be efficient to use, so that once the user has learned the system, a high level of productivity is possible
- *Memorability:* The system should be easy to remember, so that the casual user is able to return to the system after a period of not having used it, without having to learn everything all over again
- *Errors:* The system should have a low user error rate, so that users make few errors during use of the system, and so that if they do make errors they can easily recover from them. Further, catastrophic errors must not occur
- *Satisfaction:* The system should be pleasant to use, so that users are subjectively satisfied when using it; they like it

Also occurring are definitions of usability not following the way Grundin distinguishes usability and utility. In the definition from ISO 9241-11:1998 (ISO, 1998), utility is included in usability. ISO defines usability as *“The effectiveness, efficiency, and satisfaction with which specified users achieve specified goals in particular environments*."
The components are explained as follows:

- *Effectiveness:* The accuracy and completeness with which specified users can achieve specified goals in particular environments
- *Efficiency:* The resources expended in relation to the accuracy and completeness of goals achieved
- *Satisfaction:* The comfort and the acceptability of the work system to its users and other people affected by its use

Here the artefact's capacity to perform the intended task is included in the concept of usability. I.e. the total human-machine system's performance is part of the system's usability. Jordan, (1998) begins from the ISO definition of usability and makes a different division of usability into five components (Jordan, 1998). These components are designed more for numerical measurement in evaluations, whereas Nielsen's attributes can be characterised as product properties.

- *Guessability:* The effectiveness, efficiency and satisfaction with which specified users can complete specified tasks with a particular product for the first time
- *Learnability:* The effectiveness, efficiency and satisfaction with which specified users can achieve a competent level of performance on specified tasks with a product, having already completed those tasks once previously.
- *Experienced user performance:* The effectiveness, efficiency and satisfaction with which specified experienced users can achieve specified tasks with a particular product
- *System potential:* The optimum level of effectiveness, efficiency and satisfaction with which it would be possible to complete specified tasks with a product
- *Re-usability:* The effectiveness, efficiency and satisfaction with which specified users can complete specified tasks with a particular product after a comparatively long period away from these tasks

However, all these definitions imply that usability is an emerging property of the artefact in relation to the user, the goal of the task, and the context (environment). The definitions of usability have in common that usability is a function of "human performance".

Accident, risk and safety

Besides being able to understand and manage the machine, it is important that the user does not get harmed during interaction. The human is also influenced by the environment, both physically and socially, during the interaction. Accidents mean that people are harmed by the machine, the human can be exposed to violence, electricity or toxic substances.

In order to handle the situations that may arise during the interaction, risk and safety form a good framework. When you hear words such as accident, hazard, risk and safety, they are often related to physical harm to humans or the environment. Having said that, the terms are also useful in events, which are not harmful in this way. An accident or undesired event is an unexpected event with undesired results, Table 4.1. The word "unexpected" however does not mean unpredictable, as many accidents or unwanted events can be foreseen.

Table 4.1 Description of an accident or unwanted event from Hollnagel (2004)

	Undesired result	**Desired result**
Unexpected event	Accident or undesired event	Luck
Expected event	Bad luck	Effectiveness

An accident or unwanted event means that a hazard occurs. All work with risk and safety aims at avoiding harm to humans, animals, artefacts and the environment. Central to such work is the concept of hazard. Usually hazard is defined as a potential source of harm (ISO, 2000b, Dhillon, 2003, Kolluro and Brooks, 1996).

A hazard can be something physical, like a toxic material, or something abstract, like a text message sent to the wrong person. The common denominator is that a hazard may lead to some sort of injury to someone or something and that a danger is more or less always present. As long as there is toxic material left, the hazard remains and in the same way there is always a hazard when someone writes a text message that it may be sent to the wrong person.

A hazard gives rise to a risk through the harm it can create, leading to a hazardous situation. The risk itself depends on the characteristics of this harm. The factors usually taken into account are the probability that a hazard will lead to harm and the injury that may impact humans, animals, artefacts or the environment when these are exposed to the hazard. Risk is regarded as a combination of probability and consequence (ISO, 2000b, Dhillon, 2003, Kolluro and Brooks, 1996). Figure 4.3 shows a model of the relationships concerning the term "hazard".

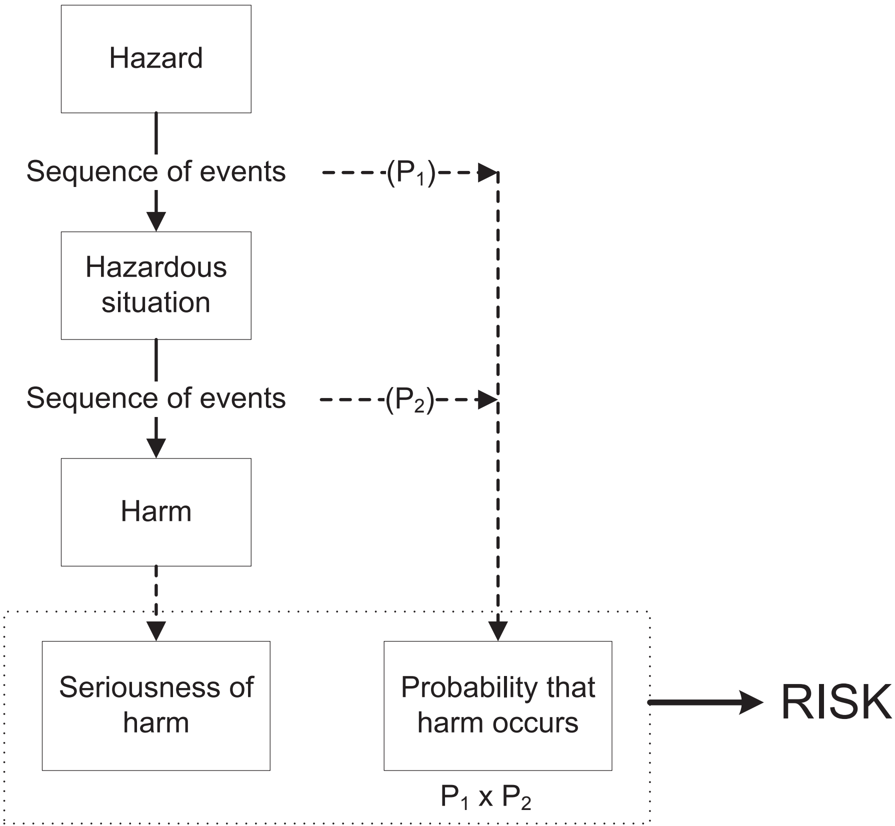


Figure 4.3 Hazard, harm and probability in the risk model from (ISO, 2000b)

Consequently there are inherent dangers in the machine. Via a sequence of events, with a certain probability, the hazard is released and the perilous situation arises, e.g. the toxic material is released or the text message is sent to the wrong person. There is then also a probability that the perilous situation actually via a sequence of events leads to harm, for example, that a person is harmed by the toxic emission or the text message that was sent to the wrong person leads to something more than the recipient's surprise.

The opposite of risk is safety and ISO 14971 (2000b) defines safety as *"freedom from unacceptable risk"*. Safety is thus determined by the level of risk acceptable in the specific human-machine system and it must be determined to give the term “safety” a meaning. Regarding something as safe is thereby a very relative measure.

Safety should be a self-evident aspect of the definition of usability. An artefact that harms people or the environment ought not to be regarded as having good usability. One way in which technology can harm people, but which has not been considered earlier in this thesis, is through stress injuries to the body, i.e. poor physical ergonomics. Products with poor physical ergonomics for the user should therefore not be able to qualify as having high usability, no matter how easy, effective and satisfying the products are to use – for these products will harm the user.

4.1.5 Input to methodology development

The goal when designing machines is that they should be able to perform their tasks and achieve their system goals, but also provide high usability, high safety and no harm on the human. These parts overlap and they can be divided according to the following points:

- **Adequacy** - does the machine have the right functions to achieve the system goals?
- **Affordance** – is the machine adapted to the human’s physical condition?
- **Effectiveness** – will the human be able to perform the interaction with the machine? Is the machine physically and cognitively adapted to the human?
- **Efficiency** – is the interaction done using appropriate resources as regarding time, steps in interaction, physical and mental workload?
- **User safety** – does the machine harm the user in short term or long term use?
- **Usage safety** – is the interaction undertaken without machines, the environment, the economy or other people than the user being put at risk?
- **Satisfaction** – is the human satisfied, without discomfort, at an accurate level of stress, before, during and after the interaction?

In order to analyse mismatch the methodology must be based on the correct way to study why this was not the case. The interaction is always done in a number of individual steps, so to examine mismatch every step needs to be examined individually. As shown in previous chapters, what affects the interaction is a variety of aspects, making a system perspective necessary. When analysis of the individual interactions takes place the big picture also needs to be taken into account.

Adequacy cannot be seen as a direct aspect of the interaction between human and machine so it does not have to be taken into account by the methodology in the analysis. Mismatch between the machine and the task and mismatch between the human and the task is not analyzed.

The human's subjective experiences such as satisfaction are interesting and relevant in itself, but more an effect of mismatch than what affects it. In the same way user safety (physiological and psychological effects on users) is more of an effect of interaction than something affecting mismatch. Usage safety is also more of an effect than that it is directly involved in the mismatch. Efficiency is also beyond the direct individual interaction, but may contribute many negative effects in case it is poor.

What remains then is affordance (physical) and effectiveness (cognitive) that are directly related to mismatch since they affect the individual steps in the interaction. If users cannot perform or understand the steps, it is a mismatch. Use and user safety are directly affected by mismatch as they are direct consequences. The way efficiency and satisfaction are affected is not equally direct, while adequacy is not affected.

Methodology requirements:

- Start from a system perspective
- Study the details of the interaction, i.e. examine each individual step
- Pay attention to both cognitive and physical aspects
- Relate to risks during the interaction

4.2 Engineering and research areas

There are a number of areas in engineering related to the creation of machines and good interaction in which the developed method will work. This thesis describes these areas and how they are relevant to the method development. The areas are:

- Product development
- Risk management
- Ergonomics and human factors
    - Physical ergonomics
    - Cognitive ergonomics
    - Human factors engineering
    - User-centred design

4.2.1 Product development

The overall framework for the methodology development is product development since this is where the function and design of the artefacts is decided. According to the National Encyclopedia (2012) product development is: *"The process that precedes the development of a new product in a company or organization. The process ranges from brainstorming, product design, consideration of environmental tolerance technology and benchmarking to construction, manufacture and marketing."* Product development thus spans from the abstract problem to the tangible solution.

Johannesson et al, 2004	Ullman, 1997	Ulrich and Eppinger, 1995	Bligård, 2011
Prestudy	Identify Needs	Concept devlopment	Needfinding
Product Specification	Develop engineering specifications		Function and Task Design
Concept Generation and Evaluation/Selection of Concept	Develop concepts		Overall Design
		System-Level Design	Detailed Design
Layout Design	Develop Product	Detail Design	Structural Design
Detailed Design			
Prototype Testing		Testing and Refinement	
Adaption for Production		Production Ramp-Up	

Figure 4.4 Four different ways to divide the product development process

Product development takes place during a product development process which consists of different phases. Figure 4.4 illustrates four ways of dividing the process according to Johannesson (Johannesson et al., 2004, Ullman, 2002, Ulrich and Eppinger, 2004, Bligård, 2011).

The overall objective of the method development described in this thesis is to increase the quality of the interaction between human and machine. It is both easier and less costly to change and improve the equipment with regard to usability and safe handling during the development process, than to modify the developed device when it is in the field. Usability and safe handling must therefore be attended to during the development process.

It is obviously during product development that the machine in the human-machine system is constructed, which makes it the natural outer boundary of this thesis. Given that the human and the environment are constant it is the design of the machine that determines the quality of the interaction. What is needed during the product development process is thus methodology to counteract the occurrence of mismatch between human and machine. This methodology should be able to function through several product development phases.

4.2.2 Risk management

The next large framework related to the thesis is risk management, since a mismatch may have consequences. The overall work in reducing and controlling risks that is carried out with various systems has been termed Risk Management. The systems in question may be purely technical ones, man-machine systems, or purely human systems, both concrete and abstract. ISO (2000, p 7) defines Risk Management as a *"systematic application of management policies, analyzing, evaluating and controlling risk"*, while Dhillon (2003, p 96)) defines it as *"the total process of risk control and risk assessment*". Figure 4.5 illustrates a Generic Risk Management process.

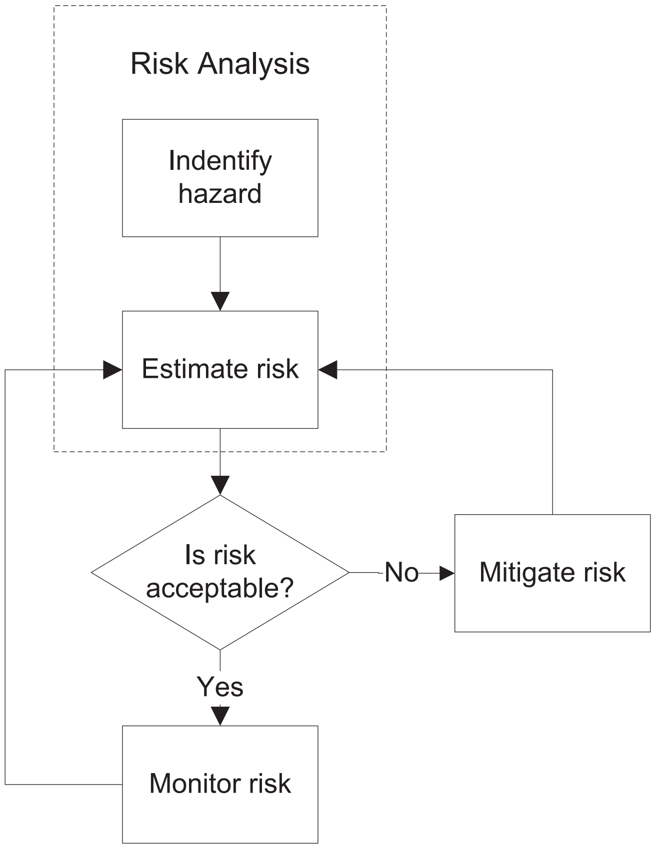


Figure 4.5 Generic Risk Management process with Risk Analysis. Adapted from (Stricoff, 1996) and (ISO, 2000)

A Generic Risk Management process consists of four main activities: identification of hazards, estimation of risk, mitigation of risk, and monitoring of risk (Figure 4.5). The first step is to identify the hazards in the system, which gives rise to risks. Next, a judgment is made of the risk that the hazard causes. If the risk is unacceptable, measures are taken to reduce the identified risk. The remaining risk in the system is then monitored, if necessary a new judgment is made of the risk and further measures are taken. In order to ensure safety in systems with both humans and machines the main issue is to not only try to reduce the cause of human error, but also explore the boundaries of safe performance (Rasmussen and Svedung, 2000).

A central part of risk management is risk analysis, which is usually defined as the effort to identify and evaluate risks in a system - *"Systematic use of available information to identify hazards and evaluate risks"* (ISO, 2000). The essence of risk analysis is therefore to identify those perilous situations/incidents, analyze their cause and examine what consequences may arise and the probability of the event occurring. This information is then used to make a risk assessment to determine if any risk-prevention actions are needed or not. The two steps in figure 4.5, which make up the risk analysis, are hazard identification and risk assessment.

Risk analysis is a natural activity in each part of the development process, but the extent is highly dependent on which machine the process is intended to produce. Risk management with risk analysis is an activity that should be conducted continuously during a machine's life, but the parts that will be discussed concern hazardous work performed during development of the machine. The goal of risk management during the development process is to reduce risks to acceptable levels. What is an acceptable level varies with each type of machine and the level must be determined by project management. To reduce the risk and increase safety, various interventions can be carried out.

4.2.3 Ergonomics and human factors

Even though Product Development and Risk Management are central frameworks for the methodology, the main framework is ergonomics and human factors, as it is specifically the interaction between human and machine that is the central aspect to consider. The International Ergonomic Society (IEA, 2006) defines it more precisely as *"the scientific discipline concerned with the understanding of interactions among humans and other elements of a system, and the profession that applies theory, principles, data and methods to design in order to optimize human well-being and overall system performance."*

As previously described, the methodology is supposed to counteract the problems that mismatch causes and thereby increase human well-being and overall system performance, which corresponds well with the definition of ergonomics and human factors.

Within the field of ergonomics and human factors there are many subgroups and one way to distribute these is described in Figure 4.6. The description assumes that the area can span two axes. The horizontal axis shows the focus from the body through the mental process to the group. Ergonomics is usually divided into physical, cognitive and organizational ergonomics. Physical ergonomics deals with human anatomical, anthropometric, physiological and biomechanical properties in relation to the tasks performed and the human's physical response to physical environmental factors. Cognitive ergonomics concerns human mental processes such as perception, memory, reasoning, and motor response in relation to the tasks performed and the human cognitive response to physical environmental factors. Organizational ergonomics deals with the optimization of socio-technical systems, including their organizational structures, guidelines and processes.

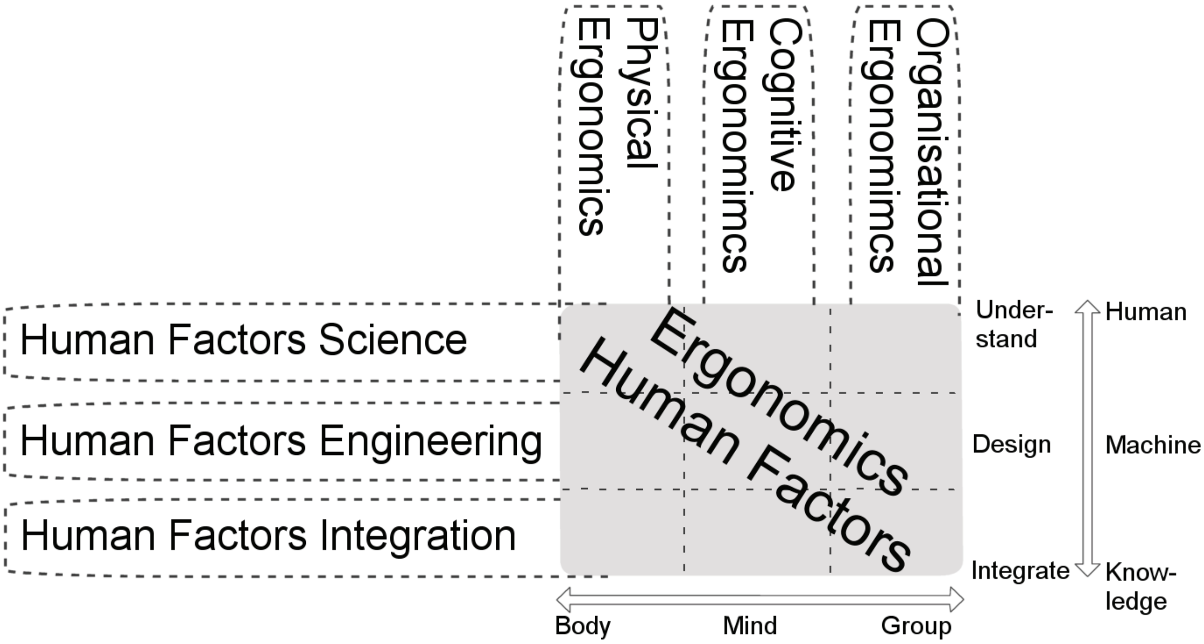


Figure 4.6 Subgroups within ergonomics and human factors

Vertically the focus ranges from understanding the human via the design of the machine to introducing knowledge (to the producer). HF is usually divided into Human Factors Science, Engineering and Integration (figure 4.6). Human Factors Science (HFS) focuses on the human and is the science that deals with understanding and describing human capacities (abilities, limitations, characteristics, behaviour, etc.) in relation to systems of different types (technical, human, etc.). Human Factors Engineering (HFE) is engineering science about how to design and construct machines that are adapted to the human abilities and that are both efficient and productive. Human Factors Integration (HFI) is about how knowledge of Human Factors Science & Engineering will be used and applied in a real context by the producers of machines, organizations, systems, etc.

The description of ergonomics/HF as a subject results in an overlap between the two fields ergonomics and human factors. HFE, HFS and HFI range from the physical through the mental, to the organizational aspect. Ergonomics on the other hand deals with understanding, developing solutions and establishing knowledge. HFE uses knowledge both from the fields of ergonomics and from HFS in the process of designing the machines. The three main sub-groups that are relevant to the thesis are Physical Ergonomics, Cognitive Ergonomics and Human Factors Engineering.

Knowledge and methods from ergonomics and human factors need to be integrated into the product development process along with other types of knowledge necessary (Bligård, 2011) and must be one of the factors that control the development of the machine. The level of ergonomics needs to be considered against the background of fundamental factors such as choice of material and manufacturing methods, but also more machine-specific factors such as functionality and performance.

Physical and cognitive ergonomics

To be able to design and evaluate human-machine interaction, knowledge about the human is needed. A typical division of that knowledge is between the physical and the mental, which is also reflected within ergonomics. IEA (2006) has defined physical ergonomics and cognitive ergonomics according as follows:

"Physical ergonomics is concerned with human anatomical, anthropometric, physiological and biomechanical characteristics as they relate to physical activity. The relevant topics include working postures, materials handling, repetitive movements, work-related musculoskeletal disorders, workplace layout, safety and health."

"Cognitive ergonomics is concerned with mental processes, such as perception, memory, reasoning, and motor response, as they affect interactions among humans and other elements of a system. The relevant topics include mental workload, decision-making, skilled performance, human-computer interaction, human reliability, work stress and training as these may relate to human-system design."

As the methodology will study interaction, both physically and mentally (to prevent mismatch), it will require knowledge about human physiology, biomechanics, and human mental processes.

Human factors engineering

In addition to knowledge about the human, knowledge about how to design the machine in adaption to the human is needed. This is known as Human Factors Engineering. One familiar definition is that of Chapanis (1985): Human Factors Engineering is *"The application of knowledge about human behaviour, abilities, limitations and other characteristics to the design of tools, machines, equipment, devices, systems, tasks, jobs, and environments to achieve productive, safe, comfortable, and effective human use."*

To enable increased usability and safe handling of equipment as early as the development process, applicable methods are needed that support human factors engineering. In order to increase safety, use errors and usability problems must be investigated and methods for studying and reducing them are thus central. Once mismatches are charted, measures can be taken to counteract these, for example through redesign.

The methodology therefore needs to function within the field of HFE and together with the methods already established there. Furthermore, the intended users are engineers since developing and constructing machinery is an engineering job.

User-centered design

To achieve the goal of human factors engineering, an approach known as user-centered design can be applied. User-centered design is an approach in the design process which bases its information on intended users. User-centered design focuses on the users through planning, design and development of a product. According to ISO 1340 (1999) user-centered design contains the following key principles and activities:

- The design is based on an explicit understanding of users, tasks and environments
- Users are involved throughout design and development
- The design is driven and refined by user-centered evaluation
- The process is iterative

- The design addresses the whole user experience
- The design team includes multidisciplinary skills and perspectives

To apply human engineering and user-centered design a process is often used to control the work. Two models for the process are ISO 13407 Human-centered Design Process (1999) (Figure 4.7) and IEC 60601-1-6 Usability Engineering Process (2004a) (Figure 4.8).

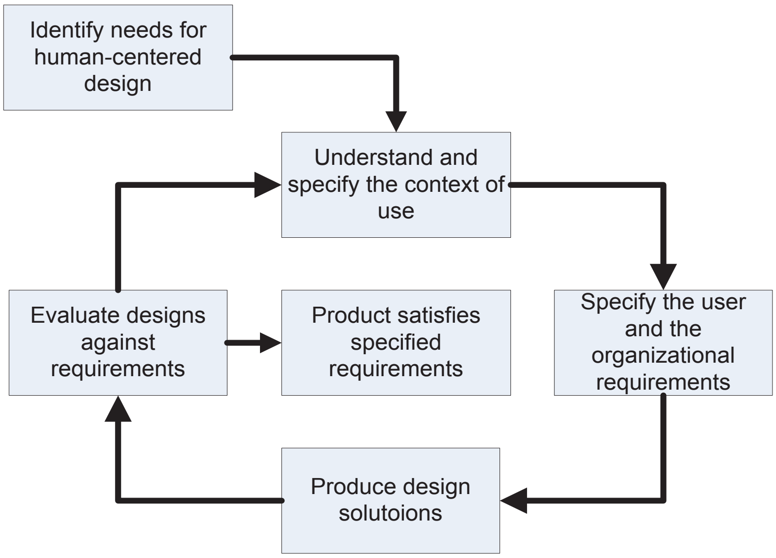
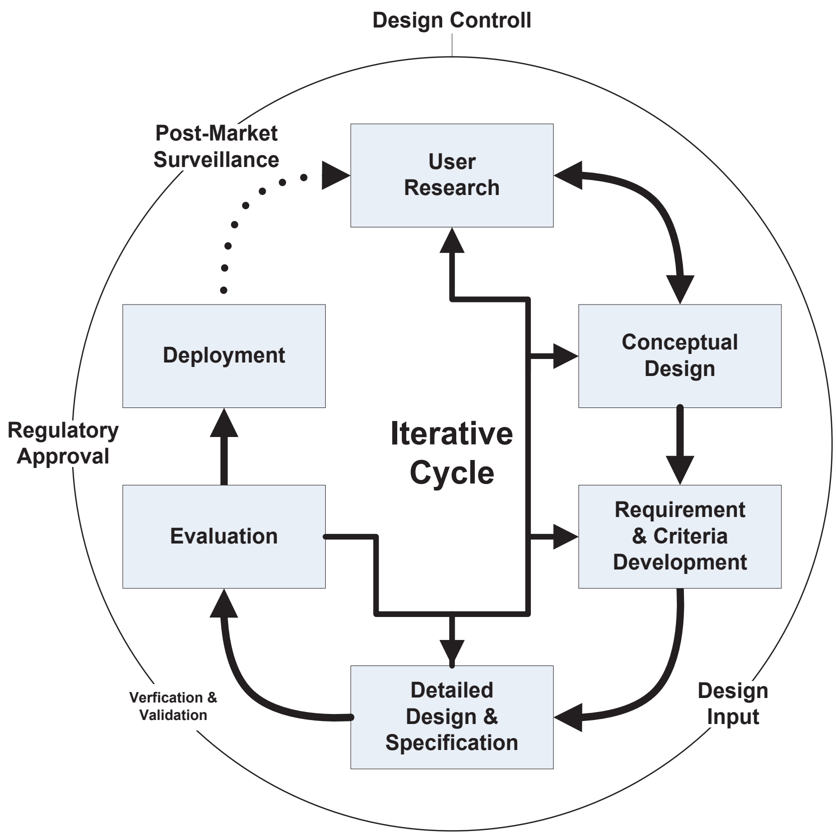


Figure 4.7 ISO 13407 Human-centred design activities (1999)



Input from users is typically obtained at nearly every stage in the circle.

Figure 4.8 IEC 60601-1-6 Usability Engineering Process (2004)

The ISO model is developed primarily for software and IT systems while the IEC model is developed for work involving medical equipment. In all steps, and for each development project, a unique assessment is made of which ergonomic activities must be performed and what level of ambition they are to have. IEC also emphasises that the same ergonomic methods can be used in many different parts of the process. Further, IEC points out that the work is seldom a serial process, even though it is described as a circle of steps. Often many of the activities in the process occur in parallel, and interaction between the steps takes place frequently, rapidly and often seamlessly. The iterative nature of the process is central for attaining successful results.

When applying Ergonomics and Human Factors in product development, it is necessary to make use of descriptions and knowledge of human characteristics, abilities and limitations, as well as a large number of methods for supporting and simplifying the work of analysis and synthesis.

The methodology during the development phase therefore needs to be able to be used within a user-centred approach by supporting the fundamental principles and working as a method in the process.

4.2.4 Input to methodology development

The areas above are vast in their own right and are usually treated separately, although they have a lot in common. The areas are directly related to the interaction between the human and the machine. The developed methodology therefore needs to relate to these areas. The methodology must function in a product development process. As mismatches in the interaction may have effects that cause serious consequences, the methodology also needs to work with the area of risk management.

The area that the methodology mostly operates within is HF / ergonomics as it concerns the human's relationship to the machine. Since the areas' physical ergonomics and cognitive ergonomics exist it is possible to develop methods for people in these areas in order to get more efficient methods than would be the case if knowledge were lacking. The people who develop machines during a product development process are primarily engineers, which mean that the method has to work for engineers, but engineers with expertise in HFE.

<u>The methodology needs to:</u>

- Be usable during the product development process
- Function together with risk management
- Be based on knowledge about physical ergonomics
- Be based on knowledge about cognitive ergonomics
- Function together with human factors engineering
- Function together with principles and processes for user-centred design
- Address users who are engineers with knowledge of Human Factors Engineering

4.3 Mismatch in interaction

The goal of the methodology is to detect mismatch in interaction. The theoretical framework thus describes this theory which is linked to the mismatch. The chapter begins by describing the mismatch in general and then continues with more detail about different types of mismatch and concludes by finally describing the main ways of relating to mismatch.

4.3.1 Lacking interaction and mismatch

In order to achieve optimal system performance and human wellbeing it is important that the human-machine interaction functions well. Lack of interaction can manifest itself in many ways and these can be divided into five groups:

- The user cannot exploit all the advantages of the technology and so does not reap the full benefit of the machine's usefulness
- The user spends too much time interacting with the machine, which leads to less time for other tasks
- The user gets stressed and uncertain, which reduces his/her ability to solve other tasks.
- The user manages the machine incorrectly, which can harm people, materials and the environment
- The user manages the machine properly, but gets harmed anyway

For errors that have consequences, there are two interesting cases to spotlight; if the user is performing correctly or incorrectly in relation to the design:

- Designed so that even if the human performs correctly, it can lead to long-term or short-term damage
- Designed so that the human has a difficult time performing correctly and/or it is easy to perform incorrectly

In the first case, therefore, the human cannot use the machine in a proper way no matter how hard he/she tries. In the latter case, the machine can be used in a good way but there are other factors relating to the machine, the human or the environment that cause errors. The reason for the inadequate interaction is that there is a mismatch in human-machine systems, i.e. that the human, machine, environment and the task do not function together in a good way. The interaction between human and machine is not always smooth and there may be a long distance between human and machine in the system's cyclic interaction. Two reasons for this can be described by two abysses (difficulties), the execution abyss and the evaluation abyss (Norman, 2002). The execution abyss is the difficulty for a user to translate a mental goal into a physical act. The user knows what he/she wants to do, but not how to do it.

- What can I do and what happens if I do something?
- What are the possible actions that can be performed in the user interface?
- Will my actions lead me closer to the goal?

The evaluation abyss is the difficulty for a user to evaluate if the response of the machine corresponds to the desired goal. The user sees, hears or feels something, but does not know what. This stops the human from performing the action with the machine.

- Is what is shown in the user interface understandable?
- Does the user interface give a good picture of the machine status?
- Is the information displayed in the user interface in accordance with reality?

Similarly, there are abysses regarding the physical, such as machines that are too big or too small to suit the individual user or that require too much force and precision. There is in other words no affordance in that case. The very word "mismatch" that is used to describe the gap between the parts of the human-machine system is taken from the book Bodyspace by Pheasant and Haslegrave (2006).

4.3.2 Sharp end and blunt end

It is consequently possible to see mismatch from two viewpoints. Firstly what arises in the use situation and has negative effects, and secondly characteristics of the machine that affect the process. The first we will call errors while the latter we will call problems.

A further description of the relationship between problems and use can be made by connecting these with the terms 'sharp end' and 'blunt end', which have been employed in regard to complex systems (Woods and Cook, 1999). The sharp end of a system is the part that directly interacts with the hazardous process, while the blunt end is the part that controls and regulates the system without direct interaction with the hazardous process. In the medical care system, nurses, physicians, technicians and pharmacists are located at the sharp end, whereas administrators, economic policy makers, and technology suppliers are at the blunt end (Woods and Cook, 1999). An error is thus something which arises at the sharp end in the medical care system, and problems originate at the blunt end – more precisely in the developers of medical equipment.

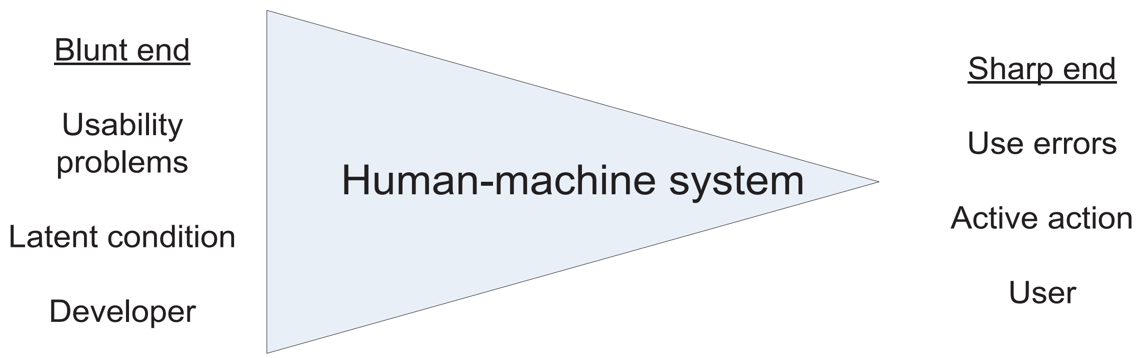


Figure 4.9 The model for sharp and blunt end, adapted from Woods and Cook (1999)

Between the blunt end, being the manufacturer, and the sharp end, being the user, there are many stages, stages that affect the user's actions. Here the organization that the user belongs to is of greatest importance for the effect that mismatch will have between human and machine. Research has shown that characteristics of organizations affect the ability of operators to perform correctly and are an important aspect, especially for providing safety. However, it is not enough to ensure safety, resilience, and so on in order to create efficiency and safety in the interaction, since it is in the actual interaction that it goes right or wrong. Many cognitive and physical features exist independently of culture and stress. Exploring the interactions at the lowest level is necessary but not enough.

4.3.3 Problems in the interaction

Ergonomics aims to optimize overall system performance and human well-being. The things that create the mismatch described above can therefore be described as ergonomic problems or interaction problems. They occur in many different forms. Two of these are usability problems and physical ergonomics problems.

Usability problems

The cause of mismatch between what the user does and what the user believes he/she is doing is called a usability problem. A usability problem is a factor or property in the human-machine system that decreases the system's usability. As shown earlier, usability is a many-faceted concept which includes, for example, both learnability and satisfaction. Nielsen (1993) describes a usability problem as any aspect of the design that is expected, or observed, to cause user problems with respect to some relevant usability measure (e.g. learnability, performance, error rate, subjective satisfaction) and that can be attributed to the design of the device.

The aspect of usability problems in focus in this thesis is what can give rise to mismatches in the human-machine system and thereby lead to use errors. Thus, a usability problem in a system can have the result that the user does not attain a goal, that the usage is ineffective and/or that the user becomes dissatisfied with the use.

Usability is thus a very important factor for ensuring high safety. Good usability counteracts usability problems that can give rise to use errors, but it also implies direct counteracting of use errors. This thesis takes its starting point in the ISO 9241-11:1998 definition of usability (ISO, 1998), yet with a clear separation between usability and functionality. Furthermore, based on the notion of risk, it is considered that a low error rate is the most essential component of usability, according to Nielsen, for ensuring safety in medical equipment. In addition, poor usability is seen as a strongly contributing cause of use errors.

Physical ergonomics problems

A physical ergonomics problem is something in the design of the machine that may cause Musculoskeletal Disorders (MSD). MSD can affect the body's muscles, joints, tendons, ligaments and nerves. There are three different factors influencing the occurrence of MSD - physical, psychosocial and individual (Winkel and Mathiassen, 1994). MSD is most often not associated with the term 'accident' as MSD often occurs after exposure over a longer duration and not due to something unexpected happening. Examples of MSD are (Pheasant and Haslegrave, 2006):

- Lifting and handling injuries
- Work-related upper limb disorders
- Musculoskeletal pain and dysfunction resulting from unsatisfactory working posture

MSD arises due to a mismatch between task demands and physical capacity in which the person is exposed to a harmful strain overload (Pheasant and Haslegrave, 2006). It may occur after a single occurrence of strain overload or after long periods of commutative strain.

What directly affects the occurrence of MSD during interaction is body position and physical strain. Body position is affected foremost by being able to see the control levers and handle actuators but also by moving all or parts of the machine or material. The physical load can be divided into three main dimensions: Level (amplitude), repetitiveness (frequency) and duration (time) (Winkel and Mathiassen, 1994). These depend upon the task and machine design.

The strain that a person is exposed to can be described by an exposure-effect model (Winkel and Mathiassen, 1994). External exposure is what the human is exposed to and this factor is independent of the single individual. Internal exposure is the effect that occurs within the human and is dependent on the individual. Active internal response is the part of internal exposure that affects the human body. Acute response is the direct effect of internal exposure, such as increased heart rate and fatigue, while chronic effects last for a longer period of time, such as MSDs.

It could also be argued that there are cognitive ergonomic problems, which in terms of the machine are something that can cause psychological harm to the user, but the long-term mental effects of use are beyond the scope of the developed methodology.

4.3.4 Error in interaction

One effect of mismatch worthy of further analysis is the errors committed by the human during the interaction with machine, interaction errors. Errors are interesting partly because they can have direct negative effects and partly because they reduce efficiency, consume unnecessary resources and make the user hesitant, i.e. four of the five listed effects. This is also clearly shown in the different definitions of usability. An error made by the user affects, for instance, all factors in the ISO definition of usability: goal fulfilment deteriorates, effectiveness decreases, and the users are less satisfied. This is clarified by Nielsen (1993), for whom the difficulty of making mistakes is one of the five properties which should concern the design. For Jordan (1998), errors made by the user affect all usability components except System Potential – which shows that here too errors have significance for usability.

For safety in particular, the difficulty in making incorrect actions is of high priority. It is the errors which expose patients to hazards. Users in such domains also believe that difficulty in making mistakes is important. Liljegren (2006) showed, for instance, in an investigation that "Difficult to make error" was the one of Nielsen's usability attributes which a selection of clinicians judged to be most important usability attributes in medical equipment. The reasons why humans err are a huge subject in its own right, Human Error.

Human error

Yet human error does not only exist in interaction with machines; it is part of everyday life. Minor mistakes occur constantly, although these are normally discovered before any serious consequences occur. Our ability to make mistakes has also inspired a well-known Latin proverb: *Errare humanum est* (to err is human) as described in the introduction. Despite the common incidence of mistakes made by people, it has become apparent that human error has a complex nature.

Much work has been devoted to defining and classifying human error. A classic definition is given by Reason (1990, p 9): human error is *"a generic term to encompass all those occasions in which a planned sequence of mental or physical activities fails to achieve its intended outcome, and when these failures cannot be attributed to some chance agency."*

Reason (Reason, 1990) subdivides human error with the aid of the Generic Error Modelling System (GEMS) and according to this subdivision, an incorrect action is one of the following:

- Slip
- Lapse
- Rule-based mistake
- Knowledge mistake
- Violation

Further, Reason makes a distinction between active failure (active error) and latent condition (latent error) (Reason, 1990, Reason, 1997). Active failures are the direct errors committed by people in a system, while latent conditions are existing defects in the system such as poor design and deficient training which, when combined with local circumstances, can result in exposure to hazard. These defects may lie latent for a long time without any harm arising, hence the term. Latent conditions can also increase the probability of active failure by creating local factors that promote error. The human errors which have been considered thus far in this thesis are of the active failure type. Reason (1997) also acknowledges the role of organisational aspects in preventing the occurrence of errors, in the same manner as the sharp end – blunt end model above (Figure 4.9).

Whereas Reason's (1990) definition of human error focuses on people's internal mental planning, Dhillon (2003) gives a definition of the concept that relates more to external guidelines. Human error, according to Dhillon (2003, p 110), is *"...the failure to perform a given task (or the performance of a forbidden task) that can lead to disruption of scheduled operations or result in damages of property."* An attempt to summarise a number of previous definitions of human error is made by Whittingham (2004, p 6): *"A human error is a failure of a purposeful action, either singly or as part of a planned sequence of actions, to achieve an intended outcome within set limits of tolerability pertaining to either the action or the outcome."*

What all these three definitions of human error have in common, however, is that they focus on the human component in the human-machine system. The definitions can be interpreted as follows: if the human had not failed, there would not be any errors. Research, though, indicates that the design of the technology is a contributory cause of human mistakes (Leveson, 1995). Leveson also concludes that it is people's ability which enables us to solve problems and make decisions based on incomplete data, sometimes leading us to commit mistakes. Without this ability, people would not be able to engage in problem-solving and decision-making. Since these traits are inherent and natural in mankind, all systems should be designed so as to be able to handle them.

If a human error were committed while using technical equipment that functioned technically as intended, the mistake was previously termed a user error. This indicated, however, that it was the user who made a mistake. During the late 1990s, a broader view of errors made by humans emerged in the sphere of medical equipment, and the term 'use error' came to be employed instead (Hyman, 1995). The term use error is now employed also by the US Food and Drug Administration (FDA) as a concept in medical equipment (FDA, 2001).

The reason why the FDA uses the term 'use error' is that errors which occur during use of medical equipment should not necessarily be attributed to mistakes by users. By speaking of use error, account is taken of the entire human-machine system and the focus is on the circumstances that led the user to make a mistake, instead of blaming everything on the user's behaviour. Use error is thus not the individual user's mistake, but an error that arises within the system. A use error may be the result of a mismatch between the different parts of the system comprised by the user, equipment, task and environment (FDA, 1999). This shift of focus from the individual user to the system as a whole shows that the FDA has placed greater responsibility for user-related incidents on the manufacturers of equipment (Patterson and North, 2006).

Standardisation organisation IEC (the International Electrotechnical Commission) has also promoted the term 'use error'. Use error is defined according to IEC (2004a, p 17) as an *"act or omission of an act that has a different result than intended by the manufacturer or expected by the operator"* in the standard IEC 60601-1-6 Medical electrical equipment – Part 1-6: General requirements for safety – Collateral Standard: Usability. The IEC justifies the term's introduction thus: *"The term was chosen over the more commonly used terms of 'user error' or 'human error', because not all use errors are the result of oversight or carelessness on the part of the operator of the medical electrical equipment. All too frequently, use errors are the direct result of poor human interface design that seduces the operator into an incorrect decision"* IEC (2004a, p 35). This agrees well with the FDA's reason for using the term.

Use Error

Returning to use error and interaction error, the two previously mentioned abysses may cause the operator to take incorrect actions during interaction with the machine, as a result of which use error occurs. Another factor that is important in the case of use error is the relationship between what the user does and what the user thinks he/she is doing (Table 4.2). If the interaction is as intended the user is performing correctly, and she/he believes to be performing correctly. But if a use error occurs, the user can either detect it or not detect it. In the case where the user detects the error, she/he is able to correct it, but if the user does not detect an error, it leads to a discrepancy between what the human and the machine register as happening. This condition is most serious when the user has no control, and there might be dangerous situations during continued use. The opposite may also occur when the user is performing correctly but believes that he/she has done something incorrectly, which means that the use is counteracted as the user does not move forward (and may instead, in a worst case scenario, make errors in the search).

Table 4.2 Relation between error and detection of error during use

	User performs correctly	**User performs incorrectly**
User believes to be performing correctly	Correct use	Incorrect use
User believes to be performing incorrectly	Missed use	Stopped use

The crucial point in the preceding section was that use errors occur because of a mismatch between parts of the system user, equipment, task and environment (FDA, 1999). (Kaye and Crowley, 2000) describe more specifically six reasons why use-related hazards occur:

- Devices are used in ways that were not anticipated
- Devices are used in ways that were anticipated, but inadequately controlled
- Device use requires physical, perceptual, or cognitive abilities that exceed those of the user

- Device use is inconsistent with the user's expectations or intuition about device operation
- The use environment affects device operation and this effect is not understood by the user
- The user's physical, perceptual, or cognitive capacities are exceeded when using the device in a particular environment

These six causes are examples of the interplay between human and machine not functioning well.

The relationship between usability problems and use errors is the same as that between active failure and latent conditions (Reason, 1990, Reason, 1997). A usability problem is a latent weakness in the system of human, machine, environment and tasks that triggers, under certain circumstances, a use error in the system. Having said that, a use error not always need be caused by a usability problem, just as not all usability problems need cause use errors.

Ergonomic errors

One type of error that may occur during the interaction is actions that can harm the user in the short or long term. These errors are covered only partially by use error as there may be times when neither the user nor the manufacturer knows that the action is harmful in the long term. An ergonomic error may thus arise even if the action is carried out as intended. Most ergonomic errors relate to physical ergonomics as it is uncommon to harm a person's cognition by single actions, even though the senses can be damaged by extreme light and sound.

The relationship between physical (and cognitive) ergonomic problems and ergonomic errors is the same as that between usability problems and use errors. A physical ergonomic problem is a latent weakness in the system of human, machine, environment and tasks that, under certain conditions, triggers an ergonomic error in the system. However, an ergonomic error need not always be caused by a physical ergonomic problem, just as not all physical ergonomic problems need cause ergonomic errors.

4.3.5 Counteract effects of mismatches

The theoretical framework that has been presented so far has shown that problems and errors in interaction are important in the work to avoid mismatch. The framework has also shown that it is relevant to work with errors and problems and that they occur in a number of areas.

What are the proper ways to operate in order to avoid problems and their effects? It is possible to differentiate between four approaches in relation to their effects.

1. Non-teaching
2. Reactive
3. Proactive 1
4. Proactive 2

To exemplify these approaches, accidents, strain injuries and incorrect actions are used.

If the description is applied to accidents it will mean:

1. Non-teaching: Taking care of the damage caused by the accident
2. Reactive: Wait until an accident occurs and until interventions are carried out
3. Proactive 1: Analyze situations to prevent accidents before they occur
4. Proactive 2: Create no harmful situations

If the description is applied to strain injuries it will mean:

1. Non-teaching: Treat injuries that have occurred.
2. Reactive: Wait until injuries occur before interventions are carried out
3. Proactive 1: Analyze work tasks to prevent injuries before they occur
4. Proactive 2: Create tasks and equipment so they do not cause strain injuries

If the description is applied to incorrect actions it will mean:

1. Non-teaching: Taking care of the consequences of an incorrect action
2. Reactive: Wait until an incorrect action occurs before interventions are carried out
3. Proactive 1: Analyze equipment to try to prevent incorrect actions before they occur
4. Proactive 2: Create equipment that cannot be used incorrectly

The best way to avoid mismatch is to initially design the machine in a way that is adapted to the human, task and environment (Proactive 2). This has been the subject of a lot of research within the field of ergonomics and human factors (Sharp et al., 2007, Boy, 2011, Monö, 1997, Pheasant and Haslegrave, 2006, Nielsen, 1993, Wickens and Hollands, 1999, Galitz, 2007, Gulliksen and Göransson, 2002). It is also important to evaluate machines to find possible mismatches before any negative consequences occur (Proactive 1), in order to avoid shortages that are not (Proactive 2) solved and that no one is aware of.

4.3.6 Input to methodology development

After a carefully study of mismatch, it becomes clear that it manifests itself in two ways. First as an active incorrect action by the human interacting with the machine, and second as a latent state being a result of the design of the machine. Thus, an analysis of mismatch must focus on both parts, i.e. both the blunt end and the sharp end.

The theory compilation also shows that mismatch may have negative effects on actual use via usability problems, but also on the user through physical ergonomics problems.

One limitation is that the methodology will not investigate physical ergonomics problems that occur during intended use. This limitation is of a practical nature as MSD is a vast area in itself, and also because no need has been found in the projects where the methodology has been developed; the already existing methods have been sufficient.

Something that the existing methods does not address, however, is the physical ergonomics problems related to ergonomics errors, i.e. when the user does not perform a task in a sound ergonomic way, although it is possible - hence the links between psychological ergonomics and usability. To evaluate the true ergonomics, not only the possible ergonomics, of a machine, it is important to ask the following questions:

- Will the task be performed in an ergonomic way?
  Why? / Why not?
- Can the task be performed in a non-ergonomic way?
  What are then the consequences?

The method development will concentrate on usability problems, use error and ergonomic errors with the focus on the cognitive part; why the user does what he/she does as well as how and why it turns out right or wrong.

The next governing factor is the level on which the analysis should be conducted. Mismatch between human and machine can be found at many levels, from the labelling of individual control levers up to the choice of conceptual models. However, it is still during the individual action that the mismatch will make itself felt. It is the individual action that is conducted correctly or incorrectly, regardless of the level of the mismatch. The methodology must therefore focus on what goes right or wrong in the detailed interactions, i.e. in the individual actions.

However, the methodology will not consider Abnormal Use, since this does not lie within the framework of the IEC definition of use error, and is therefore not the manufacturer's responsibility.

What the methodology needs to examine in greater detail, to study the mismatch, can be summed up in the form of six questions (Table 4.3). The answers to these questions can then be used to counteract the mismatch and thereby increase efficiency and safety.

Table 4.3 Requirement questions for method development

1 a. Will the user act correctly with the machine? b. Will the user act correctly with the human?
2. Why does the user act correctly?
3 a. Which errors can the user commit with the machine? b. Which errors can the user commit with the human?
4. Why does the user act incorrectly?

Of these four questions, Nos. 1a and 2 derive from the area of Usability Evaluation, while Nos. 3a and 4 originate from the area of Human Reliability Assessment. Nos. 1b and 3b derive from evaluation of physical ergonomics.

Questions 1 and 2 address usability problems, whereas 3 and 4 address use errors and ergonomics errors. It is necessary to deal with both aspects in order to improve safety, since the human can be harmed directly by use errors and indirectly by usability problems. The answers to these four requirement questions will be helpful in improving usability and safe handling of the evaluated equipment.

The next property is how the methodology provides answers to the stated questions (1-4 above). Since the statements are given in the form of questions, it was decided to incorporate the questions in the methods. Thus the methods become question-based in the analysis. Question-based analytical methods, according to the author's own experience, are simple to teach, yield fast results, and are advantageous to work with in groups (as is normally done in product development projects). Their simplicity is due to the methods' functioning as a type of check-list.

The methodology needs to:

- Relate the blunt end with the sharp end
- Investigate usability problems
- Investigate use error
- Investigate ergonomics errors
- Not investigate physical ergonomics problems during correct use
- Take no account of abnormal use
- Investigate:
    - Will the user act correctly with the machine?
    - Will the user act correctly with the human?
    - Why does the user act correctly?
    - Which errors can the user commit with the machine?
    - Which errors can the user commit with the human?
    - Why does the user act incorrectly?
- Be question-based

4.4 Interaction evaluation

The theoretical framework concludes with an examination of various methods for evaluating interaction and for detecting and identifying mismatch. The review begins with a general method theory and continues with different groups of evaluation methods.

4.4.1 Method theory

To study the human-machine interaction there are several methods within different areas and here areas related to the thesis will be addressed. Yet, before that a bit of basic methodological theory is needed to help guide the reader between the different types of methods. Methods can be divided between two factors, its type of data and its approach.

Type of data

Data is what a method collects, processes and/or presents. This data can be empirical or analytical, objective or subjective as well as qualitative or quantitative.

Empirical and analytical

Methods can be categorized according to how the data is collected (Bohgard, 2009). The origin of the data can be either from empirical or analytical studies. Empirical data comes from the direct study of real phenomena, for example from observations and interviews with users, measuring of physical dimensions and environmental factors as well as practical strength tests. Analytical data comes from indirect studies, i.e. studies that do not collect data directly from reality, but instead represent the findings theoretically. Analytical data could come from strength calculations, evaluations from templates and other analytical methods such as shape analysis.

Objective and subjective

The next categorization concerns the source of the collected data and the categories are as follows: objective, semi-objective and subjective (Bohgard, 2009). Objectively collected data is obtained by direct measurements of real-world phenomena, such as people's heights and weights, but also heart rate, EMG and oxygen uptake. Furthermore, objective data can treat observations of how many times a person jumps in and out of a truck during a work shift, or how often an operator returns to the main menu of an interface when he/she solves a predefined task.

Subjective data is dealt with when humans are the study objects. This information is obtained when they themselves verbally or in writing express what they believe, feel or think about something. Subjective data is therefore the user's experience of physical strain on a muscle group, total effort, discomfort or mental workload, for instance. Semi-objective data is obtained when a person (not a test subject) makes an estimate or assessment of a phenomenon from a template.

Qualitative and quantitative

The last category of data describes the type of data, of which there are three types: quantitative, semi-quantitative and qualitative (Bohgard, 2009). Quantitative data is direct numbers from a measurement or an observation. Semi-quantitative data is categorization or ranking from scales, for example how uncomfortable a certain seat is on a given scale or just how severe the consequences are of use error (on a different scale). Qualitative data is a description and understanding of the world and context in the form of words and images. Qualitative data answer questions parameters such as what, who, how, when and where.

Approach

The next classification of methods is based on its approach. What is interesting for the thesis is: reactive or proactive, formative or summative and analytical, expert or user.

Reactive and proactive

The first categorization of methods can be made by defining whether the analysis is reactive or proactive. According to Cacciabue (2004), studies of Human Machine Systems and Human Error and Accident Management can be divided into two different approaches: Retrospective Analysis and Prospective Analysis. These approaches complement one other and contribute equally in the design of safe technology.

Cacciabue describes *retrospective analysis* as follows: *"Retrospective analyses consist of the assessment of events involving human-machine interaction, such as accidents, incidents, or 'near-misses', with objective of a detailed search for the fundamental reasons, facts, and causes ('root causes') that fostered them"* (Cacciabue, 2004, p 24). In addition, Cacciabue describes *prospective analysis* thus: *"Prospective analyses entail the prediction and evaluation of the consequence of human-machine interaction, given certain initiating events and boundary configuration of a system"* (Cacciabue, 2004, p 24). A fundamental part of prospective analysis is creative thinking, as the analysis aims to predict the behaviour of users in a man-machine system, especially when the analytical methods do not include any real users.

Formative and summative

The next categorization is based on when the methods are used during the development process and allows the methods to be divided into formative or summative methods (Hartson et al., 2001, Leventhal and Barnes, 2008). Formative methods are performed during the product development process and their purpose is to shape the product towards the ultimate goal, among other things to detect potential problems in design and then counteract them by re-design. In the case of formative evaluations these should be performed as early as possible during the product development process because it is less resource-intensive to change a design at that stage of the process (Hix and Hartson, 1993).

A summative method is applied to the final result of the development process by summarizing it. A summative evaluation examines how well the finished product turned out. Methods can be both summative and formative, depending on how they are used in relation to the product development process.

Analytical, expert or user

The final categorization concerns the process of how data is generated in the method and the type of method performing the process. The methods may here be categorized as analytical, expert or user (Leventhal and Barnes, 2008). An analytical method applies a systematic and structured process to achieve its goal. The systematic method and structure are used to reduce the assessments done by the performer, thereby increasing the method's objectivity. An expert method relies on experts in the field being able to use their general knowledge and experience to collect necessary data. During an expert evaluation an expert assesses a product to list its good and bad qualities. The last group is user methods and this method supposes that a user with his/her specific knowledge and experience gained during use is able to obtain the necessary data. During a user evaluation it is therefore the user who assesses the product to list its good and bad qualities.

4.4.2 Evaluation methodology

When applying Human Factors Engineering (including User Centered Design) in product development, it is necessary to make use of descriptions and knowledge of human characteristics, abilities and limitations, but also a large number of methods for supporting and simplifying the work of analysis and synthesis. Important methods include gathering information from users, such as interviews, observations and surveys. However, the present frame of reference will focus on methods that are dedicated to working on mismatches in the Human Machine Systems.

Evaluation for investigating mismatch in the interaction can be conducted in different ways. The easiest way is to let a user or HF expert use a machine and then gather their opinions of possible mismatch through an interview. This method is usually referred to as Review (Bligård, 2011) or Applying Interviews to Usability Assessment (Young and Stanton, 2004).

The next step is to apply more structure to the evaluation, such as a checklist for the HF expert and a test procedure for the user. An example of the checklist is heuristics evaluation (HE) (Nielsen and Mack, 1994) and for the latter Usability Testing (UT) (Nielsen, 1993).

A third way to get a more structured process is an analytical evaluation where the systematic procedure of the method is the basis of the results. Examples include the methods Cognitive Walkthrough (CW) (Lewis and Wharton, 1997) and Predictive Human Error Analysis (PHEA) (Baber and Stanton, 1996). These methods will reoccur in the thesis, but here is a brief presentation of the specific areas of evaluation methods that are relevant for this thesis as well as common methods within them. The areas are:

- Evaluation of risk
- Usability evaluation
- Human reliability assessment
- Evaluation physical interaction

Evaluation of risk

Methodology and methods to analyze risk need to be proactive and formative to handle hazards before consequences occur. A central part of Risk Management is Risk Analysis, also termed Safety Assessment. Risk Analysis is usually defined as work on identifying hazards and assessing risks in a system (ISO, 2000a, Dhillon, 2003, Sandom, 2004, Kolluro and Brooks, 1996).

Central to risk analysis is, therefore, identification of dangerous situations or events, analysis of their causes, and investigations of their possible consequences as well as their probability of occurrence. On this basis, an assessment is made of the risk in order to determine whether risk-reducing measures need to be taken. The two steps in the Generic Risk Management process (Figure 4.5) that constitute Risk Analysis in the process are Identifying Hazards and Estimation of Risks.

The methods and approaches used in risk analysis can be divided into qualitative, semi-quantitative, and quantitative ones (Stricoff, 1996). A qualitative analysis is often the first step in working with risks. This analysis is used to investigate which risks may occur and how they occur. In a semi-quantitative analysis, risks are then ranked relative to each other or with the help of scales or matrices. An example of a simple scale for assessing consequences is found in the IEC standard for alarms, IEC 60601-1-8 (IEC, 2003):

High:	Death or irreversible injury
Medium:	Reversible injury
Low:	Minor injury or discomfort

A purely quantitative analysis means that both the probability and consequences of a risk are quantified with numbers, i.e. that the probability is calculated as a number between 0 and 1 while the extent of the consequence is described, for instance, in terms of type and number of injuries. This makes a quantitative risk analysis much more resource-demanding than a semi-quantitative one. Ranking and quantification of risks are done so that the risks can be compared, allowing for an evaluation of different ways to minimize the total risk. In relation to the Risk Management process illustrated in Figure 4.5, Identifying Hazards can be seen as a qualitative approach whereas Estimation of Risks is quantitative or semi-quantitative.

Two frequently used methods in identifying hazards are, firstly, Hazard And Operability Study (HAZOP) and, secondly, Failure Mode and Effect Analysis (FMEA) (Taylor, 1994). HAZOP involves studying deviations from a normal state. This method investigates what happens if something becomes too large or small in a system, such as the value of a parameter or the level in a tank. The deviation's consequences and causes are described. The principle of FMEA is to examine each component in a system and ask how it can break down, and what happens if it does. For instance, what are the consequences if a specific cooling pump in a nuclear power plant stops working? If the consequences are serious, a severe risk has been identified.

For estimating the risk itself, two common methods are Fault Tree Analysis (FTA) and Event Tree Analysis (ETA) (Stricoff, 1996). FTA considers an event that may cause damage (called a top event) and describes what must happen before the event can occur. For example, a car parked on a slope cannot begin to roll unless, among other things, both the handbrake and the gears are released. ETA starts with an original event and describes the process that must take place for the event to cause damage. It also shows what other consequences the original event may have, depending on how the process develops. In the case of a parked car, the consequences of its beginning to roll depend on what lies further down the slope. If nothing is there for the car to hit before it stops, no consequences exist even though a hazard has arisen.

Usability evaluation

The next group of methods exists to evaluate usability of machines. This group of methods aims to assess usability in the human-machine relation and derives from one or more usability attributes, and/or to find problems and improvement possibilities in the design. Examples of attributes are learnability, efficiency, memorability, error and satisfaction (Nielsen, 1993). An usability evaluation can be conducted, both summative and formative, and there are methods that are analytic, expert or user-oriented (Leventhal and Barnes, 2008).

Results deriving from the evaluation methods may be of either a qualitative or a quantitative nature. Examples of qualitative results are descriptions of usability problems and use errors, since they yield a descriptive result without any numbers. An instance of quantitative results is how much time it takes for the user to perform a certain task, or how many mistakes were made, i.e. results that can be stated in numbers. The results may also be semi-quantitative, meaning that the user or analyst grades different aspects of usability on a scale. One example is a scale for grading the seriousness of identified usability problems (Table 4.4).

Table 4.4 Grading of problems related to usability (Nielsen, 1993)

Grade	Description
0	Not a usability problem
1	Cosmetic problem only, need not be fixed unless extra time is available in the project
2	Minor usability problem: fixing should be given lower priority
3	Major usability problem: important to fix, should be given high priority
4	Usability catastrophe: imperative to fix this before product can be released

Three common methods for usability evaluation are the previously mentioned Heuristics Evaluation, Usability Testing and Cognitive Walkthrough. A heuristic evaluation (HE) can be briefly described as a review of a machine by a list of heuristics / principles / guidelines (Nielsen and Mack, 1994). Examples of principles are; "dialogue should be simple and natural" and "minimize the user's mental load." A heuristic evaluation often contains 5-20 key guidelines, as more can make the evaluation difficult to perform. During the heuristic evaluation the machine's deviation from the guidelines is noted and the possible problem severity is evaluated. In the literature there are heuristics for different types of machines but they can also be developed from theory specific for a project. The results of HE become a subjective qualitative and semi-quantitative list of deviations from the guidelines and as HE is usually performed by an expert in usability, it becomes an expert method.

A usability testing (UT) is an evaluation in which a user performs predetermined actions with a machine under controlled circumstances (Nielsen, 1993). The idea is that user interaction is studied (often through videotaping) to analyze parameters such as the number of keystrokes, number of errors, number of corrected errors, number of hesitations, use of the manual or if the subject gives up. During the test, subjects are often asked to speak out loud and describe what they do and how they think. It is not just the cognitive ergonomics that can be studied during a usability test, even the physical ergonomics can be studied by the user's various postures during use. The test is often followed by an interview and/or a questionnaire on how the interaction with the machine was experienced (both physically and cognitively). A usability test does not have to be done with a full-developed machine, it is perfectly possible to use simpler versions such as card game to display a user interface. UT is a user approach method and generates empirical data, subjective and objective, as well as qualitative and quantitative.

During a Cognitive Walkthrough (CW) (Lewis and Wharton, 1997) the evaluator or evaluation team simulates the user's thought process to understand how the interaction with the machine will turn out. The method is based on theory of how people want to learn by trial and error. The method is based on the correct way to do something, and subsequent use of a questioning process to find what determines whether the user will be able to perform the action sequence according to the manufacturer’s definition and in the right way. CW is basically an analytical method because of its very systematic approach, but may have elements of both expert methods and user methods as both usability experts and users are often involved when the method is performed in groups. The result of CW is a qualitative list of possible problems that may arise during the interaction. All three methods can be used both during development (proactive) and to evaluate the machines in the field (reactive), as well as a proactive approach both formative and summative.

Human reliability assessment

One special group of proactive usability evaluation methods is Human Reliability Assessment (HRA)[1]., According to Kirwan (1994, p 17), the goal of these methods and techniques is to *"assess the risk attributable to human error and ways of reducing system vulnerability to human error impact."* HRA operates through three basic functions: *"... identifying what errors can occur (Human Error Identification), deciding how likely the errors are to occur (Human Error Quantification), and, if appropriate, enhancing human reliability by reducing this error likelihood (Human Error Reduction)"* (Kirwan, 1994, p 2).

The methods and techniques used in HRA can be divided roughly into two groups: qualitative and quantitative (Stanton and Baber, 1996, Embrey, 2004). The former methods, called Human Error Identification (HEI), are used to indicate which errors are likely, while the latter, called Human Error Probabilities (HEP), are used to predict the probability that a given error will occur.

Common methods for Human Error Probabilities are the Human Error Assessment and Reduction Technique (HEART) (Williams, 1986), Technique for Human Error Reduction (THERP) (Swain and Guttman, 1983), and Justification of Human Error Data Information (JHEDI) (Kirwan, 1994).Using data showing the probability that a human will commit errors, these methods aim to assess how probable it is that a specific incorrect action will occur.

The methods employed for the qualitative analysis in HRA (HEI techniques) operate primarily within three areas (Baber and Stanton, 1996):

1. *"In design of new artefacts, so that potential errors can be identified and rectified before production."*
2. *"In risk assessment, so that the impact of safety critical errors in system operation can be reduced."*
3. *"In accident investigation, so that the cause of errors can be established."*

One feature shared by different HEI techniques is that they normally contain several or all of a set of elements listed below. This list is a compilation from Baber and Stanton (1996), Stanton and Baber (1996) and Stanton and Stevenage (1998):

- Take account of all actions that humans can perform with an artefact or a system
- Break down activities in sequences of steps
- Identify possible and plausible incorrect human actions
- Indicate possible psychological error mechanisms that underlie the incorrect action
- Judge the consequences of incorrect actions
- State ways of recovering errors
- Give proposals for how errors should be prevented or mitigated

A frequently used group of qualitative methods consists of Action Error Analysis (AEA) (Taylor, 1979, Suokas and Pyy, 1988), Systematic Human Error Reduction and Prediction Approach (SHERPA) (Swain and Guttman, 1983) and Predictive Human Error Analysis

[1] The grouping of HRA as a part of usability evaluation methods is not usually done in the literature, but it should be done since HRA investigates a central aspect of the usability of a product/system, namely whether the user will commit errors.

(PHEA) (Baber and Stanton, 1996, Embrey, 2004). These methods originate in work with nuclear energy operators and mistakes in control rooms. The analysis is undertaken to examine what potential human errors can occur in the interaction between user and machine, and why errors occur. The goal of this analysis is to identify the sequences of use where operational errors can affect the performance and fulfilment of system goals. The methods aim to break down the user's working tasks into steps (interaction) and then, for each step, to identify and investigate potential use errors that the user may commit. The procedure for the methods is similar to the one for Cognitive Walkthrough as analyses are required to simulate user actions in each step and try to predict what might happen. The results of the methods are a qualitative list of potential user error which may occur during the interaction.

Evaluation physical interaction

The groups of methods have so far been largely concerned with the cognitive interaction between human and machine. In addition, however, the physical interaction needs to be evaluated. Evaluations of physical interaction are conducted mainly with two aims, to evaluate whether humans can physically perform the task, and to investigate whether humans will be injured when the task is performed. The first aim is to achieve good efficiency during use and the latter to avoid the occurrence of MSDs.

To examine if the human can perform the task using the machine, user anthropometry can be used as reference. (Pheasant, 1990). The design of the machine is then compared to the anthropometric data of users to determine if the machine is adapted to the human. The comparison can be helped by using computer manikins (Porter et al., 1990).

To examine the likelihood of MSDs occurring, a group of methods originating in biomechanics can be used. Posture is evaluated to determine whether any harmful loads may occur (Moira, 1990). To simplify the analysis, methods based on observation of body position have been developed to assess how harmful the position is. Examples of such methods are Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett, 1993) and Rapid Entire Body Assessment (REBA) (Hignett and McAtamney, 2000). These types of analyses can also be performed with previously mentioned computer manikins. These two groups are of a more analytical nature as they do not rely on human judgments but instead give the answers directly.

A third way to examine the physical interaction is to carry out measurements on the human body such as heart rate to measure effort. Strain on individual muscles can be determined through measurements with electromyography (EMG) (Sandsjö, 2004). The methods will gradually become more empirical and proceed directly from the user, but with objective data.

Input to methodology development

The input from interaction evaluation is based on which properties the methods must have in order to effectively enable work with human safety. It is essential for the methods to detect mismatches as early as possible in product development. Hence, the methods need to be proactive and not reactive. Mismatches should thus be discovered and counteracted before they occur. A reactive method exposes humans to unnecessary risks as it demands that something must happen before measures are taken. Besides, a reactive method has the drawback that it cannot be used early in product development. Therefore, the methods that have been developed are proactive and formative, i.e. they seek to evaluate the equipment during the design process with the aim of identifying problems so that these can be dealt with.

Often no physical equipment or prototype exists early in the product development phase. This makes empirical methods less effective since the evaluation becomes less realistic. In addition, it is often difficult to obtain enough potential or actual users to be able to apply empirical methods on a large scale. The methods that are to be refined should thus be analytical because they do not depend as much as empirical methods on involving potential or actual users in actual tests. Creating a methodology that is completely analytical, that is without assessing people, is not realistic when the knowledge about users and use needed for evaluation comes mainly from human sources. It would also require a comprehensive structure to move knowledge from the human to the methodology to make the methodology able to carry out the assessment itself. It is a more reasonable approach to create an analytical framework that enables the methodology to benefit from the knowledge of experts and users, an analytical methodology that integrates with the users and experts.

One disadvantage of using formative analytical evaluation however, is that the problems detected are potential and not the "real" ones. To confirm that the problems are real, the equipment has to be actually used, but this would expose humans to needless risks. Hence, a formative analytical evaluation is unavoidable if the method is to work proactively. This also leads to generated data being subjective as it is ultimately based on judgments by experts and users.

To take measures against usability problems and use errors, the methods must detect and identify these explicitly, which means that the methods should be qualitative. Nonetheless, a certain amount of semi-quantitative work is needed to enable ranking and prioritising between the detected problems and errors. It is important for the most serious problems to be tackled first.

In all, the methods must be applicable early in the development process when there are often no prototypes (physical product representations), must not be dependent on the availability of users, and must be able to detect problems which can then be ranked and prioritised.

One conclusion from the theoretical framework was that it is important to consider use errors and usability problems in work with counteracting mismatch. The methodological framework described how work with user errors and usability problems takes place in the context of Risk Analysis and Human Factors Engineering. This framework has also shown that analysis of use errors is mostly to be found within the Human Reliability Assessment (HRA) method group, and that analysis of usability problems is preferably found in the Usability Evaluation (UE) group. The method development which is described in this thesis will therefore lie within the framework of Risk Analysis and Human Factors Engineering, and more specifically within Usability Evaluation and Human Reliability Analysis. As the methodology focused on detecting mismatch and not on assessing physical ergonomic effects, methods for Evaluation physical interaction were not considered as a starting point for the method development. These methods will instead act more as a mental support.

<u>The methodology thus needs to:</u>

- be formative and proactive
- be analytical, but include users and experts
- yield qualitative and semi-quantitative data
- be based on methods within the area of Usability Evaluation and Human Reliability Assessment

5 The developed methodology - CCPE

In this chapter the developed methodology is presented (5.1). It is followed by a description of the development of the methods and the methodology (5.2). The chapter ends by showing how the methodology was used (5.3) in different projects.

5.1 Description

The method and methodology development resulted in the Combined Cognitive and Physical Evaluation (CCPE) (Paper VIII). It contains four new methods: Generic Task Specification (GTS), Enhanced Cognitive Walkthrough (ECW), Predictive Use Error Analysis (PUEA) and Predictive Ergonomic Error Analysis (PEEA).

5.1.1 Procedure

The CCPE methodology consists of four phases: (1) Definition of evaluation, (2) Human-machine system description, (3) Work load analysis, and (4) Interaction analysis. Phases 2-4 consist of 3-4 parts each to address the various factors included in the human-machine system (Figure 5.1).

<table>
<tr><th colspan="4">1. Definition of evaluation</th></tr>
<tr><th colspan="4">2.Human-machine system description</th></tr>
<tr><td>a. User profiling</td><td>b. Task analysis</td><td>c. Context description</td><td>d. Interaction description</td></tr>
<tr><th colspan="4">3. Work load analysis (by GTS)</th></tr>
<tr><td>a. Task demands</td><td>b. Automation levels</td><td>c. Mental workload</td><td>d. Physical workload</td></tr>
<tr><th colspan="4">4. Interaction analysis</th></tr>
<tr><td>a. Usability problem analysis
(by ECW and alarm-ECW)</td><td>b. Use error analysis
(by PUEA)</td><td colspan="2">c. Ergonomics error analysis
(By PEEA)</td></tr>
</table>

Figure 5.1 Procedure of the CCPE methodology, which includes four phases of evaluation of a human-machine system

The exact content of each phase may vary depending on the purpose of the evaluation. This makes the CCPE methodology highly adaptable to all kinds of human-machine systems, from handling a consumer product such as a kitchen aid to the nurse's workplace around the patient's bed in an intensive-care unit. The methodology can be conducted by a single analyst or by a group of analysts; this may be made up of designers, software developers, marketing staff, human factors experts and users. What is most important, however, is that there is knowledge about the users and knowledge about usage of the machine among the people who conduct the assessment. Staffing is described in greater detail later in chapter 6.8.

5.1.2 Definition of evaluation

The first phase in the CCPE methodology (Figure 5.1) is to establish the framework for the analysis, which then serves as a basis for further analysis. The definition of the evaluation phase should answer the following five questions:
(1) What is the purpose of the evaluation?
(2) Which machine will be analysed?
(3) Which usage will be analysed?
(4) Who is/are the intended user(s)?
(5) What is the context for its use?

The choice of machine, use and user made before the analysis begins is decisive for the quality of the subsequent analysis. If the choice of intended use or user is inadequately or wrongly made in relation to the actual conditions, the entire evaluation of usability of the machine will be inaccurate. It is very important to understand that CCPE methodology is highly dependent on the input data upon which the analysis is based.

5.1.3 System description

The second phase in CCPE methodology (Figure 5.1) is the system description, which illustrates how the human-machine system works. The system description is most important for the work and interaction analysis because if the system description is deficient, incomplete or wrong, the results of the analysis will suffer. The system description consists of four activities: user profiling, task analysis, context description, and interaction description.

User profiling

The first part of the system description is to profile the user, i.e. the human(s) acting in the system. User profiling describes the intended user's abilities, limitations and characteristics (such as knowledge, experience and anthropometrics) that are relevant for the evaluation. The user profile captures the diversity and heterogeneity of the user group and searches for the greatest common characteristics. A user profile can include different items for investigation depending on the characteristics of the human-machine system.

As the goal of the evaluation is to find mismatches between the human and the machine, the user must be correctly and carefully described to obtain an accurate result. Since the interaction analysis is based on a simulated user, the user profile must map the internal performance-shaping factors to increase the relevance of the evaluation.

To make the user more visible the user profile can be complemented by a persona (Nielsen, 2004). A persona is a fictitious but realistic user of the machine created to represent a user group. The aim of a persona is to make the user more concrete and living and in that way help designers meet actual user needs and preferences. A persona should be written so well that the designers are enabled to perceive and identify the intended user as an authentic person and constantly be reminded to integrate user needs and preferences in their design proposals. A persona can be described as a user profile that is brought to life.

Task analysis

The second part of the system description is the task analysis, which consists of two parts: (1) selection and grading of tasks, and (2) specification and description of tasks. Task here includes both cognitive and physical operations.

Selection and grading of tasks

The first step is to choose which tasks are to be evaluated, and then to grade them according to importance. The tasks chosen for evaluation naturally depend on the aim and goal of the study. The aim of the study may be to evaluate tasks that are carried out often, or tasks that are carried out more rarely but are safety-critical. The selection of tasks must be based on the intended use, not on the design or function, of the machine. Above all, it is important that the tasks are realistic.

Each task to be evaluated is given a unique number, or *task number.* The tasks are to be graded from 1 to 5, based on how important they are in the intended use of the machine. The most important tasks are graded 1 and the least important 5. The grading is called *task* *importance*. To allow a comparison between different machines, it is important that the tasks which are selected for comparison should have the same task importance for all user machines analysed.

Specification and description of the task

The way chosen tasks should be performed in the correct way by the intended user is described in detail with the method of Hierarchical Task Analysis (HTA) (Stanton, 2006). The correct way is in accordance with the intended use of the machine described by the manufacturer. HTA describes how the overall goal of the task can be achieved through sub-tasks and plans. The HTA breaks a task down into elements, or sub-tasks. These become ever more detailed as the hierarchy is divided into smaller sub-tasks. The division continues until a stop criterion is reached, often when the sub-task consists of only one single operation (so-called progressive re-description). The result is usually presented in a hierarchical tree diagram (Figure 5.2). The breakdown into a hierarchical structure based on different levels is vital, since the subsequent interaction analysis is performed on all levels. If the task is only described as a series of actions, it is not sufficient for the interaction analysis. Further both physical and cognitive operations should be included in the task analysis if that is the purpose of the analysis.

In CCPE methodology the bottom level in the HTA, i.e. the individual steps in the interaction between human and machine, is termed *operations*. The tasks and sub-tasks that lie above the bottom level of the HTA are termed *nodes*. A node, together with underlying nodes and operations, is termed a *function* (Figure 5.2). In the function boxes the (sub-) task goal is described and in the operation boxes both operation goal and the action are entered. The nodes and operations in the HTA must be numbered uniquely in order to facilitate the compilation of results from the analysis. For example, a function can be described according to the uppermost node. When making comparisons between different machines, the design of the HTA will be different for each machine. This is not a problem for the subsequent analysis, as long as the subdivision and grading of tasks are the same for all the machines. If they are not the same, however, it will be impossible to make a relevant comparison.

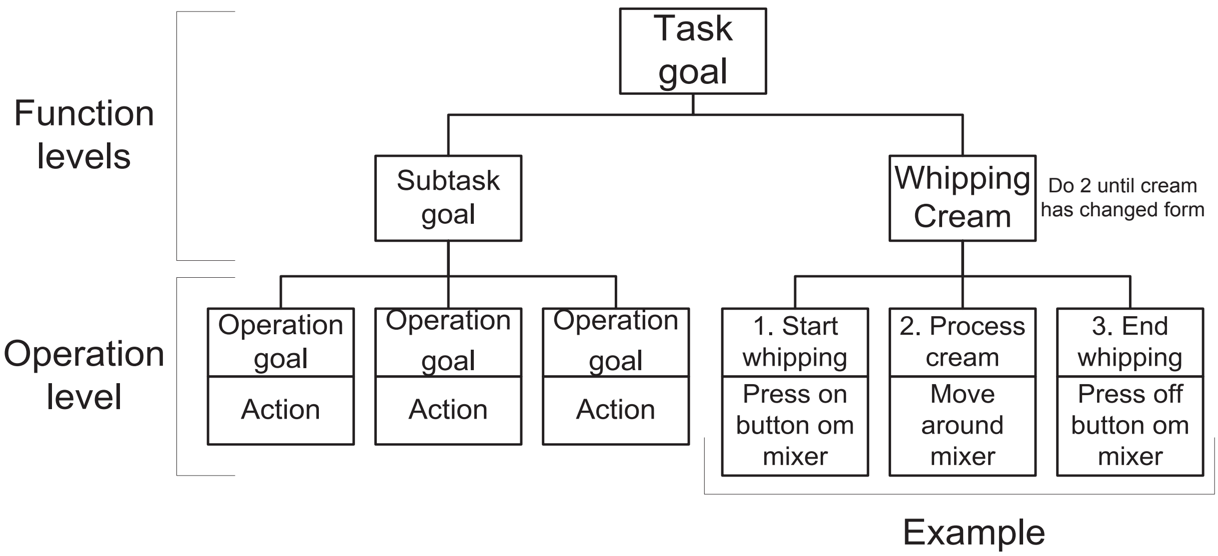


Figure 5.2 HTA hierarchy with operations and functions

It is worth noting that an HTA diagram only describes one correct way (with variants) in which a user can perform a task. Often a task can be performed in several different ways to reach the goal, but often only one way may thus be described in each HTA diagram. When the interaction analysis is performed, only one of the correct ways in which a user can perform a task is considered. The chosen correct sequence can be the common use, intended use, or risky use dependent on the scope of the evaluation. Moreover, the grouping of operations and functions may vary for the same task. It is therefore important that the design of the HTA diagram corresponds as closely as possible to the sequences of action that occur in reality.

Context description

The third part of the system description deals with the context. The user performs the task(s) in an environment which is outlined in the context description. It specifies the physical, organisational and psychosocial environment during use. The purpose is to map the external performance-shaping factors that affect the user (e.g. lighting, noise) and context characteristics (e.g. culture, organisation, location) that affect the interaction.

First the environmental and organisational environments are described in general terms. Then the performance-shaping factors in the environment are identified and how they affect the usage and the user. Added to the context description a graphic system model could be drawn presenting the elements affecting the human and the machine, and their interaction.

Interaction description

The fourth part in the system description of the human-machine system is the interaction between the human and the machine. The interaction description describes the appearance of the machine during the analysed tasks and the way the user interacts with the machine. Given the correct way (the intended use proposed by the manufacturer) to perform a task as described in the HTA diagram, the way machine appears to the user during different actions should then be stated. In this way it becomes possible to evaluate the appearance of the machine (including the user interface) throughout its use. The interaction description can be compiled in a simple way by combining the appearance of the machine (for example using screen dumps) with the HTA diagram. Another usable method for describing the interaction is Link Analysis (LA) (Stanton and Young, 1999). Link analysis can be used to graphically describe how often and in which order an operator uses a machine's controls and displays while performing a task. The method can also be used to describe the user's eye movements and thus understand where attention is focused during work. The link analysis is done by a stepwise walkthrough of the task, with each sub-step plotted on a schematic diagram of the user interface of the machine. Another more advanced way to describe the human-machine interaction is to use the User-Technical Process (UTP), suggested by Janhager (2005).

UTP (Figure 5.3) is used to describe the interplay between the user and the technical system. UTP shows parallel processes, namely the user process and the technical process. The user process consists of two parts: *mental activities* and *user actions*. The mental activities illustrate the user's thoughts and feelings, while the user actions describe what operations a user performs to reach the goal of the use. The user actions correspond to the lowest level (= operations) in the HTA.

The technical process also consists of two parts: *interface functions* and *technical functions*. The interface functions describe the appearance of the user interface (sending and receiving signals), and the technical functions describe the actions that the machine performs to contribute to a certain purpose.

The specification of the interaction must be done on the level of detail required for the subsequent analysis. There is no general rule for the level of detail needed – it is up to the analyst to decide. Each operation has to be described in such a way that the interaction between user and machine can be evaluated. Both the user's actions and the machine's responses to actions must be specified in equal detail.

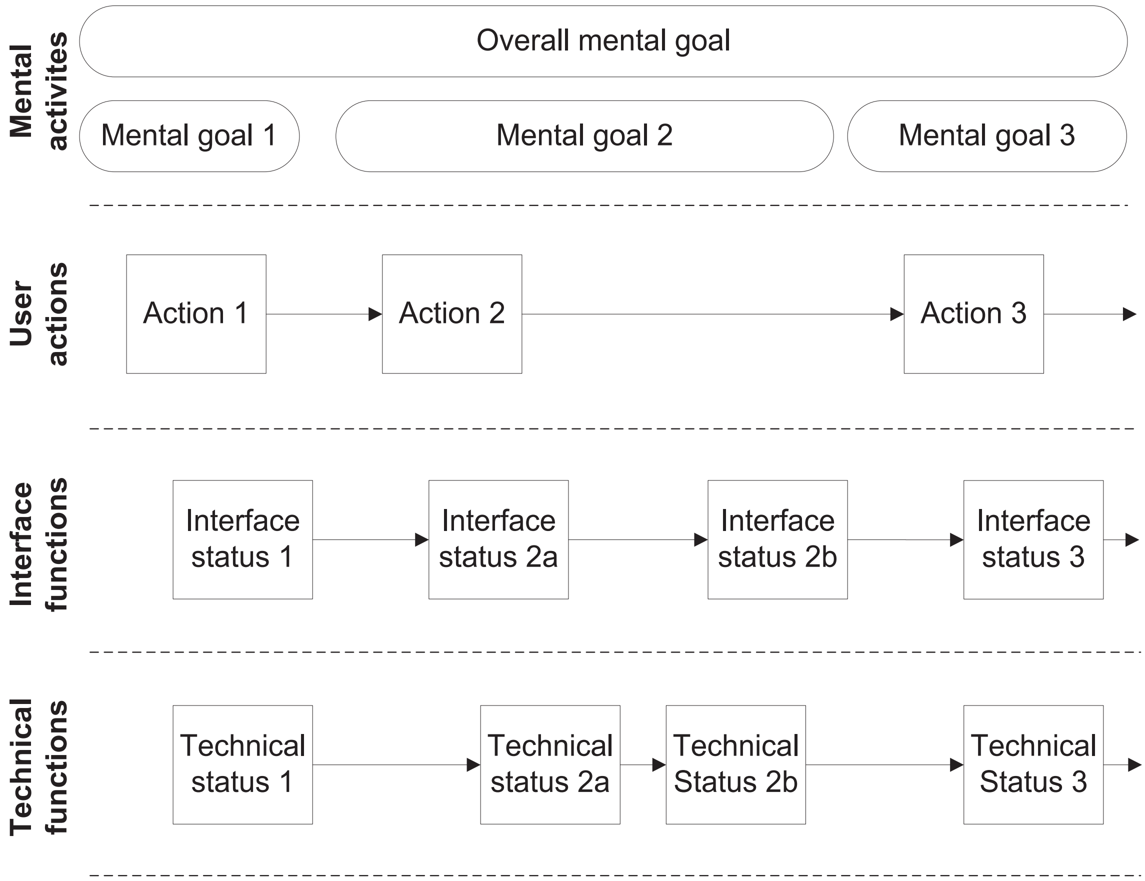


Figure 5.3 User Technical Process (UTP) from Janhager (2005)

5.1.4 Workload analysis

The third phase of CCPE methodology is workload analysis (Figure 5.1) which is performed with the developed method, Generic Task Specification (GTS) (Paper VII). The purpose of the work load analysis is to map automation levels, task demands, mental workload and the physical workload which may affect the interaction between human and machine. The aspects are divided into four parts:

(1) task demands
(2) automation levels
(3) mental workload,
(4) physical workload

When using the GTS framework, the purpose of further interaction analysis of the human-machine system is to determine the selection of aspects presented in output data matrices. If, for example, the focus of the further analysis is on cognitive demands, then aspects regarding mental workloads are categorised and aspects of physical workload could thus be excluded, and vice versa. However, this exclusion should be performed with great caution. It is important to fully understand if one of these aspects does not play an important role in the interaction. Exclusion of an aspect also counteracts the fundamental idea of the CCPE methodology, that of combined evaluation.

Categorisation and grading within each aspect may also be adapted to the specific human-machine system studied. Categories and the number of grades can be changed to better fit the overall purpose of the evaluation. However, it is important that the specific meaning of each category and grade is defined for each use.

Task demands

The first part of the work load analysis is to describe the task demands. Task Demands refers to the demands that the task make on the user for each operation in the HTA. The demands from a task/operation are independent of the user who performs the task/operation. Task demands are dependent on the task, but also on the interacting machine and the environment. If multiple machines or environments are studied for the same task, they require separate sets of task demands. The task demands are described by five aspects (type, category, accuracy, time, performance).

<u>(1) Task Type (TT):</u> The operations are classified into categories.

- **Action (A)** - e.g. pressing a button or pulling a switch
- **Retrieval (R)** - e.g. getting information from a screen or manual
- **Checking (C)** - e.g. conducting a procedural check
- **Selection (S)** - e.g. choosing one alternative over another
- **Communication (I)** - e.g. talking to another party

<u>(2) Task Category (TC):</u> Each operation is classified according to generic task categories, i.e. in this case if the operation is machine-specific or also common in other machines.

- **Common (L)** The operation is common in the use of the machine
- **Regular (R)** The operation is regular in the use of the machine
- **Uncommon (O)** The operation is uncommon in the use of the machine

<u>(3) Performance/Accuracy (P/A):</u> The accuracy of the performance.

- **High (H):** The task must be performed with high fine-motor ability and/or precise mental processing
- **Medium (M):** The task must be performed with some fine-motor ability and/or medium mental processing
- **Low (L):** The task can be performed with low accuracy

<u>(4) Time Pressure (TP):</u> How much time pressure does the user feel due to the rate or pace at which the tasks occur?

- **High (H):** Almost never has spare time. Interruptions or overlap among activities are very frequent, or occur all the time.
- **Medium (M):** Occasionally has spare time. Interruptions or overlap among activities occur frequently.
- **Low (L):** Often has spare time. Interruptions or overlap among activities occur infrequently, or not at all.

(5) Performance-Shaping Factors (PSF): Which additional factors might influence the performance of the user during execution of the tasks? The factors can either be external or internal stressors, or both, for example air temperature, fatigue or working colleges. No classification of these factors is needed, only points of view should be written in free text, since there exist a plethora of possible Performance-Shaping Factors. Lists of plausible PSFs have been written by Kim and Jung (2003), among others.

Data for the aspects can be gathered by interviews with users, by observations of use, and by use of more advanced methods such as Applied Cognitive Task Analysis (ACTA) (Militello and Hutton, 1998).

The aim of ACTA is to elicit information about the cognitive demands and skills required for performing a task. An ACTA is composed of three combined interview procedures. The structure of ACTA facilitates transformation of the information more directly into applied machines, such as user interface design or training scenarios.

Automation levels

The second part of the work load analysis is to map the automation level. Automation describes the allocation of the task between the human and the machine. The classification of automation is performed for each operation and is independent of the user who performs the task/operation. The automation levels are described in seven categories:

(1) Power (P) What/who provides the power in performing the task?

- **High (H):** The machine provides almost all the power
- **Medium (M):** Both human and machine provide the power
- **Low (L):** The human provides almost all the power

(2) Data Collection (DC) What/who collects the data from the process when performing the task?

- **High (H):** The machine collects almost all the data
- **Medium (M):** Both human and machine provide the data
- **Low (L):** The human provides almost all the data

(3) Data Analysis (DA) What/who analyses the collected data when performing the task?

- **High (H):** The machine performs almost all the data analysis
- **Medium (M):** Both human and machine perform data analysis
- **Low (L):** The human performs almost all the data analysis

(4) Decision Making (DM) What/who makes the decision based on the analysed data?

- **High (H):** The machine makes the decision
- **Medium (M):** Both human and machine make the decision
- **Low (L):** The human makes the decision

(5) Control (CO) What/who controls the performance of the task?

- **High (H):** The machine controls the performance of the task
- **Medium (M):** Both human and machine control the performance of the task
- **Low (L):** The human controls the performance of the task

(6) Execution (EX) What/who executes the task?

- **High (H):** The machine executes the task
- **Medium (M):** Both human and machine execute the task
- **Low (L):** The human executes the task

(7) Supervision (S) What/who supervises the performance of the task?

- **High (H):** The machine supervises the task
- **Medium (M):** Both human and machine supervise the task
- **Low (L):** The human supervises the task

Data for the seven automation categories can be gathered via further analysis of the interaction description and by studying the actual or proposed human-machine system.

Mental workload

The third part of the work load analysis is to analyse the mental workload. The mental workload describes the load that the user's information process is exposed to when performing the tasks. The classification of mental workload is made for each operation and is dependent on the user who performs the task/operation, which is described by the user profile. This is in contrast to task demands, where the classification is independent of the human who performs the task/operation. The mental workload is described by six aspects:

(1) Mental Processing Type (MPT): This aspect describes the mental level at which the operations are performed. Each operation is classified into categories.

- **Skill-Based Processing (SB)** Automatic actions in response to a stimulus (unintentional)
- **Rule-Based Diagnostic Processing (RBD)** Diagnostic-based, previously learned rules (deliberate or unintentional)
- **Rule-Based Action Processing (RBA)** Behaviour-based, previously learned rules (deliberate or unintentional)
- **Knowledge-Based Processing (KB)** Problem-solving and judgement (deliberate)

(2) Attention Resources (AR): This aspect shows how much perceptual attention activity is required when performing the task, e.g. when examining, searching, and monitoring data.

- **High (H):** High attention effort and concentration necessary. The activity requires extensive attention.
- **Medium (M):** Medium attention required.
- **Low (L):** Low attention is required. The activity is almost automatic, requiring minor or no attention.

(3) Memory Resources (R): This activity shows how much memory activity is required when performing the task, e.g. when thinking, deciding, calculating, remembering, and searching for information.

- **High (H):** High memory resources necessary
- **Medium (M):** Medium memory effort required
- **Low (L):** Low memory resources required

(4) Processing Resources (PrR): This aspect deals with the complexity of decision-making and shows how much processing activity is required when performing the task, e.g. thinking, deciding, calculating, remembering, and searching for information.

- **High (H):** High mental processing effort and concentration necessary. Very complex activity.
- **Medium (M):** Medium mental effort or concentration required. Complexity of activity is moderate due to uncertainty, unpredictability, or unfamiliarity.
- **Low (L):** Low conscious mental effort or concentration required. The activity is almost automatic.

(5) Frustration and Stress (FS): How insecure, discouraged, irritated, stressed or annoyed, versus secured, gratified, content, relaxed and complacent, is the user when performing the task?

- **High (H):** High stress due to confusion, frustration or anxiety. High to extreme determination and self-control required.
- **Medium (M):** Medium stress due to confusion, frustration, or anxiety noticeably adds to the workload. Significant compensation is required to maintain adequate performance.
- **Low (L):** Low confusion, risk, frustration, or anxiety exists and can be easily accommodated.

(6) Superimposed Mental Activities (SMA): What does the user think about when performing the task? Superimposed mental activities can for example be the overall aim for the user, machine parameters which need continuous monitoring, or selection of the strategy for the next main task. No classification needed, points of view just written in free text.

Data for the mental workload aspects can be gathered by interviews with users and observations of use, and by use of more advanced methods such as Subjective Work Load Assessment (SWAT) (Wilson and Corlett, 1995) and NASA Task Load Index (Hart and Staveland, 1988).

Physical workload

The fourth part of the work load analysis deals with the physical workload. The physical workload describes the load that the user's body is exposed to when performing the task. The classification of physical work load is made for each operation and is dependent on the user who performs the task/operation, who is described by the user profile. This is in contrast to task demands, where the classification is independent of the user who performs the task/operation. The physical workload is described by five aspects:

(1) Force Resources (FR): How much force is required when performing the task, e.g. when pushing, pulling, and turning?

- **High (H):** High effort required to produce sufficient force
- **Medium (M):** Medium effort required to produce sufficient force
- **Low (L):** Low effort required to produce sufficient force

(2) Fine-Motor Resources (MR): How much fine-motor ability is required when performing the task, e.g. when pushing, pulling, and turning?

- **High (H):** High fine-motor ability required
- **Medium (M):** Medium fine-motor ability required
- **Low (L):** Low fine-motor ability required

(3) Speed Resources (SR): How much speed is required when performing the task, e.g. when pushing, pulling, and turning?

- **High (H):** High speed required
- **Medium (M):** Medium speed required
- **Low (L):** Low speed required

(4) Body Loads (BL): This aspect shows which parts of the body are affected by physical loads in the operation. The body is divided into different parts, e.g. (a) Neck, (b) Shoulders, (c) Arm and Elbow, (d) Hand and Wrist, (e) Upper Back, (f) Lower Back, (g) Leg and Knee and (h) Foot and Ankle.

- **High (H):** High body load on the zone
- **Medium (M):** Medium body load on the zone
- **Low (L):** Low body load on the zone

(5) Body Contact (BC): If there is a possibility that various body parts come into contact with the machine or environment unintentionally, this may result in discomfort and injury. No classification, points of view just written in free text.

Data for the aspects can be gathered through interviews with users and observations of use, and by use of methods for load assessment such as Rapid Upper Limb Assessment (RULA) (McAtamney and Corlett, 1993) and Rapid Entire Body Assessment (REBA) (Hignett and McAtamney, 2000). Biomechanical calculations can also be made, or measurement of muscular strength can be performed using electromyography.

Result compilation

The fifth part in the work load analysis is the creation of matrices. The output data from GTS is semi-quantitative so the results can be displayed in different types of matrices. In this way, the workload and demands can be compared with each other or assessed separately. By comparing the mental and physical workload with the task demands and the user profile, presumptive mismatches between human and machine can be detected, e.g. high short-term memory load or high physical static loads on the shoulders. The result of the work load analysis is thus the basis for the interaction analysis as it describes factors that may affect the interaction.

5.1.5 Interaction analysis

The fourth and final phase in CCPE methodology is interaction analysis (Figure 5.1), where the interplay between human and machine is evaluated. Interaction analysis is based on the correct handling sequences described in the HTA diagram. In the analysis, the HTA is divided into two parts: operations and functions. Operations are the lowest level in the HTA and functions are the sub-tasks in the HTA. For each function and operation, questions are asked to identify presumptive mismatches. The question process tries to simulate how the user interacts with the machine, with the analysts in the evaluation team playing the role of the user.

The interaction analysis consists of three parts: (1) usability problem analysis, (2) use error analysis, and (3) physical ergonomic error analysis. The interaction analysis results in a list of identified human-machine mismatches. The mismatches are also graded from low importance to high importance. A grading system makes it easier to determine what it is most important to rectify in the subsequent redesign of the machine. If suitable the interaction analysis can be supplemented with analysis of physical ergonomic problems, e.g. with methods previously mentioned such as REBA and RULA. However, this is outside the scope of the CCPE methodology.

Usability problem analysis (ECW)

The first part of interaction analysis is to analyse usability problems in the human-machine system. The usability problem analysis is performed by the Enhanced Cognitive Walkthrough (ECW) method (Paper II). ECW is an analytical method which looks into potential usability problems by investigating what prevents the user from performing correct actions and why that happens. ECW uses a detailed procedure to simulate the user's problem-solving process at each step of the interaction between human and machine. There is a continual check of whether the user's goal and knowledge can lead to the next action being correctly executed. Enhanced Cognitive Walkthrough (ECW) is based on the third version of CW, presented by Lewis et al. (1997), which is one of the variants of the CW method (Mahatody et al., 2010).

Prediction of usability problems

To predict usability problems, the analyst works through the question process in ECW for all the selected tasks. The interaction analysis is based on the described correct handling sequences in the HTA. The question process then generates conceivable usage problems. The question process is divided into two levels of questions (Table 5.1). The first (level 1) is employed for the nodes in the HTA, and the second (level 2) for the operations in the HTA. In level 1, the machine's ability to "capture" the user is studied, and in level 2 its ability to lead the user to perform the function correctly is studied.

The analysis begins with the analyst asking the questions on level 1 for the uppermost node in the HTA diagram (Figure 5.2). The analysis then continues downward through the HTA diagram, where the analyst employs questions on level 1 for the nodes and questions on level 2 for the operations furthest down in the tree. The underlying nodes/operations of a given node are analysed completely before the analysis proceeds to the adjacent node.

Table 5.1 Analysis questions for ECW

	Level 1: Analysis of tasks/functions	Level 2: Analysis of operations
1	Will the user know that the evaluated function is available? *Does the user expect, on the basis of previously given indications that the function exists in the machine?*	Will the user try to achieve the right goals of the operation? *Does the user expect, on the basis of previously given indications, what is to be performed?*
2	Will the user be able to notice that the function is available? *Does the machine give clues that show that the function exists?*	Will the user be able to notice that the action of the operation is available? *Does the machine give clues that show that the action is available and how to perform it?*
3	Will the user associate the clues with the function? *Can the user's expectations and the machine's indications coincide?*	Will the user associate the action of the operation with the right goal of the operation? *Can the user's assumed operation and the machine's indications coincide?*
4	Will the user get sufficient feedback when using the function? *Does the machine give information that* the function has been chosen and *the position the user is at in the interaction?*	Will the user be able to perform the correct action? *Do the abilities of the user match the demands by the machine?*
5	Will the user get sufficient feedback to understand that the function has been fully performed? *Does the user understand, after the performed sequence of actions, that the right function has been performed?*	Will the user get sufficient feedback to understand that the action has been performed and the goal has been achieved? *Does the user understand, after the performed operation, that he/she has done it correctly?*

Grading of the answers

Each question is answered with a grade (a number between 1 and 5) (Table 5.2) and a justification for the grade. These justifications, called *failure/success stories,* are the assumptions underlying the choice of grades, such as that the user cannot interpret a displayed symbol. The grading, called *problem seriousness,* from 1 to 5 represents different levels of success (Table 5.2). Grading makes it easier to determine what it is most important to rectify in the subsequent reworking of the machine.

Table 5.2 Grading of the failure/success stories

Grade	Grade in words	Explanation
5	Yes	A very good chance of success
4	Yes, probably	Probably successful
3	Do not know	Impossible to decide if successful or not
2	No, uncertain	Small chance of success
1	No	A very small chance of success

During prediction of usability problems, each question is answered – assuming that the preceding questions are answered with YES (grade 5) – irrespective of what the real answer was for the last question. In certain cases, however, the questions may be impossible to answer, and these must be marked with a dash in the protocol.

Identification of problem

The next step is to identify the predicted problems. If the problem seriousness is between 1 and 4, i.e. not with "a very good chance of success", it points to the existence of a potential usability problem. Based on the failure story, the usability problem is then described. The problem is the cause which prevents the user from performing the correct action.

Table 5.3 Problem types employed in ECW

Problem Type	**Explication**	**Origin**
User (U)	The problem is due to the user's experience and knowledge, possibly because the user is accustomed to different equipment	Comes primarily from questions 1 and 3
Hidden (H)	The interface gives no indications that the function is available or how it should be used	Comes primarily from question 2
Text and icon (T)	Placement, appearance and content can easily be misinterpreted or not understood	Comes primarily from question 3
Sequence (S)	Functions and operations have to be performed in an unnatural sequence	Comes primarily from question 1
Physical demands (P)	The interface has too high demands on the user's physical speed, motoric and force	Comes primarily from question 4 (operation level)
Feedback (F)	The interface gives unclear indications of what the user is doing or has done	Comes primarily from question 4 (function level) and question 5

Each problem is further categorised by a problem type. The categorisation stems from the failure stories and the description of the problem. Depending on the machine and the task that the user is to solve with it, different problem types can be used. Suggestions of problem types are given in Table 5. 3.

Prediction and identification of usability problems are conducted in parallel, i.e. a problem is investigated in depth immediately after being identified. During the analysis templates are used, one for each operation and function (Appendix F). The results of the identification are then later reported in a tabular form.

Compilation in matrices

Matrices are employed to present the semi-quantitative results from the analysis. The collected answers from the prediction and identification are ordered in different ways within the matrices so as to emphasise different aspects of the analysis. The matrices can be combined in various ways and the numbers in the matrix cells show the number of detected problems distributed according to the two types of data that are compared. The information employed from the ECW consists of: *task number*, *task importance*, *problem severity* and *problem type*. The matrices can be combined in several ways (Table 5.4).

The five different matrices in Table 5.4 describe the perception of problems with the machine in different ways. The numbers in the matrices show how many problems exist in the specific combination of analysis results relating to the entire problem complex. Since the matrices only exhibit the same problem complex in different ways, the sum of the numbers is the same in all matrices belonging to a given machine.

Matrix A (problem seriousness versus task importance) shows if there are serious usability problems with the machine that can prevent its intended use. If there are many problems in the upper left corner of the matrix, this means that serious problems exist in important tasks. If the problems are in the lower part of the matrix, they come from less important tasks, and if the problems are found in the right part of the matrix they are not so serious. An example of Matrix A is shown in Figure 5.4. The figure shows that there are three problems with task importance 4 and problem seriousness 3 (the cell with diagonal lines).

Table 5.4 Matrices for presenting the results from the ECW analysis

Name	**Content**	**Explanation**
Matrix A	Problem seriousness versus task importance	*This shows the machine's general condition*
Matrix B	Problem seriousness versus problem type	*This shows the overall problems with the machine*
Matrix C	Problem type versus task importance	*This shows which problems are most important to rectify*
Matrix D	Problem seriousness versus task number	*This shows which tasks have the most problems*
Matrix E	Problem type versus task number	*This shows which types of problems are most common in the tasks*

Matrix B shows problem seriousness versus problem type. This kind of matrix provides an overview of what sorts of problems exist in the machine and how serious they are. Such a matrix may, for instance, show that most of the problems concern marking of buttons, but that the most serious problems have to do with feedback. By studying the numbers in each matrix, one can find patterns, see how serious the problems are, and understand which types of problems stem from the design of the machine.

	Problem Seriousness			
Task Importance	1	2	3	4
1	0	0	0	1
2	0	1	1	8
3	2	2	8	1
4	1	2	3	5
5	1	0	0	0

Figure 5.4 Example of Matrix A – Problem seriousness versus task importance

Matrix C (problem type versus task importance) shows which problems are most common in the most important tasks. Matrices D and E reveal more specifically how serious the problems are that occur in each task and what types of problems they are.

Alarm analysis (Alarm-ECW)

The Alarm-ECW method is a special version of Enhanced Cognitive Walkthrough with focus on alarm signals and alarm messages (Paper XII). The difference is that the question on level 2 is replaced with a new set of questions (Table 5.5). In the Human-machine system description the characteristics (visual, audible and haptic) of the investigated alarms, the task (how to attend to the alarm) and the operator all need to be described.

The usability analyst(s) answer(s) six questions (Table 5.5) after taking on the persona of the operator, i.e. the analyst puts himself/herself in the operator's situation. If applicable, questions 1 and 2 are answered on the assumption that the operator is located at a distance from the alarming device. Questions 3 to 6 are answered on the assumption that the operator is next to the alarming device.

Table 5.5 Analysis questions for Alarm-ECW

Question	**Explanation**
1. Will the operator be able to detect that there is an alarm condition?	*Is the operator likely to notice the alarm?*
2. Will the operator understand the seriousness of the alarm?	*Does the alarm alert the operator in a way that matches the alarm's priority?*
3. Will the operator be able to identify the alarm?	*Is it possible to distinguish this alarm from other alarms?*
4. Will the operator be able to interpret the alarm?	*Does the operator understand the cause of the alarm?*
5. Will the operator associate the correct measure/action with the alarm?	*Does the alarm guide the operator to the necessary corrective action?*
6. Will the operator get sufficient feedback to understand that the alarm has been attended to correctly?	*Can the operator understand, after taking corrective actions, that he/she has performed the appropriate measure(s)?*

The process after that is the same as for ECW. The answer from each question is called a *success/failure story* and it is a justification of why, or why not, the user/operator will be successful. From the success/failure story a *problem seriousness (PS)* rating is derived for each question. The problem seriousness rating is a classification of how successful the operator will be in attending to the alarm correctly.

If problem seriousness is between 1 and 4, it points to the existence of a presumed *usability problem*. The usability problem is described based on the failure/success story. The problem is the causes which restrain the user from performing the correct action and it is further categorised into a *problem type (PT)*.

Further, alarms usually have a priority which depends on how critical the alarm condition is. For example, in medical technology one way to prioritise alarms is with low, medium or high priority. In the ECW-method this priority is referred to as *alarm importance*. During the analysis templates are used, one for each alarm (Appendix F).

Use error analysis (PUEA)

The second part of the interaction analysis is the use of error analysis. Here the aim is to predict and identify presumptive use errors in the interaction. Use error analysis is performed by the Predictive Use Error Analysis (PUEA) (Paper III) method. PUEA is a theoretical analysis method for Human Error Identification. It is based on the methods of Action Error Analysis (AEA) (Harms-Ringdahl, 2001), Systematic Human Error Reduction and Prediction Approach (SHERPA), (Stanton and Baber, 2005) and Predictive Human Error Analysis (PHEA) (Embrey, 1986). PUEA employs a detailed process for breaking down the user's tasks into steps and, for each step, then predicting and identifying potential errors of use.

Prediction of use errors

To predict use errors, the analyst works through all the selected tasks. The interaction analysis is based on the correct handling sequences described with an HTA. To predict potential incorrect actions, a question process is employed. The question process is divided into two levels of questions (Table 5.6). The first (level 1) is employed for the nodes in the HTA, and the second (level 2) for the operations in the HTA. On level 1, use errors are identified that may arise when actions are performed at the wrong time or in the wrong order. On level 2, use errors are identified that may occur in the individual action.

Table 5.6 Analysis questions for PUEA

Level 1: Analysis of tasks/functions	Level 2: Analysis of operations
What happens if the user performs an incomplete operation or omits an operation?	What can the user do wrongly in this operation?
What happens if the user performs an error in the sequence of operations?	What happens if the user performs the operation at the wrong time?
What happens if the user performs functions/tasks correctly but at the wrong time?	

Guided by the questions, the analysts try to predict as many use errors as possible that can arise in the human-machine interaction. Each predicted use error is noted in a list. During this process, they also eliminate errors that are considered too unlikely to occur. This elimination is done in relation to how the simulated user is expected to make decisions and perform, in view of the machine and the social, organisational and physical contexts. However, it is important to be careful about dismissing without further investigation improbable errors that would have serious consequences, as these can also constitute a hazard. If there are no use errors corresponding to the answers to the questions, this also should be noted.

The prediction of use errors begins with the analysts asking the questions on level 2, for the operation furthest down to the left in the HTA tree. When all operations in that HTA branch have been analysed, the analysis continues to the node above, which is analysed with the questions on level 1. Next, the analysis shifts to the operations that lie below the adjacent node to the right of the previously analysed node. These operations are then analysed on level 2. The analysis continues similarly, i.e. the nodes/operations below a given node are analysed completely before the analysis proceeds to the adjacent node.

Use error identification

For each predicted use error, an investigation is made of eight items: (1) error type; (2) error cause; (3) primary consequence of the error; (4) secondary consequence of the error; (5) error detection; (6) error recovery; (7) protection from consequences of use error; and (8) prevention of use error. The first two concern the error itself, the next two its potential consequences, and the last four items concern mitigations of the errors and consequences. Four of the items also contain a categorisation (1 and 2), a judgment of probability (5), or a judgment of severity (4). This is done to facilitate the compilation and assessment of the investigation.

(1) Error type. For the description of the use error that was given during the identification (questions from Table 5.7), the error is also categorised according to the description in Table 5.7. The classification of the type of use error is based on proposals from Baber & Stanton (1996) and Embrey (2004).

Table 5.7 Error types for PUEA

Error Type		**Explanation**
Plan	**P1**	Plan preconditions ignored
	P2	Incorrect plan executed
	P3	Correct but inappropriate plan executed
	P4	Correct plan executed but too soon/too late
	P5	Correct plan executed in wrong order
Action	**A1**	Action too long/short
	A2	Action mistimed
	A3	Action in wrong direction
	A4	Action too little/too much
	A5	Misalign
	A6	Right action on wrong object
	A7	Wrong action on right object
	A8	Action omitted
	A9	Action incomplete
	A10	Wrong action on wrong abject
	A11	Unnecessary action
Checking	**C1**	Checking omitted
	C2	Check incomplete
	C3	Right check on wrong object
	C4	Wrong check on right object
	C5	Check mistimed
	C6	Wrong check on wrong object
Retrieval	**R1**	Information not obtained
	R2	Wrong information obtained
	R3	Information retrieval incomplete
Communication	**T1**	Message not transmitted
	T2	Wrong message transmitted
	T3	Message transmission incomplete
Selection	**S1**	Selection omitted
	S2	Wrong selection made

(2) Error cause. The error cause describes why the user is performing a use error. The error is also categorised according to a classification adapted from GEMS (Generic Error Modelling System) (Reason, 1990) (Table 5.8). This categorisation is made according to the criteria described by Reason (1990).

Table 5.8 Error causes classification

Error Cause	Explanation
(L) Lapse	A memory lapse, forgetting the intention. Why am I doing this? '*Forget plan or execution'*.
(S) Slip	Failure of attention during execution. A correctly planned action is not correctly executed. '*Good plan, bad execution'*.
(R) Rule-based mistake	Occurs during problem-solving of familiar situations. Misapplications of good rules, i.e. well-known rules are used incorrectly to make a decision. '*Bad plan, good execution'*.
(K) Knowledge-based mistake	Occurs during full attention to problem-solving activities, or problems never encountered before. Wrong decision based on own conclusions drawn from prior knowledge and known rules. *'Wrong conclusions, correct execution'*
(V) Violations	Intended act or omission of act that violates present regulation and/or instruction, e.g. braking rules. Error action can be cutting corners to save time, omitting safety checks etc.

(3) Primary consequence of the error. Here the primary consequence of the use error is noted, i.e. the direct effect of the error on the machine. Primary consequences can be described as what happens in the underlying technical system after the error and how the machine reacts to the error.

Table 5.9 Grading of Secondary Consequence

Grade	Grade in words	Explanation
1	Disastrous	Death, loss of function or permanent impairment or damage to body structure
2	Major	Permanent impairment or damage to body structure
3	Moderate	More severe injury requiring medical treatment
4	Minor	Minor reversible injury
5	Negligible	Inconvenience or possibly minor reversible injury

(4) Secondary consequence of the error. Secondary consequences are the effects of errors that can lead to a hazardous situation for the user or other people, or to the risk of machine damage or economic loss. Secondary consequences are graded on a five-point scale. Table 5.9 shows a scale adapted for injury to humans.

Table 5.10 Grading of Detection of the Error

Grade	Grade in words	Explanation
1	**Improbable**	Extremely difficult to detect
2	**Remote**	Difficult to detect
3	**Occasional**	May be detected
4	**Reasonable**	Likely to be detected
5	**Frequent**	Most often or always detected

(5) Detection of the error. Investigation as to whether the user will detect an already committed error before it has any secondary consequences. Also reports on which technique or method the machine employs to indicate to the user that a use error has occurred. The probability that the user will detect the error is graded according to Table 5.10.

(6) Error recovery. Investigation as to whether an already committed error can be corrected, i.e. whether the user has any possibility of undoing the error before any secondary consequences occur.

(7) Protection from consequences. Investigation of which measures the system employs to protect the user and the environment from the consequences of the error. The measures may be physical barriers or alarm systems.

(8) Prevention of error. Investigation of which measures the technical system employs to prevent the use error from occurring. Preventive measures may be anything from a "dead-man's grip" to warnings in the manual.

Prediction and identification of use errors are conducted in parallel so that an error is investigated immediately after being identified. During the analysis templates are used, one for each operation and function (Appendix F). The results of the identification are then later reported in tabular form.

Compilation in matrices

Matrices are used for presenting the semi-quantitative results from the analysis part of PUEA. The answers from the investigation are arranged in different ways in the matrices to highlight different aspects of the analysis. The matrices show the sum of use errors on the basis of the ways in which the use errors have been numbered, graded or categorised – that is, in terms of task number, error type, error cause, detection, and secondary consequence.

The matrices can be combined in several ways (Table 5.11). The number in each cell of these matrices indicates how many problems have been detected when two types of semi-quantitative results are compared. Table 5.11 describes ten possible matrices, A-J, which can be produced from the semi-quantitative results. For example, matrix A shows how many use errors exist in each task, and the consequence of the errors.

Table 5.11 Matrices for presenting the results of the PUEA analysis

Name	**Content**	**Explanation**
Matrix A:	Consequence versus task	*Shows the tasks with the most serious error consequences*
Matrix B:	Error type versus task number	*Shows which type of use error exists in the various tasks*
Matrix C:	Error cause versus task number	*Shows the causes of the use errors in the different tasks*
Matrix D:	Error type versus secondary	*Shows which error type gives rise to the highest risks*
Matrix E:	Error cause versus secondary	*Shows which error cause gives rise to the highest risks*
Matrix F:	Error cause versus error type	*Shows the coupling between error cause and error type*
Matrix G:	Detection versus task number	*Shows in which tasks errors are difficult to detect*
Matrix H:	Detection versus error type	*Shows which type of error is difficult to detect*
Matrix I:	Detection versus error cause	*Shows the causes of errors that are difficult to detect*
Matrix J:	Detection versus secondary	*Shows severity of consequences for errors that are difficult*

Which matrices are produced depends on the overall purpose of using PUEA. Matrices A, B, C and G report information connected with each task, while the other matrices show the relation between the types of semi-quantitative data from the error analysis. Matrix J illustrates an aspect of hazard with the machine, since it reports the probability of detecting errors as compared with the seriousness of the consequences. The total risk cannot be illustrated as PUEA does not investigate the probability that a specific error will occur. Matrix D presents error type versus consequence seriousness. This matrix shows the severity of each type of error. If the numbers are found in the left side of the matrix, the errors have severe consequences. If the numbers are found in the right side of the matrix, the consequences are not so severe.

Matrix H presents error type versus detection probability. The matrix shows how probable it is that errors of different types are detected. If the majority of the numbers are found on the left side of the matrix, it means that the errors are hard to detect. On the contrary, if numbers are found on the right side of the matrix, the errors are easily detected before any serious consequences occur.

	Secondary Consequences				
Detection	1	2	3	4	5
1	0	0	6	0	3
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	10	0	0

Figure 5.5 Example of matrix for detection versus secondary consequences

Matrix J (Figure 5.5) shows error detection probability versus error consequence seriousness. If the majority of the numbers are found in the left side of the upper corner of the matrix, the error consequences are severe and the errors are very hard to detect. On the contrary, if numbers are found in the right lower corner of the matrix, the error consequences are not severe and error detection is probable. The matrix uses grey markings to make it easier to read and to show which errors are serious.

Ergonomic error analysis

The third and last part of the interaction analysis is ergonomic error analysis, which studies if the interaction will be performed in an ergonomically suitable way (both physical and cognitive). Ergonomics errors are analysed with the Predictive Ergonomic Error Analysis (PEEA) (Paper VI) method. PEEA is a method which investigates if a work task is performed in an ergonomically correct way and if the task can be performed incorrectly. The method is based on two well-known methods from the field of usability and human reliability: Cognitive Walkthrough (Lewis and Wharton, 1997) and Predictive Human Error Analysis (Embrey, 1986).

PEEA is applied on the operation level (level 1 according to ECW and PUEA) of the HTA and works with questions in two steps. First, the method examines if the actions will be performed in a correct ergonomic way, and secondly if the actions can also be performed in an incorrect ergonomic way. For each identified non-ergonomic action (ergonomic error) a further examination can be undertaken.

Prediction of physical ergonomic problems and errors

The first step for every operation in the HTA is to describe how it can be performed in a correct/good ergonomic way. The next step is to examine if the actions will be performed in a correct ergonomic way and if they can also be performed in an incorrect ergonomic way. This investigation is based on four questions (Table 5.12). The results are listed in a table.

Table 5.12 Identification questions for PEEA

Problem identification	**Error identification**
Does the product give any information (cues) about how the action can be performed in an ergonomically correct way?	Can the action be performed in a non-ergonomic way? How?
Does the user know how the action can be performed in an ergonomically correct way?	
Will the user try to perform the action in an ergonomically correct way?	

Investigation of physical ergonomic errors

Every identified non-ergonomic action (ergonomic error) is then further investigated using the following four questions (Table 5.13). The analysis of the consequences can be based on the expert knowledge of the analyst or on specific methods. For example, if the error relates to incorrect body posture it can be made through the use of RULA, or REBA analysis, or heuristics concerning the user's anatomy, physiology, anthropometry and biomechanics. During the analysis templates are used, one for each operation (Appendix F).

Table 5.13 Investigation questions for PEEA

Error investigation
Which are the possible causes of the action being performed in an ergonomically incorrect way?
Which are the short-term consequences for the user?
Which are the long-term consequences for the user?
Is the product designed to prevent the ergonomic error?

Question process in the interaction analysis

To ensure efficiency in the interaction analysis a number of issues need to be observed and considered. A task can often be performed in several different ways to reach the goal described in the HTA with associated plans. When the interaction analysis is performed, only one of the correct ways in which a user can perform a task is considered. The chosen correct sequence matches the common or critical real use.

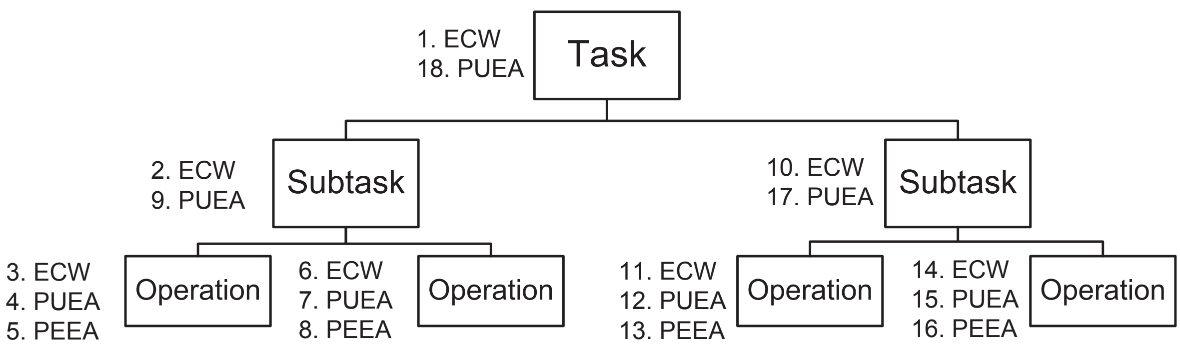


Figure 5.6 Order in which to ask questions in ECW and PUEA respectively, in the interaction analysis

The questions and the question process are most efficient if pursued in the following order. Start with ECW questions at the top of the HTA, and then use the ECW questions on the way down on the left. At the bottom of the HTA, use the PUEA questions at the Operations level and then the PEEA questions. Then work through the HTA from left to right. To sum up, apply ECW questions on the way down of the HTA, PUEA questions on the way up the HTA and the PEEA questions last at the bottom. Figure 5.6 shows, with the numbers 1-18, the type and order of questions to use when predicting usability problems and use errors.

In this way, the prediction and investigation of usability problems and use errors is conducted in parallel, i.e. a problem or an error is investigated more in-depth immediately after being predicted. The results of the investigation are then reported in tabular form.

Finally, the same use error may be identified in several operations and functions, but it only needs to be documented once. There is no need for duplicates in the resulting list of use errors.

Result compilation in interaction analysis

The result of the interaction analysis is primarily a list in tabular form of the presumptive mismatches between human and machine in the form of usability problems, cognitive use errors and physical ergonomic errors. Matrices are then employed to present the semi-quantitative results from the interaction analysis. The collected answers from the prediction and identification of errors and problems are arranged to emphasise different aspects of the analysis. The matrices can be combined in various ways, and the numbers in the matrix cells show the number of detected mismatches distributed according to the two types of data (Tables 5.4 and 5.11) that are compared.

The results obtained from the interaction analysis can also be incorporated in risk analyses of the machine. The user errors can for instance be integrated in a risk analysis on a higher system level when the Fault Tree Analysis (FTA) and Event Tree Analysis (ETA) (Stricoff, 1996) methods are used, or at a detailed machine level when the Failure Mode, Effects and Criticality Analysis (FMECA) method is used (Defence, 1980).

Result reflection in interaction analysis

Since ECW, PUEA and PEEA are all methods that analyse potential use errors and usability problems, the predicted problems and errors may not occur during a real use situation. The grading systems are also subjective judgments by the analyst(s). The detected potential problems and errors need to be reviewed and confirmed in interaction with real users before changes are made in the analysed user machine.

One way of reviewing the results is to triangulate the analytical approach with usability evaluation methods such as Heuristic Evaluation (Nielsen and Mack, 1994) and Usability Testing (Nielsen, 1993). By using these methods, the relevance of potential usability problems and use errors found can be confirmed or dismissed. Another possibility is to let actual users in a focus group discussion (Cooper and Baber, 2005) decide whether the potential problems and errors are relevant or not.

5.1.6 Methodology application

The time and resources needed for the CCPE methodology are highly dependent on the selected tasks and the characteristics of the human-machine system being evaluated. Furthermore, if the system has a low degree of usability, i.e. several design flows, this affects the evaluation process. In addition, the more complex the human-machine system, the more effort is needed to perform the evaluation. Therefore it is difficult to say how much time should be spent on a specific analysis. A plain evaluation of a simple product may require a few hours of work while an extensive evaluation of a complex product in a complex system may require a few days of work. When performing CCPE it is recommended to avoid sessions longer than two hours at a time to maintain the analysts' attention, especially in interaction analysis.

Time is also needed to learn and train in the use of CCPE methodology. If the analyst has prior basic knowledge in the area of human factors/ergonomics and has previously used evaluation methods to evaluate ergonomics or usability issues, the methodology is easy to adapt. However, as always when using a new method, some sessions are needed in order to become highly skilled in its application; consequently, a skilled analyst performs the evaluation routinely and thereby decreases the time required. Users who have basic knowledge in ergonomics and have performed evaluations before should be able to use the CPPE methodology after a one-day training course. To become a skilled moderator you have to perform at least two or three complete evaluations with the CPPE methodology to learn in practice how the various steps work together and how to adapt the methodology to a specific evaluation.

For a participant who does not have a leading role in the analyst group, it is often enough to participate for a few hours in a session with a trained session leader in order to be able to contribute to the analysis.

5.1.7 Staffing

Evaluation using CCPE methodology can be conducted by a single human factors expert, or by a group of people that may consist of people with a variety of knowledge about the actual human-machine system; i.e. designers, software developers, marketing staff, human factors experts and real life users.

The interaction analysis part can be performed by a single analyst but is best performed by a group of people, since it makes it possible to discuss presumptive user actions. If more than two persons are conducting the interaction analysis, a skilled moderator must be appointed to lead the evaluation. The role of the moderator is to keep up the pace, adapt the questions to the specific function/operation, and guide the discussion so it does not lose focus. The moderator should preferably be a person with knowledge of human factors.

To use CCPE methodology efficiently the analysts need knowledge in cognitive and physical ergonomics. For example, to employ PUEA fully the users of the method require knowledge of the Skill-Rule-Knowledge model (Rasmussen, 1983) and the Generic Error Modelling System (Reason, 1990). This renders the methods more difficult to apply than other simple evaluation methods such as CW, but in general necessary knowledge in the field of cognition is always of benefit in the usability engineering process.

No method or methodology can ever fully replace knowledge, and knowledge is required to use the results from the CCPE methodology since it focuses on details more than on the comprehensive view. Therefore, CCPE methodology cannot replace knowledge and experience of the area of human factors/ergonomics. Knowledge about cognition and human error is also needed in order to design usable machines so the entire development process is not based on "trial and error". Interaction analysis shows chiefly what is inadequate (usability problems and use errors) and what the causes are, not really how to improve the design. Hence, practical experience and theoretical knowledge about user interface design, or knowledge about the human information processes, is essential throughout a usability engineering process – which means that it is unnecessary to design a methodology for users who lack this knowledge.

Thus, the expected users of CCPE methodology are basically engineers with fundamental knowledge of cognition, cognitive ergonomics, physical ergonomics and interface design. However, there is nothing to indicate that the methods cannot be employed by other professional groups working with human-machine interaction, such as industrial designers, psychologists or physiotherapists. A discussion of the problems and possibilities associated with using methods in the area of human factors engineering in industry is presented in Andersson et al. (2011).

5.1.8 Adaptation of the CCPE Methodology

The CCPE methodology should not be regarded as a rigid framework but rather as a structure that needs to be adapted to each context for analysis. The methodology may be adapted to each application individually, but often it is recommended to standardise the methodology within a company or organisation. This makes it possible to compare different analyses with each other.

System description

Knowledge of the users as well as the tasks is fundamental in all human factors engineering work regardless of method or methodology. In the described methodology it is essential to select an appropriate user on whom to base the analysis. A single profile or multiple profiles can be used. CCPE methodology does not provide support in selecting user profiles, but it is necessary to be aware that the selection of appropriate user profiles will significantly affect the analytical result of the methodology. However, when a user profile is created, it only needs to contain aspects that are relevant for the intended use. It is important not to add too many aspects that make the profile more difficult to manage.

Other aspects that must be weighed in when describing the presumptive user include whether he/she is an expert or novice as regards handling the equipment and solving the task. For medical equipment, this is often a question of whether the user is a medical staffer (professionally trained) or the actual patient (layman). Since much equipment is developed for and sold on a global market, cultural differences must also be taken into account. Moreover, it is of interest to investigate the user's level of expertise, since this can influence his/her perception of information and the workload experienced (cf. Thunberg, 2006). However, no general way in which the user should be described through categorisation of relevant aspects, such as novice vs. expert, has yet been developed for ECW and PUEA. This is an area for further work.

Another analysis stage that the CCPE methodology does not support is the selection and grading of tasks. Here the analyst(s) has to use other methods in the area of ergonomics/human factors; i.e. interviewing the user and observing use are essential tools. It is important for some of the analysts who are going to do the interaction analysis to be present when the system description is created, in order to lodge it securely in their memory and not only documented on paper.

Input data from the artefact and context are also needed. The artefact is describable with visual aids, such as pictures of the graphic user interface and flow schemes for the logic. In addition, the methods for task analysis – HTA, LA and UTP – can describe the artefact's relation to the task. As for describing the context, no method for this has been employed in connection with ECW and PUEA. It is of interest to map the physical, social and organisational contexts that can influence the users' performance of the task. One area for further work is thus to investigate whether any suitable method of the kind exists, or to develop a method with the purpose of describing input data from the context to ECW and PUEA.

Workload analysis

In the workload analysis it is important to select which aspects to focus on. The GTS suggests a number of aspects but the analysts have to add or remove aspects to suit the evaluation. For example, if the evaluation focus is on cognitive ergonomics, some of the physical ergonomics aspects may be removed. The GTS includes a scale of three steps, but this may be altered to better fit the specific analysis.

Interaction analysis

The interaction analysis also needs to be modified to work properly. The questions in the question process are written in a general way, but when they are asked in the interaction analysis they need to be worded more specifically. The example of a light switch is used to illustrate this point (Table 5.14 and 5.15). The task is to light up a room and this is done with three switches.

Table 5.14 Rephrasing questions level 1: Analysis of tasks/functions

	Generic questions	**Rephrased questions**
1	Will the user know that the evaluated function is available?	Will the user know that it is possible to light up the room?
2	Will the user be able to notice that the function is available?	Will the user be able to notice that the room can be lit up?
3	Will the user associate the clues with the function?	Will the user associate the clues (e.g. lamps in the ceiling) with lighting up the room?
4	Will the user get sufficient feedback when using the function?	Will the user get feedback during the lighting process to understand that the lighting process is going on?
5	Will the user get sufficient feedback to understand that the function has been fully performed?	Will the user get feedback when the whole room is lit up?

Table 5.15 Rephrasing questions level 2: Analysis of operations

	Generic questions	**Rephrased questions**
1	Will the user try to achieve the right goals of the operation?	Will the user try to light the first lamp?
2	Will the user be able to notice that the action relating to the operation is available?	Will the user be able to see the switch for the first lamp?
3	Will the user associate the action relating to the operation with the right goal of the operation?	Will the user understand which switch to press to light the lamp?
4	Will the user be able to perform the correct action?	Will the user be able to press the switch?
5	Will the user get sufficient feedback to understand that the action has been performed and the goal has been achieved?	Will the user get feedback to understand the switch has been pressed and the lamp is lit?

Interaction analysis often uses a five-step scale which may need to be modified. One example is the grading of secondary consequences. If the machine does not have an obvious potential to harm the user, the highest grade should not be “Death, loss of function or permanent impairment or damage to body structure”. In this case the categories need to be rewritten. The usability problem types also need to be adapted to specific evaluation situations.

It can sometimes be hard in the interaction analysis to trace a use error or usability problem to a specific function or operation. But it is not so important that the problems/errors are assigned precisely, the important thing is rather that the problem/error is documented for further analysis and counter-measures.

5.2 Development

The section tells the story about the development of CCPE and its methods.

5.2.1 Choice of approach

The methodology and its practices were developed in the course of a number of projects which are presented in chapter 5.3. The reason why the method was even developed in the first place was because the methods used initially in the projects were not seen as good enough, or because there was a lack of methods for what had to be evaluated. In recent times, the need for a coherent methodology has driven the method development even further. The actual development can be described by the spiral process described in Chapter 3 (Figure 3.1). Spiral processes have been governed partly by the established theory and requirements (chapter 4) concerning the methods, and partly by the deficiencies found in the methods. Often the requirements and the deficiencies emerge simultaneously. The deficiencies are described thoroughly in the appendices, papers I, II and IV, but a summary of these is given below.

The main source of the deficiencies in existing methods as well as methods during development has been the results from other Human Factors Engineering methods – such as Heuristic Evaluation, Usability Tests, and interviews with and observations of users, but also from expert evaluations. Together with methods used during development these methods have been employed in different development project and by comparing the usability problems and use errors detected by other methods. Comparisons between what the methods detected and what they ought to detect, i.e. weaknesses in the methods during development, alternative methods could be identified.

On the basis of the detected weaknesses and deficiencies, the methods were supplemented so they can cover more usability problems, use errors and ergonomic errors and also make it possible to investigate the problems and errors more thoroughly. The starting point for the changes in the methods was the existing methods and theory in the area of Human Factors Engineering.

The working approach for creating proposals of improvement has been a "trial-and-error" process where various suggestions have been proposed, tested and then accepted or rejected. In this way the methods have been modified to better meet the requirements and resolve the deficiencies.

5.2.2 Development of Enhanced Cognitive Walkthrough

The foundation of the methodology is set by the Enhanced Cognitive Walkthrough method, whose development took place during the author's master thesis (Bligård and Wass, 2002). In that project there was a need for an analytical usability-evaluation method that delivers qualitative data. The choice was the existing Cognitive Walkthrough (CW) (Wharton et al., 1994, Lewis and Wharton, 1997) method. It derives from evaluations of simple technical systems, called walk-up-and-use systems. Moreover, the method is question-based. CW has also been used by many researchers in evaluation of medical equipment (Kushniruk et al., 1996, Kaufman et al., 2003, Horsky et al., 2003). Within the research group to which the author belongs, there is experience of its use both in research projects (Liljegren and Osvalder, 2004, Liu et al., 2005) and in degree work (Axelsson, 2002).

During the author's use of CW, a number of weaknesses in the method were identified. The main source of the deficiencies in CW was the results from experience of using the method in different development projects and by comparing the usability problems detected by other methods with those detected by CW. These are fully reported in Paper I. Here follows a brief compilation of the most important shortcomings.

- CW has a deficient high-level perspective in the evaluation of user interfaces, which was manifested thus
    - CW does not answer whether the user knew that the function concerned was available
    - CW does not answer whether the interface provided hints that enabled the user to discover more easily that the function was available
- The explanations for success or failure yield insufficient information about the difference in problem severity between distinct operations
- It is difficult to obtain an overview of the results, both within an interface and between different interfaces

This is also shown by the author's own experience from practical application of CW, because CW does not discover any use errors during usage, whereas outside the method usage the author discovered the possibility of the user making errors. Hence, in order to answer these questions, some other method than CW is needed on which to base the method development. The modification of Cognitive Walkthrough that has been performed according to the spiral process, Figure 3.3, and the focus of the improvements related to the question process and semi-quantitative data.

An example of the development process is how the matrices develop. In methods under development (ECW) indexing had been introduced to make them more semi-quantitative. Different indexes were needed to enable a better overview of the results. For ECW it was especially essential to be able to compare the severity of a detected usability problem with the importance of the function affected by the problem. It is obviously more urgent to solve a usability problem if it exists in a vital function of medical equipment than if it occurs in a less important function.

The first proposal was to multiply the severity of the problem (graded 1-5) by the importance of the function (also graded 1-5) and thus obtain a numerical measure of how critical it was to deal with each individual usability problem. This proposal was rejected because it meant that a less severe problem in an important function would appear as critical as a severe problem in a less important function, and no such symmetry was felt to exist. Instead, it was decided to use a matrix with problem severity on one axis and function importance on the other. Each combination could then be judged separately as to whether it needed rectification or not. The introduction of matrixes also enabled the results of the analysis to be easily overviewed.

The refinement resulted in three supplements to Cognitive Walkthrough:

1. Division into two question levels, allowing investigation not only of operations but also of tasks/functions.
2. Introduction of indexes: a grading of tasks and of failure/success stories, and categorisation into problem types
3. Presentation of results in the form of matrices for easier overview of the results

The further development of CW was called Enhanced Cognitive Walkthrough (ECW).

Although the main part of the development of ECW occurred in 2002, minor adjustments have been made over the years. The last change took place in autumn 2011, when a question was added to the query process in spring 2012 when analysis questions were updated to work better. The question which was added related to whether the user was able to perform the action (question 5 operation). There may be situations in the interaction when the user knows what to do but cannot perform the action, which came to light when ECW was used to evaluate kitesurfing (Appendix E). The questions showed all Yes (5), but both the evaluators felt that beginners would not be able to perform the task, which made it necessary to modify the method. A question related to affordance was therefore needed.

The last change to the method was made during the writing of this thesis as the questions on both levels went through a review process to make them clearer and easier to use. ECW was originally developed to evaluate user interfaces for medical devices but has also been successfully used to evaluate artefact s without a traditional user interface like stable equipment and kite-surfing gear. The questions therefore needed to be more general in order to simplify for the person performing the evaluation (while at the same time retaining the meaning of each question).

5.2.3 Development of Predictive Use Error Analysis

The next method development was performed when the author worked at Breas Medical AB; part of the work involved carrying out a risk assessment of use. It concerned examining the operational errors that may occur with the machine that was under development, for the purpose of taking subsequent action and increasing safety.

The first idea was to use ECW, but that did not work as one limitation of CW is that the method focuses on the correct sequence of actions in the interaction, i.e. on those actions which are the correct ones for reaching the goal. CW does not take account of when and how in the interaction a user commits errors, and how these errors can be counteracted. The originators of CW (Lewis and Wharton, 1997) also write that *"the CW makes no attempt to say what the users will do if and when they depart from the correct sequence."*

Furthermore, CW does not address how a user is to recover from a committed error, i.e. whether it is easy or difficult for a user to find a way back to the correct sequence of action (Blandford et al., 2004). Many researchers have pointed out that CW is poor at analysing errors in the interaction (Green et al., 2000, Kaufman et al., 2003, Blandford et al., 2004).

A decision was instead taken to use the Predictive Human Error Analysis (PHEA) (Embrey, 2004) method. This method belongs to a group of three rather similar methods: Action Error Analysis (AEA) (Taylor, 1979, Harms-Ringdahl, 2001), Systematic Human Error Reduction and Prediction Approach (SHERPA) (Embrey, 1986), and Predictive Human Error Analysis (PHEA). These methods originate from evaluations of complex systems, primarily control rooms in nuclear energy plants and the process industry. However, AEA, SHERPA and PHEA are so similar that they will be regarded hereafter in the thesis as a general method. AEA, SHERPA and PHEA investigate what errors a user can commit in the interaction, and what consequences these errors may have. AEA is an often-employed method for risk analysis in the process industry (Suokas and Pyy, 1988). SHERPA has also been applied in medical care to investigate medication errors (Lane et al., 2006).

Parallel with this was a need in the course of the author's academic work. During a project concerning comparison of different types of pulsoximetry sensors (Moric et al., 2004), as well as in master thesis work regarding an operation robot (Nilsson and Åhman, 2005) the aims and requirements for a method to predict usability problems and use errors began to emerge. In work on the robot, in particular, there was a need to investigate both what made the user act correctly and what made the user act incorrectly. The Action Error Analysis (AEA) method

During the author's use of AEA and PHEA, several inherent weaknesses were revealed. These are fully reported in Paper II. A brief compilation of the most important weaknesses is given below.

- Deficient connection to cognitive theory in order to explain the occurrence of errors
- Deficient high-level perspective in the analysis, as the methods focus chiefly on individual operations
- Difficulty in obtaining an overview of the results

The modifications were also made by the cyclic process and influenced by the earlier developed ECW. The refinement led to AEA, SHERPA and PHEA being combined as a joint method and to three supplements. In addition, the various investigation points that exist for detecting use errors were modified. The supplements are:

1. Application of the Generic Error Modelling System (Reason, 1990) to describe detected errors
2. Division into two question levels, allowing investigation not only of operations but also of tasks/functions. This division is the same as in ECW
3. Presentation of results in the form of matrices for easier overview of the results

The further development of PHEA was called Predictive Use Error Analysis (PUEA). As ECW was already developed with a structured surrounding framework, PUEA was developed with a similar framework as it was the same type of information that was needed for the method to be able to be conducted. PUEA has not undergone any major changes since its first version.

5.2.4 Combining ECW and PUEA

Already during the development of PUEA the idea of being able to combine ECW and PUEA methods was born (hence their shared structure). The reason for this was to obtain a more comprehensive analysis of the interaction. PHEA had also already been used with ECW in a number of studies (Pettersson and Osvalder, 2005, Henriksson and Strängberg, 2007). By combining ECW and PUEA four interesting questions could be considered:

1. Will the user act correctly?
2. Why does the user act correctly?
3. Which errors can the user commit?
4. Why does the user act incorrectly?

When combining ECW and PUEA a method emerges to cover all the four requirement questions. Since CW has the potential to answer questions 1 and 2, while AEA, SHERPA and PHEA have the potential to answer questions 3 and 4, a combination of the two will cover all the questions. The methods also fulfil the requirements of being formative, analytical, proactive and question-based. Their combination was thus seen as sufficiently satisfying the above-described requirements to be selected as a basis for the method development. Table

5.16 shows the joint fulfilment of requirements for the combination of CW and AEA, SHERPA and PHEA. However, the combination does not completely cover the requirements to do so; method development is needed regarding question-based analysis and semi-quantitative data.

Table 5.16 Joint fulfilment of main requirements

Requirement questions	**Joint**	**ECW**	**PUEA**	**CW**	**AEA**	**SHERPA**	**PHEA**
1. Will the user act correctly?	Yes	Yes	No	Yes	No	No	No
2. Why does the user act correctly?	Yes	Yes	No	Yes	No	No	No
3. Which errors can the user commit?	Yes	No	Yes	No	Yes	Yes	Yes
4. Why does the user act incorrectly?	Partly	No	Yes	No	Partly	Partly	Partly
Properties							
Question-based	Partly	Yes	Partly	Yes	Yes	No	No
Qualitative data out	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Semi-quantitative data out	Partly	Partly	Partly	No	No	Yes	Yes

It should be noted that the choice of CW and AEA, SHERPA and PHEA does not imply any claim to be an optimal match. The project in the course of which method development took place aimed to improve the user interfaces of different devices, not to improve the methods. Instead, the choice of methods has been based on established requirements and well-known methods that were regarded as suitable for usability analysis in the projects, i.e. the methods are sufficient for the projects' requirements. Among the methods AEA, SHERPA and PHEA, the last method became the most frequently used of these three.

As ECW and PUEA were developed along the same structural framework, there was no problem to merge them into a common methodology (Paper IV). The big benefit is being able to examine whether the user is performing correctly and incorrectly at the same time. Common templates were developed for ECW and PUEA (Appendix F) and this idea has also received feedback connected to the original methods in the other direction. Analysts are now taught to use CW and PHEA at the same time to get a more comprehensive analysis.

5.2.5 Development of Alarm-ECW

A spinoff from ECW was the variant with alarm-ECW. It was developed when an analysis tool for alarm messages was needed. The need existed in a project where alarm signals from a dialysis machine were evaluated. The regular ECW worked well enough but there was scope for improvement. The difference compared with before was that during the alarm and information signals it was the machine that initiated the interaction, and not the human, so a different set of questions was required.

The idea was that, instead of building up questions on operating level along explorative learning, the theory was based on alarm design and guidelines from alarm signals were used as a template. The guidelines for alarm design were NUREG-0700 Chapter 4 (NRC, 2004), IEC-62241 (IEC, 2004b), Bransby and Jenkinson (1998) and EEMUA 191 (1999) where the emphasis was on EEMUA which has presented important characteristics for an individual alarm as:

- Timely – in the right time
- Relevant – to the operators
- Unique – not be a duplicate of another alarm

- Prioritised – help the operators focus their attention
- Understandable – speak the operator's language
- Diagnostic and advisory – indicate what has happened and what action the operator has to take
- Manageable – not too many alarms

Using Seven Stages of Action (Norman, 2002) a description was developed regarding how an operator handles an alarm and for every step a question was formulated similar to Enhanced Cognitive Walkthrough, table 5.17. The result became a method for alarm evaluation, Alarm Enhanced Cognitive Walkthrough.

Table 5.17 Operator step and questions Alarm-ECW

	Operator step	**Question Alarm-ECW**
1	Detect the alarm	Will the operator be able to detect that there is an alarm condition?
2	Determine the seriousness of the alarm	Will the operator understand the seriousness of the alarm?
3	Identify the alarm	Will the operator be able to identify the alarm?
4	Interpret the origin of the alarm	Will the operator be able to interpret the alarm?
5	Attend to the alarm	Will the operator associate the correct measure/action with the alarm?
6	Evaluate the measure	Will the operator get sufficient feedback to understand that the alarm has been attended to correctly?

5.2.6 Development of Predictive Ergonomic Error Analysis

Previous method development has occurred within the area of cognitive ergonomics, but in a project where the physical ergonomics of a dialysis machine was evaluated the focus was instead moved to the physical aspects. The evaluation was done by classical methods such as RULA and REBA. One aspect that was interesting to explore, but that is not covered by these methods, was how the user will perform an action, not only can perform it. Although it is possible to perform the action in a good ergonomic way, the user does not have to do it that specific way. Thus, a user can perform an action in the right way in relation to the machine, but the action can be performed in a manner that is harmful to the user physically or mentally.

One idea was therefore to create a method that examines whether the user will perform the action in an ergonomically correct way and what ergonomic errors the user might make. The meaning of ergonomic errors initially only concerned the physical part, but that came to be expanded to also encompass the mental part. This was done so as to make the CCPE methodology more consistent and to adapt PEEA to the agreed definition of ergonomic error.

The method emerged from the joint ECW/PUEA evaluation and the original CW and PHEA methods, modified to instead analyse how users use their body to interact with a machine. The method examines the following questions:

- Will the task be performed in an ergonomic way?
  Why? / Why not?
- Can the task be performed in a non-ergonomic way?
  What will the consequences be?

The method was named PEEA - Predictive Errors Ergonomic Analysis. The analysis is done with the assumption that the user performs the action correctly on the machine. PEEA and ECW / PUEA thus cover different aspects of the interaction.

5.2.7 Development of Generic Task Specification

ECW and PUEA investigate what goes right and wrong in the interaction. When these methods were used to evaluate the dialysis machine under development a need arose to better map the workload on the user, to use it both as a basis for ECW/PUEA and as an independent analysis to evaluate that specific aspect of the interaction. The first drafts of GTS were already in existence during the process of developing PUEA, but it was not then seen as a separate method, more as a step towards preparing to describe the task.

There are already many methods used for assessing workloads, from dedicated expert evaluations to structured methods like SWAT (Wilson and Corlett, 1995), NASA-TLX (Hart and Staveland, 1988) for mental workload and RULA (McAtamney and Corlett, 1993), REBA (Hignett and McAtamney, 2000) for physical workload. What was needed in the dialysis machine project was a generic way to report results. As a result of this approach physical and mental workload could be accounted for and it would be possible to make comparisons between analysed machines.

The Generic Task Specification method, GTS, was designed by categories taken from SWAT and NASA-TLX, supplemented with more physical aspects and breakdown of the body using a body map. For each category there is a grading scale. GTS was further adapted to easily work with ECW, PUEA and PEEA.

5.2.8 Further development of the methodology

In connection with GTS it emerged that a number of methods could be used together. These methods relied in turn on other methods such as User Profile and HTA. Furthermore, the methods used could vary depending on the purpose of the evaluation and likewise the contents of the method could be adapted in the same way. A system of methods had thus emerged, a methodology. The methodology was named Combined Cognitive and Physical Evaluation - CCPE. The development of CCPE was hence not the result of a conscious effort, more a result of how the different methods had been developed in proximity to each other and the way they worked together. The individual methods grew to become a more systematic and structured way to study the interactions between humans and machines with the aim of detecting possible mismatches between them in the interaction.

<table>
<tr><td colspan="4" align="center">1. Definition of evaluation</td></tr>
<tr><td colspan="4" align="center">2.Human-machine system description</td></tr>
<tr><td>a. User profiling</td><td>b. Task analysis</td><td>c. Context description</td><td>d. Interaction description</td></tr>
<tr><td colspan="4" align="center">3. Work load analysis (by GTS)</td></tr>
<tr><td>a. Task demands</td><td>b. Automation levels</td><td>c. Mental workload</td><td>d. Physical workload</td></tr>
<tr><td colspan="4" align="center">4. Interaction analysis</td></tr>
<tr><td>a. Usability problem analysis
(by ECW and alarm-ECW)</td><td>b. Use error analysis
(by PUEA)</td><td>c. Ergonomics error analysis
(By PEEA)</td><td>d. Physical ergonomics analysis
(By REBA or RULA for instance)</td></tr>
</table>

Figure 5.7 Procedure for CCPE methodology, which added in physical ergonomic evaluation

One part of the interplay between human and machine that is not accounted for is an analysis of whether, from the machine viewpoint, the correct way to work is harmful to humans in the short or long term. There are three reasons for this not being done: (1) There are already many good methods in the field, (2) it is a complex area, and (3) none of the projects studied needed new methods. However, it is no problem to combine CCPE methodology with methods that examine parameters such as the physical ergonomics of a task. They can be added as part of the interaction analysis that specifically deals with the physical ergonomics in the same way as the other parts, see Part 4.d in Figure 5.7.

5.3 Use

CCPE methodology and its methods have been used in numerous projects. What is presented here is details of its use by the author. The presentation spotlights methods as well as combinations of methods, first with ECW and then with CCPE.

5.3.1 Use of ECW

Enhanced Cognitive Walkthrough has been used as a single method in a number of projects.

Master thesis: Home care ventilator

(Author analyst)

The ECW method were first used in a master thesis work in the field of human factors engineering (Bligård and Wass, 2002, Paper IX) in which the first version was developed. The goal of the work was to evaluate and redesign a family of user interfaces for home-care ventilators. In the work there was a need for an analytical method to evaluate existing interfaces; thus it was not possible to perform an exhaustive usability test of all functionalities. Cognitive Walkthrough (CW) was selected as the method for grouping the usability problems. After the redesign, the CW method was employed to evaluate the new user interfaces to make it possible to compare usability problems. In the work, heuristic evaluation and usability testing were also employed. However, problems were found with the CW method. Therefore a method development was undertaken which resulted in the Enhanced Cognitive Walkthrough (ECW) method. During the process of developing ECW, thoughts about and increased interest in the requirements of methods in real product development emerged.

Research project: Insulin pumps

(Author analyst)

The first use of ECW in a research project was a project on insulin pumps (Paper 1 and Paper XI). The project evaluated the user interface for two existing insulin pumps with ECW to list problems connected to user friendliness. ECW was described thus: *"This method is more resource demanding than the heuristic evaluation, but it does, however, more specifically show the problems that can arise during use"*(Bligård et al., 2003a).

Master thesis: Infusion pumps

(Author supervisor)

The first time ECW was used by other persons than the author was in a thesis project about infusion pumps (Gross and Foufas, 2003, Paper X). Redesign of the user interface for the volumetric infusion pump IVAC 591. Enhanced Cognitive Walkthrough was used and revealed many of the usability problems with the user interface. They wrote (Gross and Foufas, 2003, p 63): "*In this thesis the ECW was performed on all HTA data for six different devices. If another attempt to apply the HFEP*[2] *were to happen again, the input to the ECW would be narrowed down, focusing on just a few of the devices and the important functions. Before performing the ECW it is recommended that the four questions are carefully interpreted. Interpretation of the questions is of crucial importance for problem detection and since the process is very demanding it is important to think over the purpose of the ECW and adjust the interpretation to fit the device in question. The ECW was the single most important source for the design criteria and was considered to be one of the best tools in the HFEP for*

[2] Author note: Human Factors Engineering Process.

detecting problems in an existing UI. ECW improved handling of the problems detected by the classical CW. The grading of problem seriousness not only highlights the most serious problems, it also encourages the inclusion of less serious problems in the list. According to the author's own experience, the benefit of considering less serious problems is that users tend to be more capable of making mistakes and mixing up functions than expected, so the less serious problems need to be considered too."

Research project: Children's car safety seats

(Author not analyst)

The first use of ECW in a research project without the author's involvement dealt with car safety seats for children. The purpose of this study was to examine ergonomic aspects of forward-facing child car seats, considering the child's seating comfort as well as the usability of the seat for different user groups (Pettersson and Osvalder, 2005). ECW was used to investigate user problems in the interaction between parent and car safety seat and it resulted in a number of design aspects being recommended for improvement.

Master thesis: Nuclear power plant control room

(Author not supervisor)

The next thesis project to use ECW dealt with control rooms. ECW were used for analysis in one of the control rooms of a nuclear power plant (Oxstrand, 2006), more precisely to investigate what factors in the user interface needed to be examined. The following was written about ECW (Oxstrand, 2006, p 50): *"ECW was conducted after an HTA was carried out, in other words early on in the study. After this was done, it was uncertain whether the result of ECW would help carry the study forward as it focuses considerably on the user interface graphics. This is usually not a bad thing but this study focused on the entire control room as an interface and not only the physical control units. In other words, it was uncertain whether ECW was really suited for the study purpose as it only analysed a small part of what was of interest to study. The reason that the method was implemented in spite of this was to analyse the interface to see what it contributed in terms of risk of human error. When the results of all the methods were compared and analysed, it was immediately obvious that ECW had a given place in the study. Results from ECW confirmed and strengthened the results from all the other analysis methods."*

Master thesis: Carton bale machine

(Author supervisor)

ECW was also used as a method in a thesis project regarding a carton bale machine. The purpose of this thesis was to conduct an examination of the user friendliness of the machines in the Orwak 3000 series with the aim of generating improvement proposals for the machines' user friendliness (Henriksson and Strängberg, 2007). ECW was used to evaluate the existing user interface. The following was written about ECW (Henriksson and Strängberg, 2007, p 104): *"The interaction analysis added understanding of the usability problems that the machine interface had and the errors that could arise from these areas. ECW was a time-consuming method, but it was helpful since it made the authors carefully analyse each step of the tasks identified in the HTA. The results from ECWs clearly showed the steps in which the users encountered difficulties."*

Master thesis: Combat vehicle 90

(Author not supervisor)

ECW has also been used singly to evaluate vehicles. This thesis work aimed to analyse the user interface in an existing infantry fighting vehicle and then propose possible improvements, and this was done using ECW (Ellingsen and Lundmark, 2009). The following was written about the method (Ellingsen and Lundmark, 2009, p 83): *The ECW was done with haptics in mind, since the commanders rarely look at the panels they are using. "This was a new way of using this method for the authors, but the method worked very well for this kind of evaluation. The source of error in this test was the fact that it was carried out by the authors alone, with the HTA as a guide for the procedures. It would probably produce more correct results if the walkthrough was done in cooperation with a commander. The ECW could also have been performed on several commanders to get an average result."*

Master thesis: Movable incontinence inserts

(Author not supervisor)
In this master thesis the ECW was used to identify and evaluate flaws in prototypes of movable incontinence inserts. The aim was to define needs and indications as to which concepts should be developed further and where efforts should be focused in order to improve the concepts (Nilsson and Oredsson, 2010). The ECW analysis was also performed to find out if there was a need for more guidance for use. The result of the ECW showed a difference between the concepts that was useable in the development process and also underscored a greater need for guiding instructions.

5.3.2 Use of PUEA

Predictive User Error Analysis has been used as a stand-alone method in a number of projects.

Industry project: Home care ventilator

(Author analyst)
The first use of PUEA was in a project where the method was developed, when there was a need to perform a risk analysis of use of home-care ventilators at Breas Medical AB (Bligård, 2003, Bligård, 2004). The risk analysis was performed by a team. The team mainly consisted of a usability engineer (the author), a product manager, a clinical expert, and a person from the regulatory/quality department. The work started by applying AEA, but the result was not satisfactory. After that, PHEA was tried with the same result. The problems that emerged were the same as the problems found earlier when using the CW method. This resulted in a new method being developed with PHEA as the base. The new method was later named Predictive Use Error Analysis (PUEA). PUEA was subsequently used to evaluate the new home-care ventilators under development. The method was modified several times during this phase and the modifications mostly centred on which categories would be involved and in what order they would be. The method provided a systematic review of all possible human errors with the machine and showed the user-related risks that existed. These results were then used in the overall risk analysis to assess whether the machines were safe enough, or if design changes were needed.

Master thesis: Insulin pumps

(Author supervisor)
The first time PUEA was used by someone other than the author was in a thesis regarding insulin pumps. The goal of the project was to develop two models for the future of an insulin pump with integrated continuous glucose monitoring that was user friendly, has an attractive design and helps users feel healthy (Dahlén and Ullström, 2006). PUEA, then called EPHEA (Enhanced Predictive Human Error Analysis) evaluated the simple task of changing the insulin ampoule in an insulin pump. They wrote about the method (Dahlén and Ullström, 2006, p 34): *"The EPHEA method was used to control the errors that can be made. The method turned out to be a good method for detecting errors and as a result the project gained an overview of whether the disposition and design worked together with the interface. EPHEA had never previously been used by anyone in the group but it was a very effective method to*

use to detect use errors early in the product development. Unfortunately, this method is misleading to the extent that the project group itself tried to figure out which situations could cause problems with the insulin pump and remote control. For the method to cover all the problems it should have involved an insulin pump user since the project group did not have a good understanding of the types of errors that can occur."

5.3.3 Use of ECW and PUEA

Enhanced Cognitive Walkthrough and Predictive User Error Analysis have been used together in several projects.

Research project: Anaesthesia machine

(Author partially analyst)

This project evaluates an anaesthesia machine for Maquet that was in the final stages of development. ECW and PUEA were used as risk analysis for use in order to investigate whether there were hazards during use and how the design of the machine affected this. The author participated initially during the analysis phase in order to teach the company, after which the company did most of the analysis itself. The company's thoughts about the methods can be found in Paper V.

Research project: Kitesurfing

(Author analyst)

The aim of this study was to investigate how the design of kitesurfing equipment can affect safety issues from a usability perspective (Paper XVI). A focus group of six subjects analysed the task of preparing kitesurfing equipment for use, employing the Enhanced Cognitive Walkthrough (ECW) and Predictive Use Error Analysis (PUEA) evaluation methods. From the evaluation, a list of plausible usability problems was identified together with proposed design guidelines. The following was written about the methods (Lundgren et al., 2011, p 5): *"The main procedure with a focus group performing ECW and PUEA was appropriate for this study and resulted in useful guidelines to improve usability aspects in the design."*

Research project: Emergency patient stretcher

(Author analyst)

This project evaluates an emergency patient stretcher using ECW and PUEA (Lundgren, 2010). The purpose of the analysis is to identify potential improvements to make the product easier and safer to use. The analysis was performed by a group of four people where two were users of the stretcher and two were experienced users of the method. The two users were initially sceptical about the ECW / PUEA thought process, i.e. to actively look for problems and errors, but they quickly changed their minds when they realised that the methods highlighted the problems and errors and thus showed what could be improved.

The analysis resulted in seven points for the improvement of 66 identified more or less serious user problems and risks for handling errors. These were then presented to the manufacturer of the stretcher. The manufacturer's reaction to the material presented was that this type of data was very valuable for further product development and the results confirmed some of the development possibilities that the manufacturer itself had identified.

This evaluation was ideal for ECW/PUEA because it was conducted on a medical product, users of patient stretchers could attend the performance, and the evaluation was led by experienced method users. The two methods users were surprised at how many unfamiliar problems and errors could be found with the methods by critically reviewing the product. In this case, it was also a product that has many different users, often in critical and time-pressured situations and therefore places high demands on understanding and ease of use.

Research project: Public transport website

(Author not analyst)
This study evaluated a public transport website to investigate usability problems and use errors with the help of ECW and PUEA (Moradi and Pour, 2011). The study shows of numerous problems and errors that may set the foundation for a redesign of the webpage.

The following was written about the methods by Moradi and Pour: *"The outcomes of the methods offer extensive details about usability problems, such as problem seriousness ... and use error, such as error type ..." and "One of the main advantages of ECW and PUEA over their predecessors relates to result presentation facilities, whereby the results of the analysis phase are presented in several matrices in ECW and PUEA. The matrix form of results makes it possible to compare different tasks or interfaces in the same field, overview the general situation of the interface and review problematic tasks." And "The outcomes of matrices can form suggestions for website design improvements."*

The authors acknowledge that the two level analyses increase the generality of the evaluation. About the drawbacks of ECW and PUEA the authors write: *"Application of evaluation methods in this study confirmed that limitations such as tediousness, complexity, being subjective and time-demanding still exist." "Moreover, consequence severity grading for the PUEA method must be attuned for better conformation with a human–computer interaction system."*

The authors also miss a guideline for how to report the qualitative aspect of the errors and problems and they feel that ECW and PUEA need further development to work better with web pages. They conclude: *"All in all, the utilisation of ECW and PUEA methods in a Human Computer Interaction (HCI) system provided comprehensive detection of probable usability problems and use errors that the daily user of the system may encounter."*

Research project: Electronic medical records

(Author not analyst)
This project was undertaken with the aim of assessing technical and social aspects of an electronic medical records system (Tancredi et al., 2012). One part was a usability evaluation where ECW and PUEA were the two methods used. The methods functioned very well for the evaluation and the authors write: *"ECW unfolded issues with planning, learning and exploration of the system; it identified serious problems with texts and icons, hidden controls or information, unnatural sequences and insufficient or ambiguous feedback*" and "*PUEA identified potential serious errors that the interface can lead to.*" They also stated that *"Only ECW and PUEA helped to reveal issues related to physical action."* In this project, ECW and PUEA have thus worked as intended. More can be found about this study in chapter 5.4.

Research project: Car seat belts

(Author analyst)
This is an ongoing project during the writing of this thesis. The overall aim of the project is to evaluate two proposals on improved seat belts. ECW and PUEA were used to predict usage errors for the proposals until more empirical experiments were conducted. This was done so as to get an idea of what could possibly happen with the users involved. ECW and PUEA showed many possible operational errors with the belts and many of these occurred when tests

with users were conducted. As a result, the proposals for the belts underwent further development designed to make the belts easier to manage and also to reduce the probability of the user making mistakes.

Master thesis: Boat navigation

(Author not supervisor)

This thesis project aims to develop an innovative device that delivers an extended set of data about the status of the boat and engine as well as about its navigation (Strömberg and Freyhall, 2009). HTA, ECW and PUEA were used to evaluate developed design proposals. The ECW was first performed to test if the system's intended use procedure was obvious to the user. PUEA aimed to find what type of error the user would make and how severe the consequences would be.

Master thesis: User interface cardiac output

(Author supervisor)

This thesis project related to medical technology and the goal for the usability study of an USCOM Cardiac Output Monitor is to present a set of usability guidelines (with the focus on software) to increase the quality of the human-machine interaction: target achievement, efficiency, safety and user satisfaction (Mårtensson, 2010). The implementation of the guidelines is visualised by fictional screenshots of the interface. ECW and PUEA were used to find usability problems and use errors in the interaction with the user interface. Mårtensson (2010, p 137) wrote: *"The use of Enhanced Cognitive Walkthrough (ECW) and Predictive Use Error Analysis (PUEA) that were based on the Hierarchical Task Analysis, (HTA) led to the discovery of many of the usability issues. They are suitable methods for analysing the human-machine interaction of a device such as the USCOM and are certainly effective. It was interesting and rewarding to be able to discover potential causes of errors, and then watch users perform errors related to these issues in actual usage. The methods are quite time-consuming however, and might not always be the first choice for improving the human-machine interaction for a product at a small company such as Uscom Ltd."*

Master thesis: Flat saws

(Author supervisor)

This thesis project worked ergonomics in flat saws (Widing and Bui, 2012) ECW and PUEA was used to ECW/PUEA were used to evaluate and describe the existing situation, by identify possible use errors and usability. This was then used to identify user needs for the flat saws. About the methods Widing and Bui wrote (2012, p 107): *"The system was also analysed using ECW and PUEA for identifying usability problems and use errors. These methods were found to be a bit difficult to apply to this system, where they are designed for more complex systems. They were somewhat adapted to this project however it needed much analysis to be able to determine what is relevant for the actual system. The method is very good for identifying potential problems in the system however it would have been beneficial with a second iteration of contact with users to be able to verify this type of evaluation. The lack of verification had to be taken into consideration in the analysis both when including and excluding some aspects."*

Bachelor thesis: Radar surveillance software

(Author supervisor)

In this thesis ECW/PUEA is compared with another method called Rapid Usability Evaluation. The purpose of the thesis is to try to find strengths and weaknesses in Rapid Usability Evaluation and suggest interventions for the deficiencies that may exist (Gyllensvaan and Olsson, 2009). In addition, suggestions are given for how Rapid Usability Evaluation methodology can and should be used in the future. It will primarily be compared with Enhanced Cognitive Walkthrough (ECW) and Predictive Use Error Analysis, as these methods are potential replacements for RUE should this prove to be insufficient. The following was written about ECA and PUEA (Gyllensvaan and Olsson, 2009, p 47): *"When the complete methods are examined it quickly appears that ECW and PUEA take far too long to fit in the catchwords 'speed and flexibility', especially if a complete HTA needs to be set up just for them. An definite alternative is to carry out the complete analysis on selected elements that are critical to the system but this does not alter the fact that the methods are tedious and complicated. To speed up testing, especially in early stages when development can move quickly between different concepts, the authors propose that simplified methods are developed. These simplified methods must nonetheless be significantly faster than the original methods and less complex in their design, but still cover the parts that RUE currently misses."*

Course: Cognitive ergonomics

ECW and PEUA have been a part of the course since 2007 and have been used in student projects to investigate usability problems and use error. The methods there have been used in number of different areas and worked as they should. Students from 2009 were included in the study reported in Paper VIII.

5.3.4 Use of ECW, PUEA and GTS

Generic Task Specification, Enhanced Cognitive Walkthrough and Predictive User Error Analysis have been used together in two major projects.

Research project: Dialysis machine 1

(Author analyst)

This study was conducted to investigate the interaction with the bloodline system for three different dialysis machines (Bligård et al., 2006a). It was for this study that GTS was created. GTS was used to map out and compare workload between the machines, ECW was used to identify and compare usability problems while PUEA was used to map and compare use error. The usability evaluation was performed by a team consisting of a moderating usability engineer (the author), an assisting usability engineer and an experienced user, and was performed at the Gambro facility in Lund. All three methods worked well and were a good basis for finding the strengths and weaknesses of the design of the various machines.

Research project: Dialysis machine 2

(Author analyst)

This study examined a dialysis machine to highlight weaknesses in design and demonstrate possible design improvements (Bligård and Osvalder, 2006a). GTS was used to see where on the body the workload was placed, ECW to see problems with usability in the design while PUEA examined possible errors of the user. Alarm-ECW was also used to evaluate alarms from the machine. The methods worked well and gave clear indications of problems and possible improvements.

5.3.5 Use of GTS alone

Research project: Horse stable tools

(Author analyst)

GTS has also been used without ECW and PUEA in one larger study (Paper XVII). GTS was used to map the physical workload during work in stables – clearing the dung out of a horse stall. The method gave a good structure to the analytical task and set a basis for continuing with a computer manikin analysis.

5.3.6 Use of PEEA

Predictive Ergonomic Error Analysis has been used as a stand-alone method in a number of projects.

Research project: Dialysis machine 3

(Author analyst)

This study aims to evaluate the strain ergonomics of a prototype dialysis machine (Osvalder et al., 2005). It was for this study that PEEA was developed to examine how the user would actually perform actions ergonomically. PEEA was used here with HTA as a basis and together with RULA, REBA and OWAS. The method worked as expected by casting light on the ergonomic errors that could occur.

Research project: Dialysis machine 4

(Author analyst)

This study is a continuation of the study described above. The purpose is to evaluate strain ergonomics after a new prototype was developed based on the results of the previous study (Bligård et al., 2006b). The methods were the same and PEEA worked just as well here. It was now also possible to compare the results to clearly demonstrate improvements.

Research project: Computer mice

(Author analyst)

Many products are promoted as ergonomic, but for the user to make use of them it must be clear and understandable how the product should be handled. PEEA was used in this project to investigate the possible ergonomic errors in four ergonomic computer mice. This was then compared with results from usability tests. The study is described in Appendix A.

Research project: Horse stable tools

(Author analyst)

The ergonomic features of stable equipment were examined here to see if users used these for better handling of workload. First PEEA was performed and this was followed by usability tests. The study is described in Appendix B.

Course: Physical ergonomics

PEEA has been used as a method in the Physical Ergonomics master course since 2006 and has been used in student projects within a number of areas. The method has functioned well and given an insight into how users actually perform the actions. Just because it is possible to perform an action well from the ergonomic viewpoint, this does not mean that the user is using the product in the correct way, which PEEA clearly demonstrates to the students.

5.3.7 Use of whole CCPE methodology

To test the final CPPE methodology three studies were performed: with an office chair, with a vacuum cleaner and for kitesurfing.

Research project: Office chair

(Author analyst)

This study evaluated the ergonomics of a vacuum cleaner and an office chair, both of which were regarded as ergonomic. First the CCPE methodology was applied to develop potential usability problems and use errors. After this usability tests were conducted to compare the analytical analysis with the empirical. The study is presented in Appendix C.

Research project: Vacuum cleaner

(Author analyst)

This study evaluated the ergonomics of a vacuum cleaner and an office chair, both of which were regarded as ergonomic. First the CCPE methodology was applied to develop potential usability problems and use errors. After this usability tests were conducted to compare the analytical analysis with the empirical. The study is presented in Appendix D.

Research project: Kitesurfing

(Author analyst)

The full CCPE methodology has been developed in the context of medical technology. In order to challenge the method and test its limits, a study of kitesurfing was performed. The task analysed was an in-water start for beginners. The results from CCPE were then compared with the results of an observational study on real beginners who learned to kitesurf. The study is presented in Appendix E.

5.3.8 Summary

To summarise, the CCPE methodology and its detailed methods have been used in many projects since the method first arose ten years ago. Application may be greater than the projects listed above, as the author certainly is not familiar with all use of the methodology.

The methodology has grown considerably in medical technology, but it has also been applied to many other areas such as nuclear power, extreme sports, seat belts, car seats, stable tools and so on. Furthermore, the methodology has been used in industrial projects, research and student projects. The methodology has thus proved to be useful in many domains and in many situations.

6 Assessment

The chapter consists of two parts. In the first part the developed methodology is assessed by verification, validation and reflection. This is followed by a review of the methodology's relation to other methods and areas.

6.1 Verification

The overall purpose of the CCPE methodology development described in this thesis was to improve methods for handling and preventing presumptive mismatches between users and artefacts in the development process. To achieve this, a number of requirements were drawn up based on the theory concerning the methodology in order for it to work within the framework previously described (chapter 4). A first step in evaluating the developed methodology is to validate the methodology by outlining how the requirements are met and how they have been addressed in the methodology. This is done in table 6.1.

Table 6.1 Theoretic validation: Requirements for and fulfilment of the CCPE methodology

Requirement	**Fulfilment**
Start from a system perspective	In the CCPE the systems perspective is considered through the second stage of the methodology. The factors in the human-machine system that affect interaction are identified in the system description.
Study the details in the interaction, i.e. examine every individual step	Examination of each step of the interaction is carried out by basing the analysis on HTA. A number of questions are asked for each step of the interaction. Every action is thereby examined individually.
Consider both cognitive and physical aspects	Both the physical and the cognitive aspects are integrated in the methodology. During the task analysis both physical and cognitive actions are mapped. GTS shows both physical and cognitive load. ECW focuses on cognitive aspects, but also contains a question on physical aspects. PUEA concerns primarily cognitive aspects, although the effects may have physical consequences. PEEA shows clearly the relationship between the physical and the cognitive aspects.
Relate to risks during the interaction	Risk thoughts are clear in interaction analysis as PUEA and PEEA take into account the consequences of faulty actions, hence the connection between action and consequence.
Possible to use during product development	The methodology is designed to be applicable during product development in that it can be used to evaluate solutions that are not yet fully developed, and provide input to the proposed changes.
Function together with risk management	The methodology has embraced the basic ideas of risk management.
Based on knowledge about physical ergonomics	The parts concerning physical ergonomic problems and ergonomic errors are based on knowledge of physical ergonomics.
Based on knowledge about cognitive ergonomics	The parts concerning usability problems and use error are based on knowledge within the area of cognitive ergonomics.

Function together with human factors engineering	The methodology is developed from methods which derive from HFE and has been used in projects within HFE.
Function together with principles and processes of user-centred design	The methodology has been developed in a context where principles and processes of user-centred design have been applied.
To address users who are engineers with knowledge of Human Factors Engineering	The methodology is more effective to use for persons with knowledge of HFE.
Relate blunt end with sharp end	CCPE relates blunt end and sharp end concepts by including both user problems and use error. For use errors both cause and consequence are investigated.
Examine usability problems	ECW was especially developed to examine usability problems
Examine use error	PUEA was especially developed to examine use errors
Examine ergonomic error	PEEA was especially developed to examine ergonomic errors
Not examine physical ergonomic problems during correct use	CCPE does not concern MSDs since there is no analysis of whether an operation is harmful even if it is done in the best way. However, it is possible to add such an analysis if needed.
Take no account of abnormal use	CCPE does not take abnormal use into account by assuming that the user wants to perform correctly. PUEA seeks only use errors that can plausibly occur. The method makes no claim or attempt to be helpful in detecting all imaginable incorrect ways in which a product can be used. Instead, it focuses on the use errors that it is reasonable to think are possible during specified use in a specified environment. Whether or not the error could plausibly occur must therefore be evaluated in each individual case, in relation to the envisaged user and context. It is also important to point out that PUEA does not attach any value judgment to the term “error”, which means only that the user performs an action that the manufacturer has not intended. Thus, a use error means only a mismatch between the action expected by the designer and the way in which the user performs the action (compare the description by Norman (2002) of mental models in the designer versus the user)
Be question-based	ECW, PUEA and PEEA all use questions in their respective analysis.
Be formative and proactive	The methodology is designed to be used during the product development process and to generate information to improve development of the artefact
Be analytical, but include users and experts	The methodology does not require any testing with users, instead the user is simulated in the evaluation. However, it is beneficial if users are present during the execution of the methodology and the usefulness of the methodology is enhanced if HF experts are involved in the evaluation group
Yield qualitative and semi-quantitative data	The methodology generates results in the form of grading and categorisation but also in more detailed descriptions of usability problems and use errors

Derive from methods within the areas of Usability Evaluation and Human Reliability Assessment	CCPE is largely based on a further development of the Cognitive Walkthrough (CW) and Predictive Human Error Analysis (PHEA) methods

The specific questions that are described at the end of chapter 4.3 are fulfilled by the analysis questions in the interaction analysis in CCPE.

Usability problems are examined with ECW:

- Will the user act correctly with the machine?
- Why does the user act correctly?

Use errors are examined with PUEA:

- Which errors can the user commit with the machine?
- Why does the user act incorrectly?

Ergonomic errors are examined with PEEA:

- Will the user act correctly with the human?
- Which errors can the user commit with the human?

Table 6.1 showed that the requirements on the CCPE methodology are largely fulfilled and the methodology is adapted to the underlying theory. The next step is evaluation of how the methodology functions when applied in a thought-of context, in other words validation, which follows in the next chapter.

6.2 Validation

Besides the use of CCPE stated in the previous chapter, which in itself is an evaluation of the methodology, a structured evaluation was also performed. First a summary of use of CCPE is presented, then five empirical evaluations conducted by the author, followed by the results of an interview study with users. After this there is a summary of two evaluations conducted by other researchers. Finally, there are statements from two experienced users of CCPE regarding the pros and cons of the methodology and the author's own reflections.

6.2.1 Summary of use

CCPE methodology has been used in a number of evaluations, which were presented in chapter 5.4. The methodology was used to equally predict, identify and present mismatches in the interaction between user and artefact. Table 6.2 presents a summary of how the methodology was used.

Table 6.2 Summary of use of CCPE in different applications

	With the author as moderator				**Without the author as moderator**			
Application area / **Method**	GTS	ECW	PUEA	PEEA	GTS	ECW	PUEA	PEEA
Aesthesia machine		X	X			X	X	
Boat navigation						X	X	
Seat belts in cars		X	X					
Carton bale machine						X		
Child restraints in cars						X		
Combat vehicle 90						X		
Computer mice				X				
Dialysis machine	X	X	X	X				
Electronic medical records						X	X	
Emergency patient stretcher		X	X					
Flat saws						X	X	
Home care ventilator		X	X					
Horse stable tools	X			X				
Infusion pumps						X		
Insulin pumps		X					X	
Kitesurfing equipment	X	X	X	X				
Movable incontinence inserts						X		
Nuclear control room						X		
Office chair	X	X	X	X				
Public transport website						X	X	
Radar surveillance software						X	X	
User interface cardiac output						X	X	
Vacuum cleaner	X	X	X	X				

The summary in table 6.2 shows that the individual methods as well as the complete CCPE methodology were used in many application areas with the author as moderator as well as without the author participating in the analysis. The more common are evaluations within the field of medical technology and applications of the ECW method, but it is natural since the development of CCPE started with ECW to evaluate medical devices.

6.2.2 Empirical evaluation of CCPE

In order to empirically evaluate the CCPE, the author has undertaken five studies, one on computer mice (Appendix A), one on stable tools (Appendix B), one on vacuum cleaners (Appendix C), one on office chairs (Appendix D) and one for kitesurfing (appendix E). The evaluations were linked to the use described in the chapter 5. As CCPE had previously been used primarily in medical technology, other more everyday products were chosen to evaluate whether CCPE could be successful also in these cases. The choice of CCPE evaluation method in these five studies was based on how much the various methods had been used in earlier projects, which lead to more focus on PEEA, and less on ECW.

Computer mice

The study evaluated PEEA. Four computer mice promoted as featuring an ergonomic design were selected. PEEA was then used to identify potential ergonomic errors when using the computer mice. Furthermore usability tests were conducted in which 16 people got to try three of the mice (randomised choice). Their use was video-recorded and then analysed to determine the ergonomic errors that occurred in reality. These results were then compared with the results of the PEEA evaluation. The results can be found in Appendix A.

The comparison reveals that PEEA worked well for predicting the ergonomic errors that could occur, especially for the most frequent errors. PEEA predicted more errors than actually occurred but this is the nature of the method. Tables 6.1 and 6.2 describe the relationship between the results of the CCPE and the empirical study.

Stable tools

Also the second study evaluated PEEA. Six stable tools marketed as having an ergonomic design were chosen. There were two forks, a spade, a shovel and two brooms. PEEA was used to identify potential ergonomic errors when using the tools. Usability tests were then conducted where eight subjects tested all the equipment. The use was video-documented and analysed to determine the ergonomic errors that occurred in reality. These results were then compared with the results of the PEEA. The result can be found in Appendix B.

The comparison reveals that PEEA worked well for predicting the ergonomic errors that could occur, especially for the most frequent errors. PEEA predicted more errors than actually occurred but this is the nature of the method. Tables 6.3 and 6.4 present the relationship between the results of the CCPE and the empirical study.

Office chair

The study evaluated a modern office chair, which according to the manufacturer had very good ergonomic features. The interaction between the chair and the user during adjustment of the chair was analysed using the complete CCPE methodology. The evaluation resulted in a list of presumptive usability problems, use error and ergonomic error. Usability tests were also conducted. Eight people were asked to adjust the chair while being video-recorded. The usability tests were then analysed to see which adjustments caused problems for the subjects and what errors occurred. The results of the analytical evaluation with CCPE regarding usability problems, use errors and ergonomic errors were then compared with the results of the usability tests. The results are shown in Appendix C.

The comparison reveals that CCPE predicted more problems and errors than occurred during the usability tests but this is the nature of the method. However, it was not clear which of all predicted use errors would probably occur in real life. Tables 6.3 and 6.4 present the relationship between the results of the CCPE and the empirical study.

Vacuum cleaner

The study evaluated a modern cordless vacuum cleaner that could be used both for vacuuming the floor and as a table vacuum cleaner. The interaction between the vacuum cleaner and the user during normal use was analysed using the entire CCPE methodology, which resulted in a list of presumptive usability problems, use errors and ergonomic errors. Usability tests were conducted in which eight people used the vacuum cleaner. The tests were video-documented. The tests were then analysed to see which actions caused problems for the subjects and what errors occurred. The results of the CCPE in the form of usability problems, use errors and ergonomic errors were then compared with the results of usability tests. The results can be found in Appendix D.

The comparison reveals that the CCPE methodology predicted many presumptive usability problems, use errors and ergonomic with its three methods ECW, PUEA and PEEA. The results found by the ECW evaluation most closely resembled real world usability problems, and this method worked best in terms of prediction, followed by PUEA and then PEEA. In other words, predicting which of the various use errors and ergonomic errors that might occur in reality was more difficult than predicting which usability problems may occur. Tables 6.3 and 6.4 present the relationship between the results of the CCPE and the empirical study.

Kitesurfing

The study examined in-water starts during kitesurfing. First the CCPE methodology was used to perform an analytical evaluation; the result of the PUEA was a list of the possible use errors that a novice could conceivably make. After this four beginners were observed and video-recorded during their first-ever training for an in-water start. The films were analysed to determine the use errors that the novices actually made. The results from ECW and PEEA were not useful to compare with the results from the observations. The PUEA results in the form of use errors were then compared with the results from the observations. The result can of the PUEA evaluation be found in Appendix E.

The study shows that the PUEA methodology can be used also for a task requiring motor coordination and interaction, not only for a conventional user interface. Tables 6.3 and 6.4 describe the relationship between the results of the CCPE and the empirical study.

Discussion

Tables 6.3 and 6.4 show the relationship between what the CCPE methodology predicted and what occurred in the empirical studies. An initial reflection is that CCPE worked well in predicting what would happen. In all but one case, the majority of usability problems and use errors were predicted. However, the numbers also show that CCPE predicts many problems and errors that do not occur empirically. This may partly be because there have been too few test subjects in the usability tests as CCPE also takes into account errors and problems that are rare. It may also be due to the methodology prioritising identification of problems and errors that could occur instead of prioritising detection of errors and problems that are likely to occur. That is, the high percentage in Table 6.3 is due to the lower percentage in Table 6.4. CCPE thus identifies more problems than actually occur in order not to miss identifying some of the ones that do occur.

Table 6.3 Percentage predicted with CCPE that were identified during empirical tests.

	Computer mouse	Stable tools	Vacuum cleaner	Office chair	Kitesurfing
Percentage of usability problems predicted by ECW identified by empirical tests			45%	42%	
Percentage of use errors predicted by PUEA identified by empirical tests			63%	35%	42%
Percentage of ergonomic errors predicted by PEEA identified by empirical tests	56%	39%	26%	24%	

Table 6.4 Percentage identified by empirical test and predicted by CCPE

	Computer mouse	Stable tools	Vacuum cleaner	Office chair	Kitesurfing
Percentage of usability problems identified by empirical tests predicted by ECW			90%	90%	
Percentage of use errors identified by empirical tests predicted by PUEA			63%	53%	88%
Percentage of ergonomic errors identified by empirical tests predicted by PEEA	79%	78%	60%	44%	

ECW shows better results than PUEA and PEEA, which is probably because ECW investigates the only correct way to perform a task, while the other methods depend on both the user and the evaluator's creativity to make errors and predict errors. The PEEA also worked better on products with regards to physical ergonomics, i.e. where there is a large degree of physical interaction between the user and the artefact, which is reasonable since this is the focus of the method.

No other studies than the author’s comparing ECW, PUEA or PEEA with empirical tests have been found. There are however some studies done on the original methods, CW and SHERPA. In a study of a prototype military airspace scheduling system by Cuomo and Bowen (1994) CW was used. Of 24 usability problems predicted by CW, 14 (58%) were identified by empirical tests. In a study of a medication ordering system by Khajouei et al (2011) CW was also used. Of 38 usability problems predicted by CW, 22 (58%) were identified by empirical tests and of 41 usability problems identified by the empirical tests, CW predicted 22 (54%). Koutsabasis et al (2007) undertook a study of a website for a university department with two parallel evaluations using CW. In the first assessment, of 21 usability problems predicted by CW 18 (86%) were identified in empirical tests and of 70 usability problems identified by empirical tests 18 (26%) were predicted by CW. In the second evaluation, of 24 usability problems predicted by CW 17 (71%) were identified by empirical tests and of 70 usability problems identified by empirical tests 17 (24%) were predicted by CW.

Baber and Stanton (1996) performed a study on a vending machine using PHEA. Of 12 use errors predicted by PHEA 9 (75%) occurred during empirical observation and of 15 identified errors during empirical observation, PHEA predicted 9 (60%) of them. In another study of a vending machine by Stanton and Baber (2002) 36 students evaluated the machine individually with SHERPA. Of 21.6 (average) use errors predicted by SHERPA 6.2 (29%) were identified during an empirical test and of 9 usability problems identified by empirical tests 6.2 (69%) were predicted by SHERPA. Harris et al. (2005) conducted a study of a civilian aircraft flight deck using SHERPA. Of 56 use errors predicted by SHERPA 52 (93%) were identified by the empirical study while of 57 usability problems identified by the empirical tests 52 (91%) were predicted by SHERPA. Tables 6.5 and 6.6 show a comparison between these studies and the studies made with CCPE.

Table 6.5 Comparison between studies with ECW and CW

	ECW: Vacuum cleaner	ECW: Office chair	CW: Cuomo and Bowen (1994)	CW: Khajouei et al (2011)	CW: Koutsabasis et al 1 (2007)	CW: Koutsabasis et al 2 (2007)	
Percentage of usability problems predicted by the method and identified during empirical tests	45%	42%	58%	58%	86%	71%	
Percentage of usability problems identified during empirical tests and predicted by the method	90%	90%	-	54%	26%	24%	

Table 6.6 Comparison between studies with PUEA and SHERPA

	PUEA: Vacuum cleaner	PUEA: Office chair	PUEA: Kitesurfing	PHEA: Baber and Stanton (1996)	SHERPA: Stanton and Baber (2002)	SHERPA: Harris et al (2005)	
Percentage of use errors predicted by the method and identified during empirical tests	63%	35%	42%	75%	29%	93%	
Percentage of use errors identified during empirical tests and predicted by the method	63%	53%	88%	60%	69%	91%	

Table 6.5 shows that ECW is better at predicting all possible usability problems that could arise than CW, but it does not indicate whether the predicted problems are real or not. This is consistent with the goal of the development of ECW, which was to cover more usability problems. This has evidently been at the cost of ECW predicting more problems than actually occur, compared with CW. When it comes to PUEA, Table 6.6, the results of the comparison are not equally clear. PUEA is better when it comes to detecting several possible use errors, but not all of these errors actually exist in real handling situations. A reasonable conclusion would therefore be that ECW and PUEA have considerable potential for detecting most of the possible usability problems and use errors that can occur in a real interaction situation, thereby presenting more data and insights in these areas than their original methods CW and SHERPA. However, more studies are needed to determine this definitively. Nonetheless, the methods do not indicate which problems and errors are most likely to occur. As CW and SHERPA do not conduct such a thorough analysis as ECW and PUEA do, these methods seem to predict to a slightly higher degree the actual problems and errors that users might perform. However in the product development process, where the CCPE methodology is intended to be used, it is very important to get information on and understanding about all the possible problems users can face when the product is on the market. This knowledge might result in various ideas for possible redesigns to improve usability and mitigate use errors in the artefact when launched on the market.

CCPE has the aim to discover so many mismatches as possible and it is therefore interesting to reflect on why some mismatches in the studies described earlier have not been detected. Some missed mismatches, like the study of kitesurfing, where the result of that the task was not sufficiently well described, so all parts of the operation were not evaluated. Other possible causes are that the limit was too restrictive on what was a reasonable mismatch and that the users' variation of the action was greater than expected by the evaluator. Furthermore, the analysis is not better than the empathy and imagination that the evaluators had just at the time of the evaluation, which means that there is an element of randomness, since CCPE is based on a subjective assessment. Other aspects that may affect the analysis negatively are if evaluators become fatigue or bored, making it important to not do the analysis with the CCPE for longer sessions to keep the mind sharp.

CCPE was initially developed to analyse users’ interaction with the user interface in medical technology, i.e. pressing buttons in sequence. CCPE is not designed to analyse tasks such as kitesurfing in which the user interface is not as obvious and where the action steps are more complex. Nevertheless CCPE nonetheless worked well in this case. CCPE worked well also in other areas and a reasonable conclusion is that CCPE serves to analyse the mismatches between user and artefact within many different areas and is not limited to medical technology.

6.2.3 Interview study ECW and PUEA

As described previously, ECW and PUEA have been used by businesses and in education and they have also been evaluated in a study which is reported in full in Paper V.

Method

The study was conducted with individuals participating in a course at Chalmers (9 persons), and representatives of a medical technology company (5 persons). In both cases, the author first taught the method after which the participants in the study applied it to their course projects and, in the case of the company, a real development project respectively. After using the methods, the participants were interviewed and asked to fill in a questionnaire about the methods.

Results

The interviewees had in general a positive view of the method. The method was considered systematic and easy to overview, and offered clarity and awareness of the problems. The interviewees agreed regarding the strengths of the method. The main strength was that it is a structured method that allows for a run-through of the product's entire use, step by step, forcing you to consider many aspects in the interaction. The systematic approach made it easier to think critically of the product. However, this advantage is offset by some negative aspects already mentioned. The systematic approach makes it time-consuming, tedious, and unnecessarily complex for application with certain products, it is "technical" and "quantifies everything", causing it to be perceived as boring and lengthy.

Another advantage was that the method made it easier to communicate the concept of usability, which may appear fuzzy for some people not in the domain. The result of the method, in terms of a list of individual usability problems and use errors, made the product usability clear.

Other problems that emerged were difficulties when selecting tasks, prioritizing the tasks and then creating a good description of the user. There were also comments regarding the difficulty to know what to do with the result, including both the completed templates resulting from the analysis and the suggested result matrices. The participants also felt that the method would work better if the specified product user was a novice rather than an expert, since they thought it more difficult to imagine what an expert might do.

A few suggestions for improvements to the method emerged from the interviews. Some of the interviewees desired for instance more instructions on how to select and prioritize the tasks and also a way to adapt the method to their specific product.

Discussion

During the development of ECW and PUEA a number of strengths and weaknesses of the methods were identified, see Papers 1-4 and (Bligård, 2007) and the first part of the discussion will compare these with the strengths and weaknesses experienced by the participants in the study as first-time users.

Weaknesses

The main weakness of ECW and PUEA identified during method development is that they are more tedious and time-consuming to carry out than the original CW and PHEA methods. Based on the result of the study this suspected weakness is confirmed.

Another possible weakness identified during method development was that the method was perceived as complicated and difficult to learn, and this was confirmed by the study. Furthermore, the study showed that previous experience of other similar methods had a positive impact on ease of learning, as did substantial domain knowledge.

One other thing that may be regarded as a weakness is the dependence on the participating analysts, but the method was never claimed as being independent of the analysts' knowledge and skill. Rather, that the method should function as a framework for efficient use/application of expert competence.

Strengths

The main strength of ECW and PUEA identified during method development is that the methods provide a clear description of usability problems and user errors so they can be counteracted in the next step of the development process. The result of study confirmed that users find the method to elicit useful information for product development.

A third strength identified during the method development was that matrices should make it easier to compare the usability of two products. However, since comparisons between products were neither a part of the study, nor mentioned in the interviews, this strength cannot be confirmed nor rejected. What can instead be said is that some of the respondents did not understand how and why to use the matrices. This is a natural response if you do not use the method in a case when there is a benefit to be obtained from the matrices.

Aim of method

One important aim during development of the ECW and PUEA methods were that the method was not designed to be an optimisation of resources versus detected conceivable use errors, but that it should detect as many problems as possible. This approach is one reason for the method being perceived as tedious and time-consuming. Many of the participating students did not think that using the method produced a good result since they did not detect large numbers of errors and problems. In a real case that is a positive outcome. If you do not discover a lot of possible usability problems and use errors it means that the probability is high that the product is a safe and useable product. However, an uncertainty if the method was performed correctly and/or a disbelief that you have not discovered the possible problems and errors, could make the evaluators feel less trust in the method.

Method context

ECW and PUEA were not designed to be a standalone method but rather to be an integrated part of the product development process, including also other methods and tools. Much of the information needed for the interaction analysis in ECW and PUEA should, for instance, be available before using the method, since the basis for a user-centred product development is the understanding and description of the users and of the tasks. Consequently, preparatory work for ECW and PUEA should already have been completed.

As stated before, the method is highly dependent on preparatory work, i.e. the understanding and description of the users and the tasks. This is shown in the responses from the participants as most problems (main part from the time consumed during analysis) are related to difficulties in the preparatory work. This knowledge is the foundation for human factors work (Norman and Draper, 1986, Chapanis, 1985) and should be present in each project. ECW/PUEA is useful for uncovering these deficiencies, but cannot be fully deployed when they are present.

User skill

The answers were collected from first-time users of the method, which means that a considerable amount of the time spent on the method was learning it and developing a basic skill. Efficiency in employing the method will probably increase when the users have learnt the technique properly.

The authors' experience is that application of the method will be quicker with skilled moderators and when there is no need to think about or discuss how the method should be applied. The focus becomes the identification of problems and errors rather than managing the method and following the instructions. One minor drawback of the method is that it needs both skill in using the method and knowledge of the domain to work properly. However, this is a consequence of the conscious choice to create a method in which the quality of the result is in focus. There is not only a need for methods for Human Factors novices, but also a need for methods for HF experts.

Method evaluation

This study shows that it is possible to include developers as users in method development. The survey and the interviews made it possible to collect the product developers' experience of using a method and then comparing that result with proposed properties from the method development. This indicates that principles for Human Factors Engineering also apply for development of Human Factors Engineering methods.

A pre-assumption behind the study was that a method can be regarded as any product and, hence, evaluated in the same way. The study supports this assumption as the developers could assess the method in a similar way that a user considers a product. Method evaluations often examine how well the method performs objectively or how good the result is when used in actual projects, but it is also necessary to examine the method's performance as perceived by its users and how easy the method is to use. In the same way as the usefulness of a product can be considered in terms of the two aspects of utility and usability, (Nielsen, 1993), both utility and usability must be considered. When developing methods to be used outside the academic world it is necessary to study product developers or other industry users performing the method, and not forget to study the usability aspect.

It is crucial that the method is perceived as easy and reliable to use if it is going to be used in industrial projects. If no users want to or can use the method, it does not matter how good it is from an objective viewpoint. The participants in this study indicated that they would like to use the method in future projects, which is a good indicator that it was perceived as being sufficiently simple and reliable.

The way users perceive the results produced when using the method cannot directly be converted into a conclusion on how good the method is from an objective viewpoint, but it should give some indication. Users should possess some ability to assess the quality of the result. The participants in the study estimated that the results were of good quality, indicating that the method works well.

The study can be seen as part of the method validation with a focus on the users' experience of the method's usability. Additional studies are nevertheless needed to investigate the overall validity (the usefulness) of the method. It is necessary to verify whether ECW/PUEA detects a large number of existing usability problems and user errors and that it does not detect too large numbers of non-existing usability problems and user errors.

Conclusions interview study

The ECW/PUEA methods were perceived as a meaningful and well worth using but the methods were also perceived as tedious and time-consuming to perform. The output result of the methods depends on the quality of the input, such as knowledge about the user (user profile) and how the product works, this information is crucial for an analytical evaluation with good results.

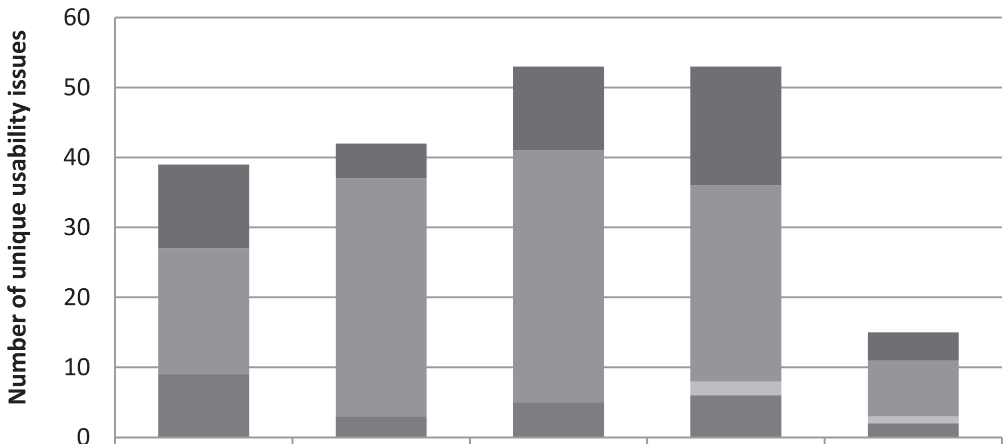
6.2.4 Theoretical comparison of ECW

In a paper by Mahatody et al. (2010) comparisons are made between the 14 versions of Cognitive Walkthrough, including ECW. The aim of the comparison was to make it easier to choose the right type of CW version depending on the evaluation to be performed. The authors' conclusion regarding ECW was that: *"It is integrated into a more global approach to complex systems."* This conclusion coincides well with the author's intention to make the method more applicable to more complex human-machine systems than that for which CW was originally designed.

6.2.5 Comparison of ECW and PUEA with other methods

In evaluations of electronic medical records as described in section 5.3.3 there was also a comparison of five different usability inspection methods: Enhanced Cognitive Walkthrough (ECW), Predictive Use Error Analysis (PUEA), Semiotic Inspection Method (SIM) and two sets of Heuristic Evaluation (HE). The comparison used four categories from the *User Action Framework:* Planning, Translational, Physical Action and Assessment. Their results are shown in figure 6.1. If ECW and PUEA could be seen as one method they would jointly have 62 usability issues and thereby be the best method of evaluation.

The following was written about the result in figure 6.1: "*The five methods with their different perspectives on human-computer interaction helped to reveal different aspects of interaction. ECW uncovered planning issues that directly influence learning and exploration of the system; it identified serious problems with text and icon, hidden controls or information, unnatural sequence and insufficient or ambiguous feedback. SIM revealed problems with translation of clues due to inconsistent and insufficient communication of the importance of objects and relations between objects. PUEA identified potential serious errors to which the interface can lead. HE1 predicted that the system will not meet the needs and expectations of users in terms of mental workload, the vocabulary used in the system and the workflow that the system imposes. The issues mentioned negatively influence the planning of actions. HE2 demonstrated that the evaluated interface does not follow the usability principles related to recognition but rather recall and visibility of system status, which leads to many problems with translation, i.e. deciding what object should be manipulated in what way*."



	HE1	HE2	SIM	ECW	PUEA
Planning	12	5	12	17	4
Translation	18	34	36	28	8
Physical action	0	0	0	2	1
Assessment	9	3	5	6	2

Figure 6.1 Percentage identified by an empirical test and predicted by CCPE (from Tancredi et al (2012))

Regarding ECW and PUEA the authors concluded that "*ECW appeared to be the most universal method; it helped to discover usability issues within all four categories of the User Action Framework*" and further "Only *ECW and PUEA helped to reveal issues related to physical action*." They described another advantage of ECW as follows: "*ECW involved physicians in the inspection work and it is therefore possible that more domain-specific issues were revealed*". In this case the possibility of including users in the evaluation is beneficial.

These are thus strong indications that ECW and PUEA work well for other evaluators. This means an improvement compared to other evaluation methods. This particular perspective has not previously been highlighted in any of the other evaluations.

6.2.6 Experience of use

In addition to the above-mentioned evaluations, the author has let two experienced users of the CCPE methodology communicate their thoughts on their experiences with the system, and the author himself will also present his own thoughts.

User 1

"*My main criticism of the methods is that their extension does not contribute enough value in comparison with CW and the methods on which PUEA is based to justify the extra work you put in to rate and categorise. It takes so much energy to remember what all the small figures and signs mean that it just gets tedious and wrong in the end".*

"It feels as though what gives CW, and perhaps other methods about which I do not know as much, such a high false alarm ratio is that you do not do the necessary preparatory work properly. Something that is more symptomatic for a general inconclusive product development process than major flaws in the methods."

"Getting started with a method requires a lot of commitment and setting aside of uninterrupted time which I think it is difficult to get from many of today's businesses, there is a big obstacle to get over in order to get started. As it requires training and a lot of knowledge to perform the methods in an effective way, perhaps a company can create a type of "task force" that comes in when an analysis is needed. With a special unit, on the other hand, you will miss the consensus the method creates among the participants. Are there risk analysis consultants? The question is also how often the methods are applicable. Are there enough opportunities in a company's product development work for people to train their ability within a reasonable time interval?"

"I've never understood the point of the matrices, for me it is not a useful way to present the results. What do I do with them? As a designer, I want to know what has to be fixed and in what way, not that the product generally has problems with recovery during common tasks; I want to know exactly what tasks have problems with recovery. The methods lack good ways to present the data."

"I can see advantages of using both methods at the same time, because it does not take that much longer to use both once you sit down with one."

User 2

"I have used the further-developed ECW and PUEA methods in several projects where we have been interested in a deeper analysis of user friendliness and risk of use error. The products evaluated were mechanical artefacts managed by people and they have high demands on safety. The methods are well adapted to that type of application and can be modified if necessary. The results that emerged underscored both major and minor defects in the product's user friendliness and many of them were remedied by simple means. It is obvious that each step in the process must be followed, and well-executed preparatory work allows more defects in human-machine interaction to be detected and corrected."

The author

"My first impression of the methodology is that it is very powerful when it comes to detecting problems and errors in the interaction, but that it is at the same time complex to initially embrace because it contains so many parts and must be adapted to function. I myself do not consider the methodology to be tedious or time consuming but I assume that is because I am thoroughly familiar with the methodology and can therefore use it effectively. On the occasions when I have moderated sessions it has been possible to maintain a good pace since there have been no doubts about how the method is to be used or what everything means. One reflection is that the methodology requires good knowledge of the user, the use and the machine to operate efficiently, but it is easy for ignorance in these areas to cause long discussions among the session's participants about how things really are. However, one of the positive aspects of the methodology is that it helps participants to gain a common picture of the whole."

After studying how CCPE methodology is applied in use with products and technical systems, the next step in the evaluation is to reflect on the advantages and disadvantages of the approach that the methodology is based on, which is done in the next paragraph.

6.3 Reflection on method development and method approach

The development of the CCPE methodology and its methods has come to be led by three general approaches in evaluating the interaction between human and machine: (1) the proactive approach, (2) the analytical approach, and (3) to combine the cognitive and physical aspects of the interplay in a joint evaluation. The section below offers reflections concerning the three approaches.

As described in Chapter 4.3.5 there are different approaches to deal with mismatches. A reactive methodology demands that something must happen in a test or the field before measures are taken, which means that the machine first has to be manufactured before a reactive approach can be applied. It is both easier and less costly to change and improve the machine with regard to ergonomics and usability during the development process than to modify the manufactured machine when it is out on the market and is handled by users. Early in the product development process there is seldom a physical prototype. A proactive method can detect the problems in earlier stages of product development, and the less detailed prototypes the method can evaluate, the more proactive the method becomes. CCPE does not need a fully manufactured machine and the intended context to search for mismatches in the interaction, since it work with both high and low fidelity prototypes, as well as abstract descriptions of the machine and a fully functional machine.

An important factor for a good evaluation of the interplay between human and machine is the involvement of users in the evaluation (Engelbrektsson, 2004, Garmer, 2002, Launis, 2001). One way to achieve this is through empirical evaluation in usability tests, which can be performed successfully during the whole product development process with different types of prototypes. But with the aim to do an extensive and systematic search to cover the whole range of presumptive mismatches in the interaction between human and machine, the empirical test become less effective if the prototypes are of low fidelity or if there is only an abstract description of the machine. A low fidelity prototype may hide some presumptive mismatches while creating mismatches that are not real. In addition, even if a high fidelity prototype does exist it is often difficult to obtain sufficient numbers of potential or actual users for large scale empirical test method to perform an extensive and systematic search for mismatches.

There are however other ways than empirical tests to involve users as evaluators than in the product development process, for example in focus group interviews (Obert and Forsell, 2000). To counter act the limitations with empirical tests, CCPE takes an analytical approach, which is less dependent on involving many potential or actual users compared to empirical test methods. The structure of the CCPE makes it possible to evaluate low fidelity prototypes as well as abstract product descriptions. But, CCPE's analytical approach does not exclude the user from participating in the evaluation, on the contrary CCPE is constructed to utilise users' knowledge through their participation.

Furthermore, CCPE is not entirely an analytical methodology since all information required to perform the analysis comes from empirical studies. Without basic and accurate information about users and use CCPE is not useful. The strength of CCPE is in gathering all the knowledge available about the system (e.g. human, task, environment and machine) in the present time to predict what will happen in the future. Because of this, the methodology becomes better if the user is a member of the evaluation team. In conclusion, CCPE is an analytical methodology with large elements of empiricism.

The strength of employing an analytical methodology is that it provides a possibility to remove deficiencies in the machine’s design before testing with users begins. In this way the users’ time is not wasted on a deficient product design; a person who has participated in a test often cannot be re-consulted as he/she may be biased by the first test. However, it should not be forgotten, as stated before, that real users play an important role in determining the needs and requirements, which underlie the evaluated design. This is especially true for the satisfaction component of usability, which the analytical methods often do not cover. Another strength is that the analytical methodology may detect hazards in use before they have happened in reality, i.e. lower the risk associated with the machine.

One disadvantage of using proactive analytical evaluation however is that the deficiencies detected are potential and not “real” ones. To confirm that the deficiencies are real, the machine has to be actually used, but this may expose humans and machines to unnecessary hazards. Hence, a proactive analytical evaluation is crucial if the method has to be used before actual usability tests are made or if it is impossible to do empirical tests. Moreover, it is only after a deficiency has been identified that it is possible to decide whether the error is plausible or not. Exposing even improbable deficiencies for further evaluation is also beneficial, as these may have serious consequences that can otherwise be overlooked if only the plausible deficiencies are investigated.

The primary strength of analytical methods is, therefore, that an analysis can detect potential mismatches in the interaction before empirical tests with real users are necessary. The empirical tests can then be employed more effectively, once the identified use errors and usability problems are remedied. In all, this is why the CCPE methodology is useful in both the early development process as well as later, where it can support design work.

The approach of CCPE was to unite the physical and cognitive aspects of the interplay between human and machine. Such an approach is needed because there is currently an obvious separation of body and mind, which is clearly visible in the division between cognitive ergonomics and physical ergonomics. The division between body and mind goes back in history to when Plato and Aristotle, and later Descartes, made a clear distinction between the spiritual and physical (Hansson, 2011, Berminge, 1993). The division between body and mind is also clearly visible in our time. An example in academia is that psychology focuses on the mind, while medicine focuses on the body.

Division of body and mind is not preferable for the subject of ergonomics as it causes human interaction to not be seen as a whole but instead be divided into a physical and a cognitive part. An example raised in the thesis of why this division is negative deals with products that are designed to have good physical ergonomic properties, but the lack of cognitive ergonomics prevents the human from taking advantage of them. This was proved, for example, in the analysis of the office chair (Appendix D). Another example is that only because a person knows how to handle a particular product this does not automatically mean that the person will be able to perform the action. This became clear during the analysis of kitesurfing (Appendix E). Users were expected to know exactly how they would do (in theory) but it was also expected that they would fail in the execution (in practice). The problem did not lie in the cognitive understanding, but in the physical action. Without studying both the physical and cognitive aspects together, the analysis of human-machine interaction will not be complete and accurate.

The CCPE methodology is an attempt to bridge the gap between body and mind and create a coherent approach that includes both physical and cognitive aspects. The use of the methodology also shows that it is possible to combine physical and cognitive ergonomics in the same analysis because the interplay occurs in an integrated way. For evaluation during product development there are two major advantages to a unified approach. First, it is possible to make evaluations more time-efficiently as both the physical and cognitive aspects are treated simultaneously. Secondly, the evaluations will offer improved quality since mismatches in the interaction featuring both physical and cognitive aspects will be better detected.

An interesting and more general question is whether it would be appropriate to combine the disciplines of physical and cognitive ergonomics to better adapt to reality. The human is unified where there are no sharp boundaries between body and mind. Depending on the situation, there are various human factors that are worth considering in greater or lesser depth and together. One appropriate approach would be to first study the whole picture and then go into the individual parts of the system. In other words, start from a system perspective instead of dividing into physical and cognitive aspects. This is very important in cases where ergonomics have to be applied in order to change and hopefully improve the reality in which people live.

6.4 Summary, evaluation of the methodology

The goal of the work was to develop an improved Human Factors Engineering methodology which was based on existing methods and which focused presumed mismatches in the interaction between user and artefact. The specific aims of the methodology were:

- Prediction – investigating when, where and how mismatches exist
- Identification – determining the type and properties of the predicted mismatches
- Presentation – describing the identified mismatches in a manner that facilitates counteractive measures

The question is if the goal has been fulfilled. From the description of the methodology in Chapter 5 and verification in Chapter 6.1 it is shown that the CCPE: (1) is based on existing methods, (2) uses a query process to predict mismatch, (3) uses categories to identify mismatches, and (4) uses matrices to present the results. According to the verification the methodology is regarded as having met the specific aims set.

The application of CCPE in different areas is presented in Chapter 5.3 and the summary in Table 6.2 shows that CCPE has been used as intended in a variety of applications. The evaluations presented in Chapter 6.4 examined the CCPE methodology in different ways and from different angles. The conclusions that could reasonably be drawn from these evaluations can be summarised as follows:

- The methodology detects mismatch between the human and the artefact in many different application areas
- The methodology works in projects to evaluate or develop products and technical systems
- The methodology is generally perceived as useful by the users

The reflection in 6.3 showed that the proactive approach, an analytical method, and integration between physical and cognitive aspects, are a suitable way to work with mismatch issues in product development and, further, that it can support design work, as stated in the title of the thesis. However, the evaluation has also shown that there are overall pros and cons with CCPE methodology.

The major advantage of CCPE is that it is a methodology and not several methods that do not relate to one another. CCPE provides a coherent structure with methods that are developed together, making it easier to use. For example, both the physical and cognitive aspects can be considered together. Further, CCPE offers a systematic and structured way of working that supports the user of the methodology in the analysis. CCPE is also developed to be adaptable to many different dimensions and is thus flexible and suitable for use with different types of products, machines and systems at different times and for different purposes.

Because the methodology is so comprehensive and flexible, this also comes with a disadvantage. Because of its complexity and the knowledge needed by its users to manage the methodology in an optimal manner, it requires training and time to learn. To exploit the benefits of CCPE, the person conducting the evaluation (the user) needs to understand the methodology. First the user needs to adapt the methodology to the situation, possibly exclude selected areas and then implement the chosen approach. Otherwise, it is possible that the process of following the methodology structure takes over and becomes the focus, instead of the actual evaluation that the methodology is intended to support. Much of the success of CCPE thus depends on the person conducting the method, i.e. the success in using the methodology in an optimal way is dependent on the moderator's skills and abilities.

When it comes to ECW and PUEA the method development was based on deficiencies in the original methods, that is CW and AEA, SHERPA and PHEA, which the development of ECW and PUEA tries to counteract. There are, however, some remaining weaknesses in the developed methods. The main weakness, apart from the fact that the methods are not fully validated, is that the method development made ECW and PUEA more tedious, more complicated and more time-consuming to carry out. This type of criticism was previously also levelled at the original methods. For example, Miller and Jeffries (1992) state that one of the drawbacks of CW is its being tedious, while Stanton and Young (1999) say that one of the drawbacks of PHEA is that the method can be tedious and time-consuming for complex tasks. Since the method developments of ECW and PUEA largely entailed additions to the original methods, one may assume that they have become more tedious, complicated and time-consuming, but also that they yield very extensive and usable results when the analysis is completed. The final result is methods of better quality and higher usability.

Further, ECW and PUEA must not be seen as independent units. When they are employed in conjunction with other methods in CCPE and in the human factors engineering process, they become less tedious and take less time. Much of the information needed for analysis in ECW and PHEA should already be available, and the results of the analysis provide input data to other methods in the HFE process. Consequently, these other methods should, in turn, take less time to carry out. One fear is that, if ECW and PUEA have become more tedious and time-consuming, they will be difficult for users to learn, accept and adopt. And seen as an independent unit the methods are harder to learn and to use. But if persons are used to work with structured and systematic methods ECW and PEUA will be perceived as less tedious and time-consuming. When the methods were taught to students in advanced master courses in Cognitive Ergonomics and Human-Machine Systems, as well as in Master theses work; the students did not perceive the methods as very difficult to understand and employ. The methods were for them one part of a larger integrated method package.

Another way of considering the above-mentioned aspects is connected with safety and when CCPE is used in more safety-critical areas. To reach as high a safety level as possible, the need is not for methods that quickly detect the most frequent problems. Methods are needed that also discover the infrequent problems; even a seldom-occurring use error can kill or injure a person. As pointed out earlier in the thesis, ECW and PUEA employ a systematic approach specifically in order to detect as many potential usability problems and use errors as possible. Since the methods can be applied with the aim of avoiding injury and death, their being relatively tedious and time-consuming should be outweighed by the increased safety.

It was stated earlier that CCPE can take time to execute, but a method should also be considered in cost-benefit terms. During CCPE the user can be involved in an evaluation in a different way than as test subjects in empirical usability tests. Users can participate as members of the evaluation team and they can share their knowledge and discuss aspects in a more extended way than in an ordinary usability test. The users can thus be involved in a more comprehensive way. CCPE provides a detailed description of how to detect mismatches, which can be used to improve the evaluated artefact. Further, the reflection described in 6.4 underscores that a major advantage of the CCPE methodology is that it does not need access to detailed and functioning prototypes but instead works with simpler forms of representation.

Another advantage is that the qualitative results from CCPE can be useful information for those who design the user manuals and educational materials, since the methods show where it can be difficult to act correctly (usability problems) and where it is possible to act incorrectly (use errors). If usability problems and use errors cannot be counteracted through the design, the only alternative is to inform and train the users.

The major disadvantage of CCPE is that the mismatches that are discovered are merely presumptive and that the methodology only indicates that they may occur during actual use. As CCPE is designed to detect as many presumptive mismatches as possible, this means that many of the mismatches detected will not occur during actual use. That too many presumptive errors are detected means, on the other hand, that all probable errors will very likely be found in the evaluation. Nevertheless, CCPE needs to be complemented by other methods, such as usability tests and focus groups, to determine the reasonableness of the detected mismatches. The comprehensive results of CCPE, however, promote knowledge about where in the interaction most problems can occur, so this part can then especially be examined with a usability test if desired.

To sum up, the main strengths and weaknesses derived from the use and evaluations of CCPE are as follow.

Strengths:

- Systematic, structured and coherent way of working that includes physical and cognitive ergonomics
- Adaptable methodology with considerable flexibility
- Comprehensive result without having to perform empirical tests with users or detailed prototypes
- Detailed description of discovered mismatches that can be used to improve the machine
- Information about where in the interaction most problems exist, which serves as a basis for usability test

Weaknesses:

- Large framework to grasp in order to use the methodology in an optimal way; use can then become complex and time-consuming
- Knowledge and experience are needed to adapt the methodology to suit the specific situation
- Dependent on the moderator if sessions are to be effective
- Since possible mismatches are the result of a theoretical judgment by evaluators, they do not need to occur during real use
- Need to complement CCPE with other methods such as usability tests and focus groups to determine the probability of detected mismatch

6.5 Comparison between methodology approaches

CCPE represents an approach based on an analytical method perspective, but also with significant contributions from human factors (HF) and expert users. Other approaches with more or less structure, as described in Chapter 4, include a usability test in which actual users are viewed and heuristic evaluation by HF experts. It is interesting to compare CCPE methodology with these other approaches to see what the pros and cons are.

In order to compare CCPE with other methodologies, a range of approaches were examined by the author:

- No work to detect and identify mismatch in the interaction
- Unstructured evaluation by users, represented by the user review (UR) method
- Unstructured evaluation by HF expert, represented by the HF expert review (ER) method
- Structured evaluation by users, represented by the usability test (UT) method
- Structured evaluation by HF expert, represented by the heuristic evaluation (HE) method
- Analytical evaluation, represented by the combination[3] of Hierarchical Task Analysis (HTA), Cognitive Walkthrough (CW) and Predictive Human Error Analysis (PHEA)
- The developed methodology, Combined Cognitive and Physical Ergonomics (CCPE)

6.5.1 Method evaluation properties

Table 6.7 lists a compilation made by the author of the properties of the method that have been discussed in the theoretical framework in Chapter 4. Each of the seven methodological approaches above was then assessed according to:

- Y= yes, property included
- N = no, property not included
- P = reasonable possibility to include

Table 6.7 Comparison of method properties

Method property	None	UR	ER	UT	HE	CW PHEA	CCPE
Can be employed proactively	Y	Y	Y	Y	Y	Y	Y
Can be employed formatively	Y	Y	Y	Y	Y	Y	Y
Includes actual users	N	Y	N	Y	N	P	P
Includes HF experts	N	N	Y	N	Y	P	P
Includes employment of theory	N	N	P	N	Y	N	P
Includes realistic use tasks	N	N	N	Y	N	Y	Y
Includes active and exhaustive search	N	N	N	N	N	Y	Y
Task-independence	N	Y	Y	N	Y	N	N
High-level perspective (Top-Down)	N	P	P	P	Y	N	N
Low-level perspective (Bottom-Up)	N	P	P	Y	P	Y	Y
Gives qualitative output data	N	Y	Y	Y	Y	Y	Y
Gives semi-quantitative output data	N	N	N	Y	Y	N	Y
Gives quantitative output data	N	N	N	Y	N	N	Y

When the table (6.7) is studied from left to right a pattern emerges with more method properties met to the right, where more advanced methods are presented, than to the left,

[3] These methods are the base for CCPE

where the simpler methods are shown. For CCPE only two properties are not included, to be task-independent and to have a high-level perspective. The approach that is closest to CCPE is Usability Testing (UT) with four missing properties and Heuristic Evaluation (HE) with five missing properties. The comparison in Table 6.7 shows that CCPE meets the set goals better than the other six, already existing, approaches to which it is compared.

6.5.2 Strengths and weaknesses of the respective approaches

The different approaches have different advantages and disadvantages in their actual execution, which may be interesting to study in comparison with CCPE. This has also been done by the author and presented in Table 6.8. The list originates from Stanton and Young (1999), Stanton et al (2005a), Stanton et al (2005) and from the author's own experiences.

Table 6.8 Comparison of strengths and weaknesses of the methods

Strength	Weakness
None	
No learning needed No resources needed	No mismatch is detected
User Review Represents the unstructured evaluation of user approach	
Not complicated and short learning curve, the method is well known to most Includes real users A structured interview provides consistency and accuracy Flexible method as follow-ups can be made during the evaluation	No structured search for mismatch Very dependent on the user's ability to find mismatches Does not consider knowledge of the subject area The analysis can be time-consuming The circumstances of the review may lead to erroneous results
Expert Review Represents the unstructured evaluation of HF expert approach	
Not complicated and short learning curve, the method is well known to most Considers knowledge of the subject area A structured interview provides consistency and accuracy Flexible method as follow-ups can be made during the evaluation	No structured search for mismatches Highly dependent on the expert's ability to understand the use The analysis can be time-consuming The circumstances of the review may lead to erroneous results
Usability Test Represents the structured evaluation of users approach	
A simple and flexible way to evaluate usability Systematic process that has control over the variables Possibility of obtaining quantitative data, and not only qualitative Involves users Based on natural human behaviour Provides a picture of how real use is performed	Time-consuming to perform, can be costly Much data is collected which can lead to time-consuming analysis Needs training to plan and carry out the test well Hard to get appropriate users to test with Needs some sort of prototype to be effective Can be difficult to generalize as it is a special situation that is tested Can be difficult to know why something goes wrong and to find all possible errors

Heuristic Evaluation Represents the structured evaluation HF expert approach	
Simple method to perform and requires little training Resource-efficient method Takes known knowledge of the subject area into account	Very subjective assessments Unstructured and unsystematic Lack of reliability, no comprehensive analysis and difficult to examine
Hierarchical Task Analysis, Cognitive Walkthrough and Predictive Human Error Analysis Represents the analytical evaluation approach	
HTA Simple, fast and general method to learn and to use Provides a comprehensive description of a task Flexible in its use as detail depth can be varied Creates a structure for further analysis by other methods **CW** Has a structured approach Easy to learn and does not need deep knowledge of cognition Also useful for HF experts Ability to analyze proposals early in the development process (no need for a testable prototype) Based on cognitive theory Relatively efficient in terms of time and resources Effective in finding significant and realistic usability problems **PHEA** Structured and comprehensive procedure, which is simple to follow Taxonomy leads to an analysis of possible errors Provides relatively valid and reliable data A lot more time-efficient compared to observations	**HTA** Provides mostly descriptive and not analytical information, cannot be used alone to provide design solutions Can be difficult to get cognitive elements of the task Can be time-consuming if it is a complex task that is to be described **CW** Limited focus, studying only ease of learning The results need to be validated by other methods Can be time-consuming for complex tasks The results are based on subjective assessments Needs a task analysis as a basis Low coherence between different evaluators **PHEA** Can be tedious and time-consuming for complex tasks Needs a task analysis as a basis Does not take the cognitive components in use error into account Some discovered possible errors will not occur Taxonomy shortfall, being general in nature **Generally** As possible mismatches are judgments by the evaluators, they do not need to occur during actual use Need to complement CCPE with other methods such as usability testing and focus groups to determine the reasonableness of detected mismatch

Combined Cognitive and Physical Evaluation The developed methodology	
Systematic, structured and coherent approach that includes both physical and cognitive ergonomics Adaptive methodology with high flexibility Extensive results without having to perform empirical tests with users or detailed prototypes Detailed description of the detected mismatches can be used to improve the artefact	Large framework to learn to use the methodology and optimal use can be complex and time-consuming. Knowledge and experience are needed to adapt the methodology Depends a lot on the moderator for the sessions to be effective As possible mismatches are judgments by the evaluators, they do not need to occur during actual use Need to complement CCPE with other methods such as usability testing and focus groups to determine the reasonableness of detected mismatches

The summary in Table 6.8 shows that each methodological approach has its distinct advantages and disadvantages, which can be related and that they belong to three areas:

- Practice and competence of the executor
- Time and resources to execute the method
- Quality of the results from the method

In general, the methodological approaches presented in the upper part of Table 6.8 do not require as much training and time, but they also do not generate such high-quality results. The methods described further down in the table require more expertise and resources but generate at the same time somewhat better-quality results. For the development of CCPE in particular, the quality of the result was the most important factor of all, primarily to find the mismatches that may have the most serious consequences. While it is important that a method is easy to learn and simple to use, during the development of CCPE no method adjustments were undertaken that would impair the quality of the result. The idea behind the development of CCPE is that as many as possible of the mismatches should be found, including the most serious mismatches, and that the identified mismatches are completely realistic. Many of the weaknesses of CCPE are thus an effect of the deliberately chosen direction for the method development.

The question is, however, whether the selected path is successful. As regards searching for potential mismatches, the active and exhaustive character of the method is essential. The author's own experience is that the search for potential use errors can be compared with looking for needles in a haystack. The key here is not to find the needles but to ensure that no needle will proceed to animal food where they can do serious harm. One has to find not only the big needles that are easily detectable on the surface of the haystack, but also the small needles that lie far below. The only way of doing it is to look actively and exhaustively through the entire haystack. A humorous analogy might be that seeking potential use errors with usability tests is like letting unknown people jump on selected parts of the stack in order to feel for needles. Similarly, heuristic evaluation means letting some “needle experts” search through the stack according to a general pattern of where needles tend to be found. However, usability tests and heuristic evaluation can be employed to investigate whether it is possible to jump in the stack. If these methods discover needles, the stack is definitely not a good one to jump in. To summarise, it is important to find all the needles, not just most of them. The last needle is also important, and if you find a few straws and think they are needles, and then throw them away, it does not matter so much.

6.6 Complementing methods

Chapter 4 describes the four areas of the CCPE evaluation methodology: Evaluation of Risk, Usability Evaluation, Human Reliability Analysis, and Evaluation Physical Interaction. The methods useful for each respective area were developed into CPPE, but the basic ideas from these areas have also been used. The main part of the methodology was Usability Evaluation (with CW) and Human Reliability Analysis (with PHEA) but the development was also designed to allow the methodology to work together with the Evaluation of Risk and Evaluation of Physical Interaction. CCPE is not developed to function as a single tool but to facilitate interaction between the methods of the four areas. CCPE can be assigned to the group of usability evaluation methods.

CCPE is not the only methodology that should be used in the human factors engineering process. Other methods are needed to be used in series and in parallel. There are methods that provide input data to CPPE as well as methods that use the output data, and there are other ergonomic evaluation methods, which work in different ways compared with CPEE.

6.6.1 Input methods

The methods providing input data to CPEE, to the system description, must elicit information about the users, the artefact, the task and the context. As pointed out in the description of CPPE, it is the quality of these input data that determines the quality of the result of the analysis. If the input data are deficient, the result of the analysis using CCPE will also be deficient. The main methods for this are observation and interviews, but other methods can also be suitable.

In the workload analysis, with GTS, the methodology is dependent of good assessment. This can be done by expert judgement but other methods can also be used. Examples of methods for analysis of the user's mental workload are SWAT (Meshkati et al., 1995) and NASA-TLX (Hart and Staveland, 1988). For more background information and a deeper understanding of how the user thinks, the Applied Cognitive Task Analysis (ACTA) method (Militello and Hutton, 1998) can be employed. For the physical workload, poster analysis methods as RULA and REBA can be used, although more advanced simulation software as well as biomedical calculations can also be used.

6.6.2 Parallel methods

When evaluating the ergonomics of a machine there may be a need for methods other than CCPE in order to investigate problems and errors in the interaction. This is due not to deficiencies in CPPE but to the methodology has an analytical, task-based, low-level perspective in the analysis. Hence, to obtain a more comprehensive approach for usability, supplementary and triangulating methods are required that have other ways of dealing with the analysis (compare Table 6.2) and different strengths and weaknesses (compare Table 6.3). Moreover, it should be noted that the design of an artefact does not automatically yield good usability just because no usability problems or use errors are discovered with CCPE; however, if the methodology detects many serious problems and errors, it is a strong indication of deficient ergonomics in the evaluated equipment.

CCPE methodology has an analytical, task-based, low-level perspective in evaluation and to obtain a more comprehensive approach for usability problems and use errors supplementary and triangulating methods with other ways of dealing with the analysis are required. The methods that are supplementary and triangulating for ECW and PUEA should be task-independent, have a high-level perspective and, if possible, involve the users. In some studies where ECW has been used (Liljegren et al., 2003, Bligård et al., 2003b, Bligård et al., 2004), it has been supplemented by Heuristic Evaluation (HE) and/or Usability Tests (UT) with good results. This combination of (E)CW, HE and UT is also recommended by the researchers who developed CW (Lewis and Wharton, 1997). They emphasise: *"We think a practical usability process needs to include both task-independent evaluation, for which Heuristic Evaluation would be our choice, and task-specific analysis of key-tasks, using a combination of CW (early) and user testing (later)."* In addition, employing supplementary and triangulating methods enables the identification of more aspects of usability to be evaluated (cf. Garmer et al., 2002).

Tancredi et al (2012) write the following about ECW and PUEA in relation to other methods: "*The five methods with their different perspectives on human-computer interaction helped to reveal different aspects of interaction. Work with categorisation of the usability issues showed that the selection of inspection method affects the type of problem that is revealed. ECW reveals more planning issues and SIM and HE2 help to find more translation problems instead."*

Since it is obvious that CCPE needs support from other methods, the question is whether these supplementary methods can replace CCPE. Table 6.9 shows a comparison between CCPE, UT and HE. All the listed methods can be employed proactively and formatively, so they are applicable during the development process – but it is impossible to replace CCPE with UT or HE and to search for problems and errors in such an active and exhaustive way. Conversely, neither can CCPE replace UT, since CCPE does not include experiments with the users who provide actual problems, errors, judgments or experiences. Nor can HE be replaced, since CCPE is not an evaluation and assessment of user interface design theory, i.e. so-called usability heuristics.

Table 6.9 Comparison between properties of interaction analysis, usability testing and heuristic evaluation (derived from table 6.7)

Method property	**Interaction analysis**	**Usability testing**	**Heuristic evaluation**
Can be employed proactively	Yes	Yes	Yes
Can be employed formatively	Yes	Yes	Yes
Includes experiments with users	No	Yes	No
Includes employment of heuristics	No	No	Yes
Includes active and exhaustive search	Yes	No	No
Task-independence	No	No	Yes
High-level perspective (Top-Down)	No	Possible	Yes
Low-level perspective (Bottom-Up)	Yes	Yes	Possible
Gives qualitative output data	Yes	Yes	Yes
Gives semi-quantitative output data	Yes	Yes	Yes
Gives quantitative output data	Yes	Yes	No

Table 6.9 shows that a combination of UT and HE will be task-independent, have a high-level perspective and involve the users. UT includes experiments with users who provide actual problems, errors, judgments or experiences. HE has a high-level perspective and includes assessment of user interface design theory.

When investigating use errors the goal is to determine whether a probability exists, in both common and rare cases that can lead to negative consequences. If it is a safety-critical case one important point is to ensure that the machine does not cause damage even in rare cases. To achieve this, there must be an active and exhaustive study of the interaction, which is not accomplished by UT or HE. Moreover, CCPE implies that the analyst plays the role of a user and that the task for evaluation is defined, which is not the case in HE.

It should be noted that the design of a machine does not automatically yield good usability just because no usability problems or use errors are discovered with the CCPE methodology; however, if the methodology detects many serious usability problems and use errors, it is a strong indication of deficient usability of the evaluated equipment.

6.6.3 Output methods

CPPE primarily identifies individual usability problems and use errors, which mean that they belong to the Human Error Identification (HEI) group, where CPPE is assigned to the Human Reliability Assessment (HRA) area. To follow the methodology in HRA, methods for Human Error Probability (HEP) are a natural continuation of CCPE in risk analysis. ECW and PUEA do not calculate the probability that a user will commit errors. Examples of methods that calculate probabilities of human error are HEART (Human Error Assessment and Reduction Technique) (Williams, 1986) and Technique for Human Error Reduction (THERP) (Swain and Guttman, 1983).

Information from the CCPE can also be included in risk analysis methods at a higher system level. Frequently employed methods having this aim include FTA (Fault Tree Analysis) and ETA (Event Tree Analysis) (Stricoff, 1996). FTA has for example been used in human factors work for improving patient safety for radiation therapy systems (Israelski and Muto, 2005). The risk analyses themselves have, in turn, no intrinsic value – the results of all methods and analyses are fed back into the usability engineering process, and thereby to the product development process in the form of requirements or design changes. The goal is to create a safer product.

6.7 Relationship with research and engineering areas

That CCPE should function together with the reference framework's theory and processes/methods was a requirement in the methodology development. This means that ECW and PUEA should fit into, and work co-operatively with, other methods in these processes.

- Human factors and ergonomics
    - Physical ergonomics
    - Cognitive ergonomics
- Human factors engineering
- Risk management
- Product development

6.7.1 Human factors and ergonomics

CCPE methodology works in the areas of Human Factors and Ergonomics as its goal is to optimize overall system performance by discovering the mismatch between user and artefact. With use errors and usability problems arising, the human-machine system functions less efficiently than it should. The methodology also relates to human well-being as problems in the interaction can affect people negatively, both physically and mentally.

CCPE is specifically related to Physical Ergonomics by considering the human physical prerequisites in the interaction, but also by bringing out the physical effects that problems and errors can have on humans. The goal of good physical interaction between user and artefact is to prevent musculoskeletal disorders (MSD). CPPE's relation to Cognitive Ergonomics is clearly identified as a GTS map mental workload while interaction analysis is based on cognitive theory and tries to simulate the user's thoughts.

6.7.2 Human factors engineering and user-centred design

The application of CPPE is to be seen as part of a larger human factors engineering process in which several methods are employed, which was discussed earlier. As noted in the frame of reference, the central activities are collection of data from and about the users, conversion of these into needs and requirements, and design of the technical system with regard to both utility and usability.

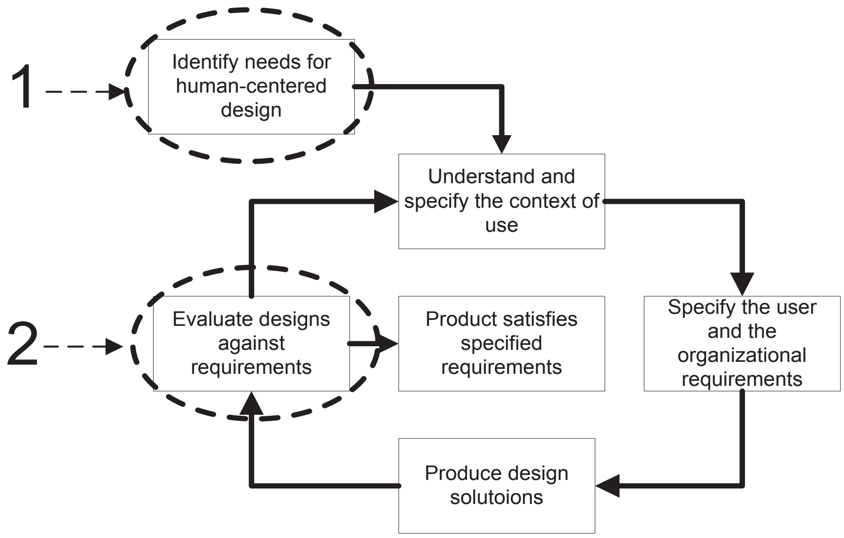


Figure 6.2 Application of CCPE in ISO 13407 human-centred design activities (ISO, 1999)

Related to ISO 13407 human-centred design activities, the CCPE-methodology can be employed on two occasions in the process (numbered 1-2 in Figure 6.2). The first (1) is to evaluate existing equipment during the first activity in order to find existing mismatches. This information is then taken as input data for the requirement specifications and for production of design solutions. Next, (2), is an analysis of the design during the process. CCPE analyses can be conducted in each of the iterations, from concept solutions to final prototypes to evaluate if the product satisfies the specified requirements.

Related to the IEC usability engineering process, described in Figure 6.3, CCPE can be employed on three different occasions in the process (numbered 1 to 3). The first (1) is to investigate (during the initial phases of the process) existing equipment in order to find existing usability problems and potential use errors. This information is then taken as input data for the design of the new equipment. Next, (2), is an analysis of the artefact under development. CCPE analyses can be conducted with equipment throughout production of the design, from concept solutions to final prototypes. The last application, (3), is as a validation tool to confirm that the equipment released does not contain too serious potential use errors or usability problems. However, CCPE is never intended to replace validation with real users or to be the only method in the development process.

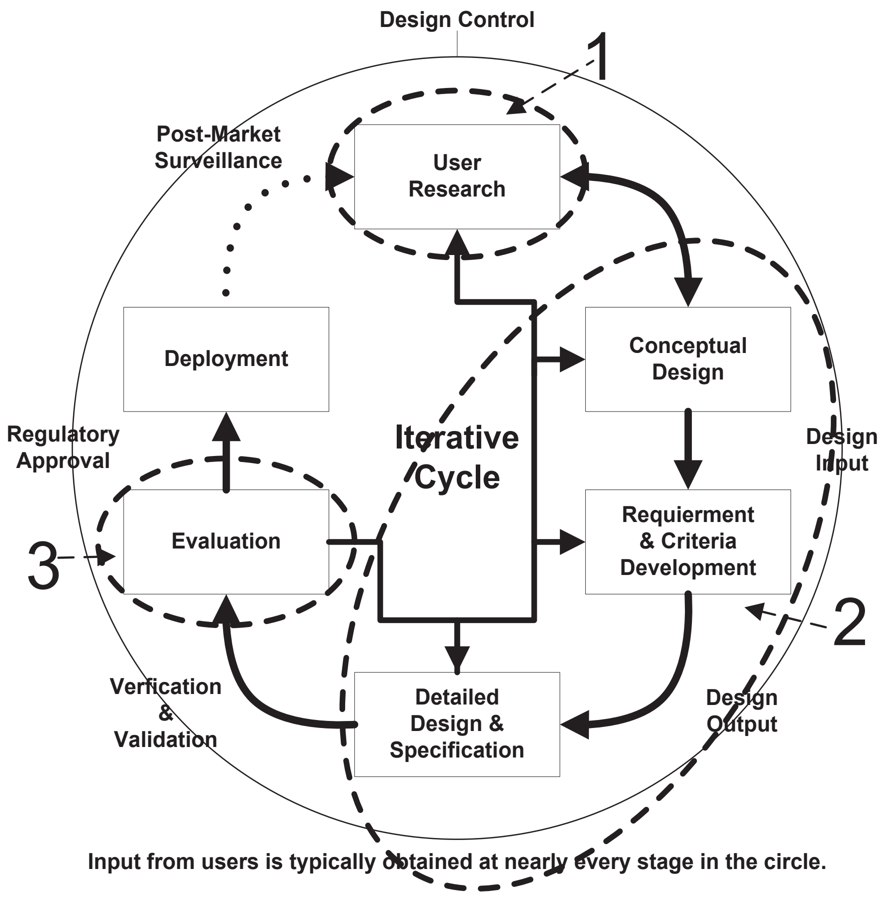


Figure 6.3 Application of CCPE in the IEC usability engineering process (IEC, 2004a)

6.7.3 Risk management

Furthermore, CCPE can be related to risk analysis and the risk management process when the prevention of usability problems and use errors improves safety. The methodological coupling between risk, use errors and usability problems becomes clear when the overall processes for their management are compared. As the survey in Chapter 4 showed, risk handling is undertaken by identifying hazards and estimating the risks of hazards. If a risk is too high, measures are taken to reduce it. It is then kept under surveillance and, if necessary, a new assessment of it is made (Figure 6.4). As for the handling of mismatches (Figure 6.5), it begins with identification of a mismatch. Next a judgment is made of how severe the mismatch is, and if severe enough it is remedied. The mismatch is then kept under surveillance, and assessed anew if need be.

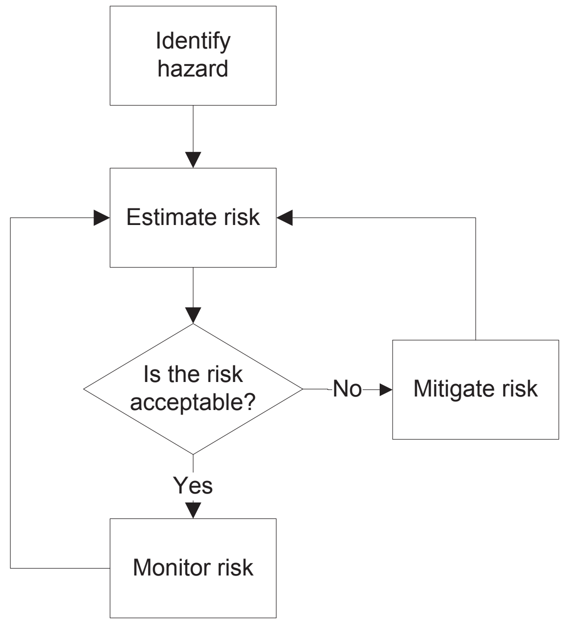


Figure 6.4 Generic Risk Management Process Adapted from Stricoff (1996)and ISO (2000b)

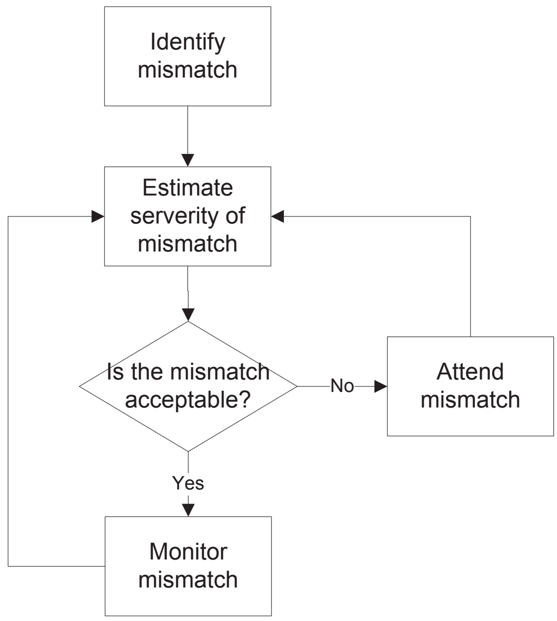


Figure 6.5 Process for managing mismatches in the interaction. Derived from Figure 6.4

Working with risks and mismatches therefore has much in common, which demonstrates that risk management is coupled with use error and usability problem management. It should thus be possible to integrate these processes and handle their respective issues in parallel in the product development process.

6.7.4 Product development

For CCPE to be really useful the methodology must work within the framework of product development. The methodology can be applied in four different ways in a development process, three formative and one summative (Numbers 1-4 in Figure 6.6).

The first way is evaluating existing machines on the market (formative) (Number 1 in figure 6.6). The CCPE methodology is then used to *identify needs* in the development process, with the aim of informing the requirement specification of a new machine. In this case both the function and operation levels of the interaction analysis are employed (Chapter 5, Tables 5.1 and 5.6).

The second way is to evaluate concepts of the machine (formative) during *the conceptual design phase* in the development process (Number 2 in figure 6.6). The aim here is to examine the design for possible improvements. In this case, only the function level of the interaction analysis is employed (Chapter 5, Tables 5.1 and 5.6).

	Johannesson et al, 2004	Ullman, 1997	Ulrich and Eppinger, 1995	Bligård, 2011
1	Prestudy	Identify Needs	Concept devlopment	Needfinding
	Product Specification	Develop engineering specifications		Function and Task Design
2	Concept Generation and Evaluation/Selection of Concept	Develop concepts		Overall Design
3			System-Level Design	Detailed Design
4	Layout Design	Develop Product	Detail Design	Structural Design
	Detailed Design			
	Prototype Testing		Testing and Refinement	
	Adaption for Production		Production Ramp-Up	

Figure 6.6 CCPE applied in the development process

The third way is to perform an evaluation of the detailed design prototypes of the human (formative), which is done during the *detailed design phase* in the development process (Number 3 in figure 6.6). The aim is to detect mismatches in the interaction before the machine is implemented, in both software and hardware. Here, both the function and operation levels of the interaction analysis are employed (Chapter 5, Tables 5.1 and 5.6).

The fourth way is to perform an evaluation of the implemented machine (summative), which is done during *the implementation phase* in the development process (Number 4 in figure 6.6). This is done to assess that the developed machine is good enough regarding effectiveness and safety in interaction. In this case both the function and operation levels of the interaction analysis are employed (Chapter 5, Tables 5.1 and 5.6). Hence CCPE works as well within the framework of product development, which was previously also shown for the areas of Human Factors and Ergonomics, Human Factors Engineering and Risk Management.

6.8 Relationship with mismatches in interaction

The starting point for the purpose of developing the methodology was that a mismatch often occurred between human and artefact. How does CCPE ultimately relate to this? The first step takes place in the GTS where both task demands and workload, both physical and mental, are mapped. Here it shows if the task demand or workload is too high, which indicates that the task exceeds the user's ability. The next step takes place during the actual interaction analysis.

CPPE investigates the problem of mismatches in two ways: directly through errors and indirectly through problems. Occasions in the interaction which involve a considerable risk are those where an error can have serious consequences at the same time as the user has a low probability of acting correctly. It can therefore be said, with some simplification, that CCPE studies the risk of the various operations. CCPE both elicits the consequence of errors and elicits a judgment of the probability that the user not will act correctly.

Much of the theoretical frame of reference is related to use error, and this is also important for CCPE. The methodology is based on the manner of acting that is officially prescribed by the manufacturer of the equipment or the workplace where the equipment is found/is to be used/is used. The methods investigate deviations from this prescribed manner of action, not from any optimal way of using the equipment. The attitude toward use error that the methods embody is thus in agreement with the view of use error that the IEC has[4]. The analysis in CCPE does not take a position on "whose fault" a use error or usability problem is. The analysis shows only that there is a misfit between user and artefact. Further investigation and proposals for measures against this misfit lie in the parts of the usability process that occur when CCPE has been completed.

The theoretical section of this thesis has treated the areas of risk and safety as well as usability and physical ergonomics, and demonstrated that they are interrelated; low usability leads to increased probability of errors, which in turn increases risk and decreases safety but also can lead to poor physical ergonomics. There is also a reverse link: a risk exists with every usability problem. One can judge the probability of a user acting incorrectly and what consequence this may have. The concept of risk can thus be employed to grade the severity of detected problems in a usability evaluation, and to rank the problems relative to each other.

Consequently CCPE tries to take a holistic approach on mismatch in interaction. Not only does CCPE examine whether the correct way of use (intended by the manufacturer) results in poor physical ergonomics; CCPE examines mismatches both in the conditions for interaction and in the actual interaction itself, both physical and cognitive. What is special for CCPE is that the methodology also focuses on mismatches due to the link between cognitive and physical factors, and in this way physical and cognitive ergonomics are integrated.

---

[4] IEC 60601-1-6 defines use error as an *"act or omission of an act that has a different result than intended by the manufacturer or expected by the operator".*

7 Discussion

The chapter begins with a discussion of the thesis work in relation to important factors relating to the development of new products and technological systems. Thereafter follows a discussion of the approach for method development and research.

7.1 Deployment

CCPE methodology was developed to be useful in the development of new products and technological systems and to serve as support in design. The central questions raised are the extent to which it is possible to generalise the methodology and how well it fits into the process of product development.

7.1.1 Generalisation

The developed CCPE methodology is based on two further-refined methods, ECW and PUEA, and on the premise that they can work together. One relevant issue is then how compatible ECW and PUEA really are – i.e. whether they can be employed simultaneously on the same type of equipment with good results, or whether occasions exist when one method does not function together with the other.

This issue is relevant as ECW derives from evaluation of simple technical systems, so-called "walk-up-and-use" systems, whereas PUEA originates from evaluation of complex systems – primarily control rooms in nuclear power plants. The methods have been employed in other contexts than those for which they were originally designed, and with expected results, i.e. prediction and identification of use errors and usability problems. For instance, ECW has been used for evaluating more complex medical equipment such as interfaces on dialysis machines (Bligård and Osvalder, 2006b, Bligård et al., 2006a), and PUEA has been used to evaluate the simple task of changing the insulin ampoule in an insulin pump (Dahlén and Ullström, 2006).

More generally, Nielsen (1993, p 24) notes that *"…many usability methods apply equally well to the design of other complex systems, and even to simple ones that are not simple enough"*. Even so, the developers of CW are not so certain that the method is suitable for more complex technical systems. Lewis and Wharton (1997) write for instance that CW is designed to assess support for "walk-up-and-use" and that *"beyond this level the CW provides nothing but a crude inventory of material that users must know to operate the interface successfully."* Kitajima and Polson (1996), on the other hand, write that steps in an interaction which are hard for novices to learn are also "highly error-prone" for experts – i.e. those tasks that are hard to learn will also be the ones where errors may occur most easily, an argument that contradicts Lewis and Wharton. This indicates that it is relevant to employ ECW also for analysis of complex systems with expert users in order to investigate use errors.

CCPE methodology was for the most part developed and applied during development of medical equipment. The areas of use were then spread in different directions, including more complex areas such as nuclear power plants (e.g. Oxstrand, 2006). There the ECW analysis resulted in valuable information to the operators about feedback from the feed water pumps in the control room. The methodology has also spread to simpler consumer products such as office chairs and vacuum cleaners (Appendix C and D). The most divergent application was in extreme sports where CCPE was used to analyze kitesurfing, as presented in Lundgren et al. (2011) and in Appendix E. In Table 6.1 in Chapter 6.1 there is a comprehensive compilation of the different application areas that shows the width of implemented applications for CCPE and its included methods.

Development in the end then shifted from medical technology, where a complex machine with simple actions such as pushing buttons in the interface was evaluated, to kitesurfing, where a simple machine with complex actions in the form of motor balance exercises was evaluated. The methodology was judged successful in all evaluations as it was able to predict and explain the mismatches between human and machine.

One reasonable conclusion to draw from the evaluations conducted so far is that the methodology is useful in all areas where there is active interaction between human and machine. Having said that, the methodology needs to be adapted to each individual analysis situation to function optimally.

7.1.2 Suitable for development work

One highly relevant question to ask is whether there is a need for the methods included in the CCPE in the development of new machines, that is, a need for methods to detect mismatches between human and machine. The central methods of a product development process are evidently the methods that help design the new solution or collect the information needed to find a new solution. Methods to evaluate the mismatches between humans and machines do not contribute, in themselves, to development of the product; rather, their role is to evaluate the proposals in which the development methods have resulted. The CCPE methods are thus a form of inspection to ensure that the developed product is good enough. This was precisely the goal of the company, described in Paper V, when they used ECW and PUEA in their evaluation work.

However, what distinguishes a good method suitable for product development? Norell (1992) has developed a list of characteristics for a method to work well in a product development process (translation by Almefelt (2005, p 48)). In general the method should:

- *"Be easy to learn, understand and apply.*
- *Contain accepted, non-trivial knowledge within the areas of interest.*
- *Provide support to identify weak spots.*
- *Be rewarding to use for different disciplines, leading to the establishment of a common reference and shared views.*
- *Support co-operation and facilitate a learning effect for users.*
- *Contribute to a systematic work procedure.*
- *Have a positive and preferably a measurable effect on the outcome of the product development work within the area of interest."*

For the individual method user, the method should offer:

- "*A high degree of perceived freedom of action*
- *Possibilities for a holistic view, solidarity and mutual understanding.*
- *Possibilities for personal growth and learning through work*"

Another important characteristic for a method, which is highlighted by Rexfelt et al. (2011) is that the method should be fun to use. If the method is boring during use, it is less likely that someone will use it. For methods within the area of ergonomics specifically, Stanton (2002) has listed six challenges faced by developers of ergonomic methods:

- Developing methods that integrate with other methods
- Linking methods with ergonomic theory
- Making methods easy to use
- Provide evidence of reliability and validity
- Showing that the methods lead to cost-effective interventions
- Encouraging ethical application of methods

Regarding why methods, especially new methods, are not employed in the product development process, Araujo (2001) has summarized the reasons:

- Lack of reason and interest by the organization in using methods
- Lack of understanding of how methods can be useful
- Lack of "appeal"
- Lack of resources such as time, staff and competence
- Defects in the design of methods
- Poor promotion practices
- Fear of changes
- Too many alternative methods to choose from
- Negative attitude towards introducing new methods

This shows that there are causes within the method itself and causes linked to the users and the organization that explain why a method is used or not. One prominent factor is that methods need to be simple (and preferably fun) to learn and to use. As shown in chapter 6 CCPE was experienced as long-winded and overly demanding to learn and to perform, due to the systematic procedure on which the methodology builds. The fact that the methodology is perceived in this way entails a need to further develop the methodology so that it becomes easier to learn and use, and thus be more appealing. However, this future modification must aim to not simplify the analysis. Chapter 7.3.2 discusses how the methodology can be developed further.

Another aspect concerns the fact that the methods should be relevant to use and give a reliable and useful result in a product development process. CCPE has been used in several development projects in industry and academia and in several master thesis projects at the university, as reported in Chapter 5.3. In these studies CCPE has led to results that were useful in the development processes.

CCPE has a clear, systematic way of working, which is one quality stressed by Norell (1992). Norell further points out the importance of a method to support collaboration between different disciplines, create a common picture and understanding, and support learning and personal development. More specifically, knowledge transfer and the creation of consensus is an important aspect of CCPE and this is dealt with more explicitly in Chapter 7.2, together with the need for knowledge. What is of interest to touch on here, however, is that one benefit of CCPE during the development process is to communicate usability more tangibly, through a systematic method of working, to other engineering disciplines relevant to the process. This was an issue raised during the interviews at the company described in Paper V. The respondents who used ECW and PUEA stated:

- *"...good to prove to others that there are problems and where they are (e.g. project leaders, others who like numbers)..."*
- *"...it is good to have something tangible to point to in this type of engineer-dominant company, usability is otherwise usually a bit fuzzy..."*
- *"...documentation of what is otherwise done spontaneously, good to have for knowledge transfer and backtracking..."*
- *"...it is an advantage that it fits into the engineering field, it is easy to convince people, quantifiable..."*

Focusing on the specific challenges facing methods of ergonomics, the development of CCPE has meant improved integration between methods of ergonomics and stronger linkage with ergonomic theory (than the original methods). However, the methodology is not perceived as easy to use. This thesis has presented preliminary indications of reliability and validity of CCPE methodology, but this requires more studies for proper definition. In addition, more work needs to be done to judge the cost-effectiveness of CCPE's application for the development process. The ethical aspects of method implementation have not been addressed in the development of CCPE. The method development tackled some of these challenges well, others less well.

As primarily Araujo (2001) points out, there are additional reasons for a method not to be used, reasons that are related to the organization and/or the individual, and not the method itself. A method is often considered as an isolated phenomenon in its use, and not as part of a whole in a development process. The development process consists largely of methods and other activities closely related to each other. Chapter 6.6 shows what other methods are closely linked to CCPE. It is important, in order for the method to work well, that previous steps have been performed in the development process and that they can provide the method with the information required. For CCPE to be effective in work with development, it is necessary for the development process to have a suitable level of HF work. Otherwise extensive preparatory steps will be needed for CCPE before the analysis can be performed. If task analysis and user profiling are undertaken during development this will favour a CCPE evaluation. Furthermore, it is important that the CCPE results can easily be implemented in the ongoing development process. The company dealt with in Paper V used the results from CCPE as part of its overall risk management and thus had a suitable way of receiving the results, where they were useful. If there is no preparation for these prior and subsequent steps when a method is to be used, it easily reduces its effectiveness. A further discussion of the methods required for a working process to be able to function can be found in Chapter 7.2.

Based on the discussion in this chapter, 7.1, the theoretical reflection in Chapter 6.3 of CCPE's starting points linked to the development process as well as the detailed description in Chapter 6.7 which shows when and how the methodology can be used in the development process, it is reasonable to conclude that CCPE is a methodology that is suitable for use in a product development process and provides good support for design work.

7.2 The users

Apart from the process and the methods, which have been discussed in Chapters 6 and 7.1, staffing is important for effective work (cf. Kaulio et al., 1999). Furthermore, as shown in Chapter 7.1 Norell (1992) and Araujo (2001) have presented a number of aspects related to the users of methods. It is especially important to discuss the following aspects:

- Knowledge
- Trust and acceptance
- Consensus and knowledge transfer

7.2.1 Knowledge

Knowledge of the method is a central feature for all use of methods. For example was 'Lack of knowledge' identified as by far the biggest reason why the methods are not used, in a web survey on use of ergonomic risk assessment methods (Mattes et al., 2012). Other prominent reasons were 'Too time consuming' and 'Too complicated to use'. One of the requirements of the developed CCPE methodology was that the expected users are engineers with knowledge of Human Factors Engineering. When the original CW method was developed, one of the goals was to make insights from cognition research available for designers and developers without their needing to have in-depth knowledge of this subject. In the case of CW, it was knowledge about explorative learning that was made more available. Some studies show that novice users of the method, after a short introduction, could use CW and thus profit from this knowledge through the method (Green et al., 2000). Nonetheless, much criticism has been directed at the fact that CW, especially in its earlier versions, requires knowledge of cognition in order to function well. For instance, Wharton et al. (1992) write: *"it will be difficult to eliminate the need for a cognitive science background both to make sense and to take full advantage of the technique"*.

Changes made from CW to ECW do not decrease this need for knowledge in cognitive psychology; it is not influenced, and perhaps is even increased, by the categorisation of problem types. In regard to PUEA, the need for cognitive knowledge has grown since the categorisation of incorrect actions was introduced. For example, users of the method require knowledge of the Skill-Rule-Knowledge model (Rasmussen, 1983) and the Generic Error Modelling System (Reason, 1990) in order to employ PUEA fully. By also including aspects of physical ergonomics in the evaluation the method will demand also this type of knowledge of the evaluators. Thus, CCPE places higher demands on its users than do the original methods. This renders the methodology more difficult to apply, but the necessary knowledge in the field of cognitive and physical ergonomics is of benefit in the usability engineering process at large.

A method can never replace knowledge, and knowledge is required especially to counteract a focus on details as is the case in CCPE. Nevertheless, knowledge about cognition and human error is needed primarily in order to design the artefacts, so that the entire development process is not based on "trial and error". CCPE cannot contribute to this by replacing knowledge and experience of the area. The CCPE analysis shows what is bad (problems and errors) and what the causes are, not directly how to make it good. Hence, practical experience and theoretical knowledge about the design of user interfaces, or knowledge about human information process is needed throughout the usability engineering process – which means that it is unnecessary for CCPE to be designed for users who lack such knowledge.

Thus, the expected users of CCPE are engineers with fundamental knowledge of cognition, cognitive ergonomics and interface design. However, nothing indicates that the methods cannot be employed by other professional groups working with interaction between humans and technology, such as industrial designers, psychologists and ergonomists.

It is rarely possible to find a single person who possesses all the knowledge needed to perform CCPE optimally. It is therefore a better solution, as previously described in Chapter 5.1.7, for CCPE to be performed by a group of evaluators, as knowledge about the methodology is needed, i.e. both cognitive and physical ergonomics, as well as knowledge of the machine, users and usage. The author of this thesis has successfully served as moderator of CCPE evaluations without knowledge of the machine, the user or use. In these evaluations, other people in the expert group possessed this knowledge, such as real users. An example of such an expert group is in the analysis of kitesurfing (Paper Paper XVI and Appendix E). The inclusion of users as part of the evaluation team is very valuable, as they can also be seen as users of the methodology.

7.2.2 Trust and acceptance

For the method to be useful, the users must accept the method and have confidence in it. Results from the study in Paper V showed that the users involved could accept CCPE methodology and that they had confidence in it. Nevertheless, it is interesting to discuss various factors of CCPE and its use that influence trust and acceptance.

One factor affecting acceptance and trust is the relationship between costs and benefits. One of the major disadvantages of CCPE (reported in 6.4) is that the methodology both is, and is perceived as, time- and resource-demanding due to its systematic structure. However, the hope is that the extensive results that the methodology entails are worth the effort needed to perform CCPE. Results from the study in Paper V point in this direction as CCPE is perceived as a useful methodology where the quality and the quantity of the outcome are seen as valuable.

The usefulness of a result from a method also affects trust and acceptance. As described in Chapter 7.1 a result obtained from a method must be identified and taken into consideration in the development process. If a user observes that a method generates a lot of results that cannot be used, this can lead to reduced trust in the method. An example from the study in Paper V shows that professional users of the methodology from the company saw direct benefits in the long-term perspective with the advanced search for mismatches between human and machine, since it brings about safer development of medical devices. The same does not apply when students use the methods for training purposes in order to evaluate a product. Their generated results will rarely benefit an ongoing product development process. However, the systematic process and the structure that define the quality of the results have a downside in that the study in Paper V showed that users felt the methodology was exhausting and that it was difficult to maintain the level of concentration needed to achieve good levels of performance.

One aspect that clearly affects trust and acceptance of the method is the relationship to the result. What is especially interesting for CCPE methodology is when there is a "zero result", i.e. when few serious mismatches are detected during analysis. For a developing company a "zero result" is the desirable result, showing that the product has been developed with regard to usability problems and use errors. Finding few errors and problems may increase acceptance of, and confidence in, the method, as it can also show that something is good. For students, however, a "zero result" can be less good since they may wish to find mismatches that can be presented in the project report and find problems and errors as a basis for improvement in the redesign. The interview study in Paper V indicated that students may become suspicious if the methods really work or wonder if they made a mistake when they performed their analyses. There was also feedback to the effect that a "zero-result" made the methods even more boring to perform since they do not provide any interesting output, which means that it is easier to lose willpower and concentration during execution. The result therefore is that acceptance of and confidence in the methodology can be reduced.

Something that also affects users' relationship to a method is recognition and familiarity. It is easier to gain acceptance of and trust in something you are familiar with if it works well. A clear indication of this is the study in Paper V, where the majority of the students indicate that the methodology was easy to learn, which is probably due to their habit of using these kinds of methods. During their education they had come into contact with the original methods, Cognitive Walkthrough (CW) and Predictive Human Error Analysis (PHEA), from which ECW and PUEA evolved. The same was true for the company in Paper V when professional evaluators used PUEA. This method has considerable similarity with the method for technical risk analysis, FMEA (Failure Mode and Effect Analysis) (Taylor, 1994) that they used before. It was basically just a matter of saying that PUEA was similar to FMEA, but for evaluation of human error rather than technical errors. Then they understood what the method was about. Both PUEA and FMEA are types of tabular data analyses, and basically work in the same way.

CCPE is designed to be a flexible methodology that can be adapted as needed. However, this also creates a problem as it becomes a method that cannot be used "off the shelf" but instead requires adaptation, which takes time and requires knowledge and experience. The fact that initially the methodology is difficult to use can reduce its acceptance, since it becomes more difficult to fit into a stressed development where there is little time for adjustment, education and training. Flexibility is therefore a double-edged sword in this case.

When it comes to CCPE there is another aspect that affects acceptance and trust. This is the relationship between the formal and the non-formal, whereby the methodology can create an illusion of objectivity. The methodology has a systematic, structured and complex framework that can be interpreted as though the analysis is independent of the person or persons conducting the analysis. However, at the analysis stage itself it is still a subjective assessment that forms the basis for the errors and problems listed. A possible reaction may thus be 'Why do all this groundwork for ultimately just guessing?'. This attitude can reduce confidence in the methods as you do not see the benefit of following a structure for your thought process, especially when empathy (being able to think like an actual user) and imagination are essential skills for a good analysis.

Another effect that can arise is that the performer of the analyses can drown in the process, i.e. applying the methods according to the template and following the methodology structure becomes more important than performing the analysis itself in an effective manner. This may reduce confidence in the methodology. Examples of this are described in the study in Paper V where one of the respondents said regarding ECW / PUEA: *"... time-consuming, much of the same thing, the same answer to many questions, especially if it is a good product ..."*. As CCPE has the extensive structure that it has, there is a danger that it becomes difficult to absorb. The point that CCPE is designed to be adapted for every occasion is thus missed. One possibility not explored is to create guides and other support to lead the user into choosing an appropriate level of analysis. This is a possible future project for developing the methodology further.

An interesting reflection on changes in trust and acceptance was made by the author in the analysis of the emergency patient stretcher described in Chapter 5.3.3. The author was the moderator and in the interaction analysis there was a user of the existing product as well as an educator on the product. When the CCPE analysis began, these people were initially very sceptical about the procedure of actively looking for problems in a product that they used and thought worked well. However, after they realized that the problems and errors detected often were relatively simple to remedy in a redesign, they became more sympathetic towards the methodology. After the analysis, they commented that they never imagined there were so many opportunities for improvement that they had never thought of before, for a product with which they came into contact regularly. The CCPE methodology and its process had led them to think in completely new ways. If they have done the analysis themselves, without the author as moderator, they would very likely have got stuck in the structure and the systematic approach of the methodology.

7.2.3 Knowledge transfer and consensus

For a successful product development project, knowledge processing, communication and collaboration are essential building blocks (Sun et al., 2010, Büyükzkan and Arsenyan, 2012, Westling, 2002, Persson, 2005, Cross and Clayburn Cross, 1995), which makes it interesting to discuss if the use of methods as CCPE contributes to this. It can be argued that methods like CCPE contribute through knowledge transfer and created consensus within the project group as the method is performed in groups.

During method use there is an exchange of knowledge among users. A dialogue is generated in which the evaluation team together learns more about users, usage and the machine. This leads to the creation of a common picture that evaluators can use as a starting point for their continuous work in product development. If end users participate in use of the method, this means that they convey their knowledge to the technical developers, while developers in their turn can pass on knowledge about the potential and limitations of the technology. This leads to a learning process for everyone involved in the evaluation, something which Norell (1992) emphasized as an important characteristic of a good method. Norell also highlighted the importance of the establishment of a common reference and a shared point of view. This means that if the evaluation team reaches consensus on what the detected problems and errors are and why they exist, then that is an important part of the method's content. Consensus makes it easier to repair any detected mismatches in design work. If different images exist among evaluators, this complicates the design process.

The results of the study in Paper V show that both professionals and students believe that CCPE is a good basis for group discussions and that it is good at creating consensus among its executors. Getting the evaluation team involved is something that Wixon (2003) argues in favour of when he claims that the method encourages participation. Buy-in and collaboration are crucial for a usability method to work in a development process. CCPE not only provides information about the mismatches in the interaction; the actual process of using the methodology also helps in the development process through the creation of knowledge transfer and consensus.

This coincides well with the author's own experiences of using CCPE in development work in the field of medical technology, primarily described in Bligård (2003, 2004). Using HTA and

PUEA resulted in extensive knowledge dissemination within the evaluation team, both about use and user, as well as the machine. One can say that execution of the methods forced the project to clarify and define the actual use of the product, which resulted in everyone having the same view. It is the author's belief that the benefits in terms of knowledge transfer and creation of consensus in the group was at least as significant as the utility of the identified mismatches, seen from the development project in its entirety. The effect of the method is not only the protocol reporting the results but also the development of knowledge and understanding among performers that arises from use of the methodology.

7.3 Usefulness

Even if a method can be generalised and fits into the development process, and the users have knowledge of and confidence in the method, it must also lead to a useful result. Chapter 7.3 discusses how and why methods like CCPE work and how to go ensure that they work even better.

7.3.1 How the methods work

One central criticism that is often presented by users regarding methods that are not based on empirical studies is that they lack a connection to reality and that they therefore do not provide realistic results. Criticisms of methods such as Cognitive Walkthrough and Heuristic Evaluation maintain that the reliability of the method results can be questioned since the results are based on a subjective judgment of the evaluator (Stanton et al., 2005). Petrie and Power (2003) clarify this criticism:

- low overlap between usability problems proposed by expert evaluations and user evaluation;
- different experts or groups of experts produce different problem sets;
- expert evaluations over-emphasize low-severity problems at the expense of high-severity problems

However, when using the methods in the development process, it is important to consider whether a particular method gives any value compared to not using any method at all, or using a different method.

The comparison in Table 6.8 in Chapter 6.5 showed that there were benefits to using expert methods compared to not using any method at all, while Table 6.7 shows that different method approaches cover different aspects of the evaluation of interaction. Furthermore, Chapter 6.6 shows how the methods can be combined to provide a more comprehensive interaction analysis. For example, in a study where Heuristic Evaluation was compared to Usability Testing, de Cook et al. (2001) showed that Heuristic Evaluation gave information about why and when, while usability testing provided information on what and how. Furthermore, in terms of resources it is not reasonable to use usability testing to examine all possible user errors that may occur during interaction with a machine. What is crucial is to be able to assess the use of each method, to get the best analysis results and not to exclude a group of methods if there is no better alternative.

It is also interesting that within the more classic engineering sphere there are a number of methods based on subjective assessments for analysis of technical systems (Stricoff, 1996). For example, classical engineering often uses methods such as Hazard And Operability Study (HAZOP) and Failure Mode and Effect Analysis (FMEA) (Taylor, 1994) which work in the same way as ECW and PUEA, guide the user in the evaluation through a structure and questions. The advantages of HAZOP and FMEA are that it can be difficult, costly and extensive to test how technical products/systems can fail in reality. Moreover, in some cases it is dangerous and even impossible to carry out real experiments, for example during risk analysis in nuclear plants.

Specifically for the thesis, Chapter 6.5 demonstrated the pros and cons of different approaches for methods and examined what it is that makes theoretical methods, such as CW, PHEA and CCPE, actually work, something that the summary in Chapter 5.2 indicates that they actually do. Although the methods provide a systematic structure to search for problems, it is still the evaluators (the humans) whose own assessments deliver the results, i.e. assessments which by nature are subjective. Thus, it is the human who uses the method that makes it useful, as is the case with most methods.

However, if it is human participation that is important in the method, why not simply ask users and experts about potential use errors and usability problems? This is evidently a possible solution, but the systematic structure of the method gives added value that allows more potential problems to be detected. In Chapter 6.2 the comparison between the different methods is also discussed. But why do the logic and structure of a method in this case cause more use errors and usability problems to be detected? Possible explanations can be found when studying human thinking in method use. Thinking is described as *"... a process by which a new mental representation is formed through the transformation of information by complex interaction of the mental attributes of judging, abstracting, reasoning, imagining, and problem solving" (Solso, 2001, p 417).* What is particularly interesting here is the actual problem-solving.

The structure of a method implies that humans think in new ways compared with free thinking without structure. We know that people can get new ideas through assistance to associations (Solso, 2001). A structure and a systematic approach also help the user understand the problem better, in this case to detect and identify presumptive use errors and usability problems by focusing the user's actions and providing guidance in thinking. Understanding the problem itself is the basis of problem-solving (Maltin, 2009, Reisberg, 2001).

One thing that is important to avoid is for people get stuck with a fixed idea during problem-solving, that is to say being left in old thought processes and unable to think differently. This is termed functional fixedness (Solso, 2001, Eysenck, 2012, Groome and Brace, 2006). The structure and logic of the method can force the user to think in new ways by asking questions related to all aspects of the interaction, so that the user does not only focus on certain parts. The method thus helps users to think in a new mental set, where the mental set is *"a readiness to think or act in a given way ..."* (Eysenck, 2012, p 314).

The structure and logic of a method may also facilitate in that it represents the problem-solving strategy that the user needs in order to perform the analysis optimally. The way a problem is represented plays a big role in how easy it is for a human to solve it (Solso, 2001, Groome and Brace, 2006, Maltin, 2009). Here the method helps steer the analysis in such a way that user knowledge will be used better, compared to not having any structure and systematic approach.

One aspect linked to problem-solving is that the method also provides the user with the right amount of information, since both too little and too much information would make problem-solving less effective (Solso, 2001). A method's structure and systematic way helps the user focus on the right things by removing unnecessary information. At the same time it clearly spotlights what information needs to be considered. In the case without a method, the user gets no support regarding what information to consider.

The above reasoning is supported by statements from persons in the interview study described in Paper V. There the respondents regarded ECW/PUEA as follows:

- *"... a structured method for considering one situation at a time..."*
- *"... really must go deep and think through all situations..."*

- *"... give an objective view, need to analyse all steps of the task that has become obvious to oneself..."*
- *"... a very methodical way, have to consider every part of its use..."*
- *"... extensive method where you consider everything, even if not assessing everything to do with the method.."*
- *"... being forced to think about everything that can go wrong ..."*
- *"... good structured way to think through the product, makes one take it seriously..."*
- *"... the systematic method of assessing the problem, a common judgement, makes it easier to be critical toward something you have developed yourself..."*

To summarise, it can be said that although the structure and scheme of the CCPE methodology is perceived as a barrier by some users (see for instance Paper V), this is what gives CCPE its value. However, as pointed out many times previously, the users of CCPE have to have knowledge of the users of the machine and the use, to deliver a good analysis results. CCPE functions as a catalyst in the method users' cognitive process.

7.3.2 How to get the methodology to work

The reports in Chapter 6 and Chapter 7 have this far touched on many aspects that show how the developed methodology fits into a development process and various weaknesses in the development methodology, and also discussed problems that methods in product development generally face. One key issue in this context is whether anything other than the methodology itself can be altered to counteract its weaknesses and get the methodology to work better. Table 6.8 in Chapter 6 shows a link whereby the better the analysis of the method is, the more time the method demands and the more time it takes to learn the methodology. Araujo (2001) pointed out that there are factors in the organization and the individuals that obstruct use of the method. Methods generally require good input to function and output must be taken care of for the method to be effective. This is a clear indication that external conditions are also necessary for the method to work.

The PRE-process (Kaulio et al., 1999) identifies methods, processes and staffing as three central elements for successful development work. It can therefore be argued that many of the weaknesses perceived of CCPE probably would not be as prominent with ergonomics in mind in the processes and trained users. This has also been discussed in Chapters 7.1 and 7.2. Changes that would be likely to occur with trained users and matching product development processes are the following:

- The methodology could easily be adapted to suit the intended purpose, so it will not be unnecessarily complex
- The moderator for the interaction analysis would adapt the questions so that the analysis is not perceived as tedious and boring
- The information in the system description phase already exists so no extra time needs to be added before starting the workload analysis or interaction analysis
- All results produced by CCPE can be utilized in some part of the development process so that no step in the methodology is perceived as unnecessary

As previously discussed, it is difficult to make changes in the actual execution of CCPE to make it easier and quicker to perform and learn, since it is the structure, the systematic way and flexibility that create the usefulness of the methodology. However, even if the CCPE methodology itself does not have to change, there are other possibilities. Three areas where significant benefits can be obtained by making CCPE easier to learn and perform are:

- Create a computerized version that reduces the need for manual administration. For instance, creating lists and matrices directly from filled out templates.
- Create practice material that helps users grasp the methodology in steps small enough to not be overwhelmed by its complexity and terminology, and provide support in how the methodology will be adapted to different evaluation cases.
- Create a selling description of CCPE which clearly highlights all the benefits of the methodology, including factors such as consensus and knowledge transfer, and clearly describe the usefulness of the results provided by the methodology provides. Also, clarify what is required of the development process and methodology users for CCPE to work optimally.

If these three points are met CCPE has a larger possibility than today to become a more widely used methodology during product development. However, much of the success of a method such as CCPE depends on factors that are beyond the methodology, such as the structure of the development process and the user skills. It can be argued that methods should be adapted to the processes and the users that already exist, which is true in many cases. However, if methods with better performance are desired, users and processes must also be challenged so they become better as a whole. Otherwise there is a risk that there will be a lock-in within method development and sub-optimization towards existing users and processes. Methods, processes and users (staffing) together generate the usefulness of product development (Kaulio et al., 1999). These therefore need to be considered in a system where all parts must be optimized together. A method is never better than the process in which it is used, and never better than the knowledge held by those using it.

7.4 Method development and research approach

Chapter 6 and Chapter 7 have thus far discussed the developed methodology. This section discusses the actual methodology development and the approach to research. CCPE is a structured and systematic methodology, but the work to develop the methodology was not structured and systematic. As described previously, the methodology emerged from a number of implementation projects in industry and academia where the goal was to evaluate machines. In these projects there were insufficient resources to implement more systematic method development and evaluation of the included methods. Although the developed methodology achieved its goals, there are weaknesses in the way the method development was performed.

The first weakness is that there has been no analysis of whether the original methods, CW and PHEA from which the methodology originates, actually are the best methods from which to initiate development work. There may well be other methods that are better in their basic design and thus may have resulted in a development that delivered further progress. Furthermore, there was no structured and systematic evaluation of these methods to state what really are their pros and cons in a process. Another weakness is that the actual development work of the new methods was not systematic but rather "trial and error", where various ideas were developed and then tested. Various proposals for change to improve the original CW and PHEA methods were not set against each other to decide which one was best, and so on.

Another weakness in the method development is that there was no detailed documentation of how the whole process actually came about – for instance, how different proposals for change were selected, how they were evaluated, or why they were kept or rejected. After the diverse supplements to the methods were made, however, there was a recapitulation of what did or did work. As described earlier, the reason for this procedure is that the methodology development was a spin-off from a number of projects with other aims than that of developing methods. This makes it hard to reflect on the development carried out and to draw conclusions – about how method development is conducted, and what functioned poorly or much better – that are more specific than the general terms of the discussion.

The biggest drawback of the development work is that there was no evaluation of whether the usefulness of the methodology was improved by the achieved development or whether the usefulness was already embedded in the methods on which the methodology is based. In other words, no systematic validation has been undertaken of ECW as being better than CW at detecting and identifying usability problems, or of PUEA as being better than AEA, SHERPA and PHEA at detecting and identifying use errors. Nor has the methodology been tested to find out if different evaluators come to the same end result or if evaluators performing the methodology two different times would generate the same results. Reliability is therefore not explored.

So what then is the strength of the method development that makes it worthy of a thesis? There are of course a number of strengths. The first is that the method development was based on the problems that existed in real projects. CCPE thus solves real problems. The development of ECW and PUEA occurred during real projects in industry that directly demonstrated that the methods do work. CCPE is thus not a desktop product developed away from the context in which it will be used, instead CCPE is a direct result of the contextual conditions. The same applies to the author of the thesis that developed the methodology. The methodology grew out of the needs that the author experienced as a user of existing methods, and he personally experienced that the different methods work. The methodology presented in the thesis is thus not a possible proposal, but something that actually has been tested and used by both the author and several other evaluators – research colleagues, engineers in companies and students.

The full methodology and parts of it have thereby been used in many evaluations. The methodology has also been used to study human-machine interaction in several different types of machines. Starting with medical equipment and its user interface to kitesurfing equipment and seat belts that require a mixture of physical and cognitive interaction. The fact that the methodology has been successfully used by many people in various applications is a strong indication that the method development was successful, although there are a number of weaknesses in the actual development work.

Another strength of the method development is that it is founded on a deep theoretical basis, in terms of method, theory and theory of interaction between human and machine. This means that the developed methodology does not float freely, but instead builds on existing knowledge and methods which have made method development easier and simpler to perform. Method development has also taken place within the thought of application area (Product Development Human Factors Engineering and Risk Management), and with this area in mind so that the developed method fits into the framework.

The research approach used in the method development was successful in the sense of having led to a result in terms the methodology. The approach, similar to action research, where the method developer himself/herself employs the methods undergoing development means that he/she personally experiences and discovers the respective strengths and weaknesses of the methods. This yields first-hand information instead of second-hand experience, for instance by interviewing others who have used the method. The immediate interaction with the method also enables the developer to directly test proposals for changing methods and to evaluate the proposals.

Another advantage of this way of working is that it helps bridge the gap between theory and practice – in this case, the gap between research and industrial applications. Since the method developers personally employ the methods in actual industrial projects, knowledge can be disseminated to the industry at the same time as experience from the industry can be distributed to research. If several individuals apart from the method developers are involved in an industrial project, this also means that the method know-how is spread within the organisation.

A further benefit of applying the methods in real development projects is that it becomes possible to discover pros and cons from a developmental perspective. The methods and the methodology are intended to be employed in industry, so it is essential that the method development be based on this, not on advantages and drawbacks, which would arise if the methods were employed as research tools. Application in the right environment allows the methodology to be adapted to this environment, with the problems that may arise there. The methods' usability is important and to ensure it, the methods must be evaluated in their intended environment with the intended users, not in a research environment with researchers as users. Further, applying the methods in their correct environment implies that the intended users get to try the methods and gain trust in them.

However, one weakness with this approach, and consequently with the methodology development, is that the central source of information was a single person, the author. Having said that, the development occurred through an exchange of ideas with colleagues who worked in parallel. Nonetheless, the methodology was also used by other people than the author, contributing input information from several people to the method development.

The fact that the research was problem-driven has had a major impact on the development work as outlined in Chapter 3.1. That the problem, i.e. the shortcomings in existing methods has directed the research, has meant that research has not directly been driven by theory or method, but the focus has been to find a solution instead of developing new knowledge. The advantage of this approach has been that a relevant problem has been solved by developing an improved methodology. The disadvantage has been that there is no template to work with, i.e. no given theory or method to relate to, which made the development workless systematic and structured. For example, there were no predetermined research questions that could lead the work.

The research work was problem-driven and occurred in different projects, which had a purpose other than to develop the methodology. This meant that it was more difficult to evaluate the reliability and validity of the methodology. It was rarely possible to compare the developed methodology with the original methods as the projects have not contained resources for using new methods that may not analyze what the project asked for. Due particularly to lack of time and resources it has seldom been possible for other individuals to use the methodology in parallel with the developer so as to compare the results of the evaluation between two users. Furthermore, it was difficult to compare the results of the methodology with empirical data, the results of the methodology were used to improve the machine before any empirical attempts were made. Therefore, the evaluations presented in Appendices A-E were conducted so that a comparison between the analytical evaluations of CCPE could be undertaken with usability tests to see if the same usability problems and use errors were found.

A more classical approach of method research could be to begin by studying the literature on existing approaches, and then conduct separate studies of these methods to examine how they have to change, and finally suggest what improved methods should look like. If the CCPE methodology had already existed, and the goal of the research had been to evaluate it to determine its qualities, it would also have involved a different approach than the one described in this thesis, with more comparative testing of the methodology. Although research at this detailed level differs from more traditional research approaches based on the overall theme of the hypothetic-deductive approach, which is a common way of conducting research, this thesis can be seen as a different and hopefully a complementary way to conduct research and generate new knowledge.

The effect of the scientific knowledge built up from the research approach is that the new knowledge developed is not within a specific range, but is instead in many different areas. In other words, there is more width than depth to the scientific knowledge built. It can be seen that the work here has not moved up top on the research front, but that it has instead woven together loose parts of the front and integrated these elements into a whole. This is seen clearly in the quantity of theories that form the basis for the methodology development, which is reported in Chapter 4, and by the methodology linking the different theories on which it is based, for example physical and cognitive ergonomics. I regard this as a strength in the construction of knowledge. Research often focuses on creating new narrowly focused knowledge at the front and on top of the research. If this alone occurred the knowledge would eventually consist only of small islands of deep knowledge that had no common connection. I consider it important to develop a desirable web of integrated knowledge that creates a whole and gives meaning. This was done in the development of CCPE methodology.

7.5 Fulfilment of purpose and goal

The purpose of this thesis work was to provide improved support for the developers in handling and preventing mismatches between user and artefact early in the product development process. The goal of the work was to use existing methods to develop an improved Human Factors Engineering methodology for predicting, identifying and presenting presumptive mismatches in the interaction between user and artefact.

The work resulted in a methodology with four refined methods, Generic Task Specification (GTS), Enhanced Cognitive Walkthrough (ECW), Predictive Use Error Analysis (PUEA) and Predictive Ergonomic Error Analysis (PEEA). This methodology was named Combined Cognitive and Physical Evaluation (CCPE). The main starting points were to develop a methodology to try and detect as many potential mismatches as possible, to increase the likelihood that no mismatch that can have serious consequences goes unnoticed, even if it is not frequent or probable. CCPE was developed based on analysis of medical technology, where security is of paramount importance, and for it to then be extended to other domains. CCPE also helps create the basis for consensus and knowledge when it is performed in groups.

The thesis has presented the methodology and the included methods in detail and the way they have evolved over a long period of continuous further development, through the use of a number of projects both in industry and academia. Along the way, through this long method development process, a need for further in-depth evaluation of the interaction was realized in order to get a complete picture. This means that CCPE methodology ultimately provides analysis of both physical and cognitive aspects of the interaction, which must be considered together to gain an overall perspective. The thesis has also pointed out the theoretical basis of the methodology and the way the methodology has been applied in a number of projects, where the individual methods contributed by investigating mismatches in the interaction between humans and machines. The strength of the presented work is that the methodology has strong connections to both practical application and theoretical depth. On this basis, a reasonable assessment should be that the purpose and goal of the thesis have been fulfilled.

To sum up the performed methodology development, CPPE is an efficient, but not yet sufficient, tool for investigating mismatch between human and artefact. However, machines (and products and technical systems) with few usability problems and use errors found in an evaluation with CCPE cannot automatically be said to have high usability. As explained previously, there are many aspects of usability besides a low probability of errors and problems. Rather, CPPE must be seen as a verification that the products will not cause use errors. Other methods are also needed that work both in series and in parallel with CCPE, since the latter does not cover all aspects of usability. Furthermore, the quality of the results from the analysis with CCPE also depends on the methods for providing the input data.

8 Conclusion

The chapter describes the general conclusions and possible future work.

8.1 General conclusions

The goal of the work was to develop an improved Human Factors Engineering method for predicting, identifying and presenting presumed mismatches in the interaction between user and artefact, based on existing methods. This goal has been fulfilled with the development of the CCPE methodology. The CCPE methodology proved to be useful in that it was successfully used in projects in different domains. The main idea behind the CCPE methodology is to offer a work process for detecting as many potential mismatches as possible in order not to miss any serious problems and errors.

The CCPE methodology detects and identifies mismatch in interaction through a structured and systematic process that supports executors' cognitive processes by leading and broadening the thinking. One strength of the CCPE methodology is that it can detect mismatch in the interaction between human and machine without requiring empirical tests with users. This simplifies the development process because after a CCPE analysis has been done, it is possible to sort out bad design solutions before implementing empirical tests with real users. The usability tests can thus be performed with an improved prototype or machine. The idea of CCPE is thus not to replace usability testing but instead to contribute to usability testing being used more effectively in the development process.

One further strength of the CCPE is integrated analysis of mismatch in interaction. CCPE searches both for interaction errors and interaction problems and integrates the Usability Evaluation and Human Reliability Assessment areas. Consideration was given to both physical and cognitive aspects, and physical and cognitive ergonomics were also integrated. CCPE analysis thus becomes coherent and demonstrates a holistic perspective because it spotlights many aspects that work together during interaction and that often influence each other.

The CCPE methodology and its constituent practices are perceived by users as a useful tool that provides relevant results. However, CCPE methodology is also perceived as complex and time-consuming as it contains many elements and concepts. CCPE methodology is therefore not a "quick and dirty" method, but requires training and basic knowledge in cognitive and physical ergonomics to be efficient to use. This disadvantage is compensated by the fact that CCPE generates a comprehensive result (both in content and scope). The methodology was further developed primarily to be used in development processes by engineers with training in human factors engineering, but other professionals such as industrial designers, psychologists and ergonomists can also make use of the method in their respective areas. CCPE also shows an additional significant advantage in that it is a methodology that can build consensus and provide knowledge exchange between participants during execution.

The CCPE methodology is based on and relates to theory in the area of methodology. The theoretical examination has shown that there are clear links between risk / safety, interaction problems (physical ergonomics and usability) and interaction error (use error and ergonomic error) and that these are dependent on one other. Furthermore, the theoretical examination highlighted that the management of usability problems and use error has much in common with the management of risk and that there is a probable link between risk management, management of usability problems and use error. In order to create safer machines, potential use error and ergonomic errors must be analyzed and corrected.

The developed methodology and its methods are intended to be used during the product development process. In this process, there are often other human factors engineering methods that work both serial and parallel with CCPE, such as usability testing. CCPE can be seen as a good but not sufficient tool for examining interaction problems and interaction errors in the product development process. CCPE needs to be complemented in the process in order to cover all areas of ergonomics (both physical and cognitive), with other methods and theory which not only focus on the possibilities of making errors in the user interaction. CCPE thus functions better if there is a product development process with a user centred focus from the beginning.

The basis for CCPE to become a method that works and focuses on the whole system is due to has been that the methodology was developed iteratively, it was used and modified several times in the industrial, research and student projects. As a result, its structure and content can be said to be thoroughly tested. Furthermore, the development of CCPE methodology shows that a research approach based on action research may be an appropriate approach to use in method development. Without a method developer personally using the methods and experiencing its advantages and disadvantages, it is difficult to create good and useful methods. Without this experience, it is likely that the proposals will be only desktop proposals. After the developer himself has tested the methods, other people used the methods too. The developer participated in sessions to gain information on how the methods worked, using this information to iteratively improve them in several stages.

The conclusion is that CCPE methodology has much to contribute to the evaluation process of design solutions and prototypes in product development's various phases in the product development process, so that the machine ultimately improves from a user perspective and that fewer problems and errors arise in interaction when the product is on the market. The benefits of technology will thus benefit humans better. The main reason for this is that CCPE evaluates cognitive ergonomics and physical ergonomics in an integrated way. This makes it easier to capture the whole picture of the interaction and prevent a split between mind and body. The human being is a whole unit in which both body and mind work together. This needs to be taken into consideration during product development to avoid mismatch between human and artefact. Studying physical ergonomics and cognitive ergonomics separated from one another makes the connection more difficult to handle.

8.2 Future work

Continuous work with CCPE is possible within several areas:

- A promotional text about CCPE's use and advantages
- Better description of the methodology and how it can be adapted
- Making of computer programs for tables, matrices and so on
- Inclusion of body posture and body load
- Extended user testing
- Validate CCPE against other evaluation methodologies
- Testing of CCPE during the development process
- Collaboration with other HFE methods
- Increase the ability to generalise

The first three areas above show possible interventions to make CCPE more suited for a product development process. A description promoting CCPE is needed that clearly highlights all the benefits of the methodology, including factors such as consensus and knowledge transfer, and a clear description of the usefulness of the results provided by the methodology. It should also clarify what is required of the development process and method users to make CCPE function optimally.

It is necessary to create educational material that helps users grasp the methodology in large enough steps not to be overwhelmed by its complexity and terminology, while at the same time providing support in how the methodology can be adapted to different evaluation cases. There also need to be examples of both simple and complex machines that show CCPE's ability to generalize and how it works for different applications.

Creating a computerized version that reduces the need for manual administration is an improvement to reduce the time required in result presentation. Tables and matrices would thus be created directly from the templates loaded directly in the computer program. Such a software program would also be very helpful in both learning and implementation of the methodology.

In the current version of CCPE there is no evaluation of body posture and body load during interaction analysis. One suitable area for future work is to thus integrate this in CCPE in a structured way.

Since CCPE methodology was formed its present design in the last iteration of the thesis, it needs to continue being tested by new users to find further improvements. As CCPE is meant to be an active methodology, the action research spiral is meant to be an integrated part of CCPE.

Another area for further development is use of the methodology by other users. The analyses with ECW and PUEA were mainly carried out by the author. The methodology needs to be used by more people to ascertain if the method's results do not only depend on the author's own knowledge.

The full CCPE methodology also needs to be validated in order to better identify the strengths and weaknesses against other evaluation methodologies and methods. CCPE cannot currently be said to be a fully completed validated methodology; instead, further development is needed for additional improvement and adaptation to different target groups and evaluation situations.

CCPE methodology can be adapted to evaluate interaction between human and machine. This ability to adapt makes the methodology complex and a possible path of future work is therefore to provide clear descriptions of how CCPE can be adapted for analysis of different systems. A clear description of the adjustments will make it easier for new users of the methodology to use it effectively and thus benefit more from it.

Even if CCPE was developed with the product development process in mind, no formal testing of its possibilities to function a real process has taken place. Further work could focus on applying CCPE in a structured way during a product development process in industry, and then evaluate the results.

No formal testing has been done on how CCPE interacts with other HFE methods, for example in a real development process in industry. CCPE is not intended to work individually, which means that studies are needed about which other methods work with CCPE and how integration between the different methods will be conducted.

Although CCPE has shown to work in many different domains, it is still important to subject CCPE to further testing in other areas in order to better map when, where and how CCPE works. This gives CCPE increased ability to generalize but also helps future methodology users in different domains to know if, when and how CCPE is to be used.

9 References

ALMEFELT, L. 2005. *Requirements-driven product innovation: methods and tools reflecting industrial* PhD Thesis, Chalmers University of Technology.

ANDERSSON, J., BLIGÅRD, L.-O., OSVALDER, A.-L., RISSANEN, M. & TRIPATHI, S. 2011. To Develop Viable Human Factors Engineering Methods for Improved Industrial Use. *Lecture Notes in Computer Science,* 6769**,** 355-362.

ARAUJO, C. S. 2001. *Acqusition of Product Development Tools in Industry: A Theoretical Contrubution.* PhD Thesis, Technical University of Denmark.

AXELSSON, F. 2002. *Utveckling av användarvänlig medicinteknisk utrustning: utvärdering av användargränssnitt på dialysapparat med Cognitive walkthrough och designprinciper.* Master Thesis, Chalmers University of Technology.

BABER, C. & STANTON, N. A. 1996. Human error identification techniques applied to public technology: Predictions compared with observed use. *Applied Ergonomics,* 27**,** 119-131.

BERMINGE, K. 1993. *En vetenskapshistorisk översikt,* Göteborg, Institutionen för vetenskapsteori, Göteborgs Universitet.

BIRKLER, J. 2008. *Vetenskapsteori: en grundbok,* Stockholm, Liber.

BLANDFORD, A., KEITH, S., CONNELL, I. & EDWARDS, H. Year. Analytical usability evaluation for digital libraries: A case study. *In:* Proceedings of the ACM IEEE International Conference on Digital Libraries, JCDL 2004, 2004. 27-36.

BLIGÅRD, L.-O. 2003. Risk Analysis on Use Vivo 30/40, Internal Report Beras Medical AB, Classified.

BLIGÅRD, L.-O. 2004. Risk Analysis on Use iSleep 20/21, Internal Report Beras Medical AB, Classified.

BLIGÅRD, L.-O. 2007. *Prediction of Medical Device Usability Problems and Use Errors – An Improved Analytical Methodical Approach.* Licentiat Thesis, Chalmers University of Technology.

BLIGÅRD, L.-O. 2011. Utvecklingsprocessen ur ett människa-maskinperspektiv. Chalmers University of Technology.

BLIGÅRD, L.-O., ERIKSSON, M. & OSVALDER, A.-L. 2006a. Cassette versus Conventional Bloodlines - Analytical Evaluation of Three Dialysis Machines, Internal Report Gambro Lundia AB, Classified.

BLIGÅRD, L.-O., ERIKSSON, M. & OSVALDER, A.-L. 2006b. Work Physiology Evaluation of New Machine Prototype II, Internal Report Gambro Lundia AB, Classified.

BLIGÅRD, L.-O., JÖNSSON, A. & OSVALDER, A.-L. 2003a. Utvärdering och konceptdesign av användargränssnitt för insulinpump. Göteborg: Chalmers University of Technology.

BLIGÅRD, L.-O., JÖNSSON, A. & OSVALDER, A.-L. Year. Eliciting User Requirements and Evaluation of Usability for Insulin Pumps. *In:* The 36th Annual Congress of the Nordic Ergonomics Society Conference, 2004 Kolding, Denmark.

BLIGÅRD, L.-O. & OSVALDER, A.-L. 2006a. Identification of use errors with Prisma Flex: A pilot study, Internal Report Gambro Lundia AB, Classified.

BLIGÅRD, L.-O. & OSVALDER, A.-L. Year. Using Enhanced Cognitive Walkthrough as a Usability Evaluation Method for Medical Equipment. *In:* IEA, 2006b Maastricht, Netherlands

BLIGÅRD, L.-O. & WASS, S. 2002. *Analysis and Development of User Interface for Home Care Airflow Generators.* Master Thesis, Chalmers Universtity of Technology.

BLIGÅRD, L., -O., WASS, S., LILJEGREN, E. & OSVALDER, A., -L. . Year. Using a Human Factors Engineering Process to Develop New User Interfaces for Home Car Air-Flow Generators. *In:* The 35th Annual Congress of the Nordic Ergonomics Society Conference 2003b Reykjavik, Island.

BOHGARD, M. 2009. *Work and technology on human terms,* Stockholm, Prevent.

BOY, G. A. 2011. *The handbook of human-machine interaction: a human-centered design approach,* Farnham, Surrey, England, Ashgate.

BRANSBY, M. & JENKINSON, J. 1998. Alarming performance. *Computing and Control Engineering Journal,* 9**,** 61-67.

BÜYÜKZKAN, G. & ARSENYAN, J. 2012. Collaborative product development: A literature overview. *Production Planning and Control,* 23**,** 47-66.

CACCIABUE, P. C. 2004. *Guide to applying human factors methods: human error and accident management in safety-critical systems,* London Springer, cop.

CHAPANIS, A. 1965. *Man-machine engineering,* Belmont, Wadsworth.

CHAPANIS, A. 1985. Some reflections on progress. *Human Factors Society 20th Meeting.* Santa Monica CA.

COOPER, L. & BABER, C. 2005. Focus Groups. *In:* STANTON, N. A., HEDGE, A., BROOKHUIS, K., SALAS, E. & HENDRICK, H. (eds.) *Handbook of Human Factors and Ergonomics Methods.* London: CRC Press.

CROSS, N. & CLAYBURN CROSS, A. 1995. Observations of teamwork and social processes in design. *Design Studies,* 16**,** 143-170.

CUOMO, D. L. & BOWEN, C. D. 1994. Understanding usability issues addressed by three user-system interface evaluation techniques. *Interacting with Computers,* 6**,** 86-108.

DAHLÉN, C. & ULLSTRÖM, T. 2006. *Design av morgondagens insulinpump.* Master Thesis, Chalmers University of Technology.

DEFENCE, U. D. O. 1980. Procedures for Performing a Failure Mode, Effects and Criticality Analysis. *Military Standard. MIL-1629a.* US Department of Defense.

DESMET, P. M. A. 2003. Measuring emotion; development and application of an instrument to measure emotional responses to products. *In:* BLYTHE, M. A., MONK, A. F., OVERBEEKE, K. & WRIGHT, P. C. (eds.) *Funology: from usability to enjoyment.* Dordrecht: Kluwer Academic Publishers.

DHILLON, B. S. 2003. *Engineering safety: fundamentals, techniques, and applications,* River Edge, World Scientific.

DICK, B. 2003. *Action Reseach Resources* [Online]. Available: http://www.scu.edu.au/schools/gcm/ar/arhome.html [Accessed 2006-04-01 2006].

EEMUA 1999. 191 Alarm Systems - A Guide to Design, Management and Procurement. London: Engineering Equipment and Materials Users Association.

ELLINGSEN, K. & LUNDMARK, E. 2009. *Study of the Operation and Information Environment for Commander in Infantry Fighting Vehicle.* Master Thesis, Chalmers University of Technology.

EMBREY, D. 1986. SHERPA: a Systematic Human Error Reduction and Prediction Approach. *International Topical Meeting on Advances in human factors in nuclear power system.* Knoxville: American Nuclear Society.

EMBREY, D. 2004. Qualitative and quantitative evaluation of human error in risk assessment. *In:* SANDOM, C. & HARVEY, R. (eds.) *Human Factors for Engineers.* London Institution of Electrical Engineers.

ENGELBREKTSSON, P. 2004. *Enabling the user: exploring methodological effects on user requirements elicitation.* PhD Thesis, Chalmers University of Technology.

ENGESTRÖM, Y. 1987. *Learning by expanding: an activity-theoretical approach to developmental research,* Helsinki, Orienta-konsultit.

EYSENCK, M. W. 2012. *Fundamentals of psychology,* London, Psychology Press.
FDA. 1999. *Proposal for Reporting of Use Errors with Medical Devices* [Online]. Available: http://www.fda.gov/ohrms/dockets/98fr/992075bk.pdf [Accessed 2006-04-12].
FDA. 2001. *Questions and Answers from the February 14, 2001 Teleconference: Integrating Human Factors Engineering into Medical Device Design and Development* [Online]. Available: http://www.fda.gov/cdrh/humfac/qanda_021401.html [Accessed 2006-03-12].
FLOOD, R. L. & CARSON, E. R. 1993. *Dealing with complexity : an introduction to the theory and application of systems science,* New York, Plenum.
GALITZ, W. O. 2007. *The essential guide to user interface design: an introduction to GUI design principles and techniques,* Indianapolis, IN, Wiley.
GARMER, K. 2002. *Participatory ergonomics in practice: an evaluation of three methods for supporting end-user participation in the development and choice of technical equipment.* PhD Thesis, Chalmers University of Technology.
GARMER, K., LILJEGREN, E., OSVALDER, A.-L. & DAHLMAN, S. 2002. Arguing for the need of triangulation and iteration when designing medical equipment: Nature of requirements on medical equipment derived from a Human Factors approach and from usability tests: Methodological aspects. *Journal of Clinical Monitoring and Computing,* 17**,** 105-114.
GIBSON, J. J. 1979. *The ecological approach to visual perception,* Boston, Mass., Houghton Mifflin.
GREEN, T. R. G., BURNETT, M. M., KO, A. J., ROTHERMEL, K. J., COOK, C. R. & SCHONFELD, J. Year. Using the cognitive walkthrough to improve the design of a visual programming experiment. *In*, 2000 Seattle, WA, USA. IEEE Comput. Soc, 172-9.
GROOME, D. & BRACE, N. 2006. *An introduction to cognitive psychology: processes and disorders,* Hove, Psychology Press.
GROSS, A. & FOUFAS, J. 2003. *Developing a Human Factors Engineering Process for Medical Device Design. Methodological considerations.* Master Thesis, Chalmers University of Technology.
GRUDIN, J. 1992. Utility and usability: research issues and development contexts. *Interacting with Computers,* 4**,** 209-17.
GULLIKSEN, J. & GÖRANSSON, B. 2002. *Användarcentrerad systemdesign: en process med fokus på användare och användbarhet*, Studentlitteratur.
GYLLENSVAAN, N. & OLSSON, L. 2009. *Rapid Usability Evaluation - en metodstudie.* Bachelor Thesis, Chalmers University of Technology.
HANSSON, B. 2011. *Skapa vetande: vetenskapsteori från grunden,* Lund, Studentlitteratur.
HARMS-RINGDAHL, L. 2001. *Safety Analysis - Principles and Parctice in Occupational Safety,* London Taylor & Francis.
HARRIS, D., STANTON, N. A., MARSHALL, A., YOUNG, M. S., DEMAGALSKI, J. & SALMON, P. 2005. Using SHERPA to predict design-induced error on the flight deck. *Aerospace Science and Technology,* 9**,** 525-532.
HART, S. G. & STAVELAND, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *In:* HANCOCK, P. A. & MESHKATI, N. (eds.) *Human Mental Workload.* Amsterdam: Elsevier.
HARTSON, H. R., ANDRE, T. S. & WILLIGES, R. C. 2001. Criteria for evaluating usability evaluation methods. *International Journal of Human-Computer Interaction,* 13**,** 373-410.

HENRIKSSON, U. & STRÄNGBERG, G. 2007. *A usability study of Orwak balers - Methodology, analysis and design guidelines for the machines in the Orwak 3000-series.* Master Thesis, Chalmers Universty of Technology.
HIGNETT, S. & MCATAMNEY, L. 2000. Rapid Entire Body Assessment (REBA). *Applied Ergonomics,* 31**,** 201-205.
HIX, D. & HARTSON, H. R. 1993. *Developing user interfaces: ensuring usability through product & process,* New York, Wiley.
HOLLNAGEL, E. 2004. *Barriers and accident prevention,* Burlington, Ashgate.
HORSKY, J., KAUFMAN, D. R., OPPENHEIM, M. I. & PATEL, V. L. 2003. A framework for analyzing the cognitive complexity of computer-assisted clinical ordering. *Journal of Biomedical Informatics,* 36**,** 4-22.
HYMAN, W. 1995. The issue is 'use' not 'user' error. *Medical Device& Diagnostic Industry,* 17**,** 58-60.
IEA. 2006. *International Ergonomics Association Web Page* [Online]. Available: http://www.iea.cc/ [Accessed 2006-04-02 2006].
IEC 2003. IEC 60601-1-8:2003 Medical electrical equipment - Part 1-8: Collateral Standard: Alarms Systems—General Requirements, Tests, and Guidance for Alarm Systems in Medical Electrical Equipment and Medical Electrical Systems. Geneva: IEC.
IEC 2004a. IEC 60601-1-6:2004 Medical electrical equipment - Part 1-6: General requirements for safety - Collateral standard: Usability. Geneva: IEC.
IEC 2004b. IEC 62241:2004 Nuclear power plants - Main control room - Alarm functions and presentation. Geneva International Electrotechnical Commission.
ISO 1998. ISO 9241-11:1998 : Ergonomic requirements for office work with visual display terminals (VDTs) - Part 11: Guidance on usability. Geneve: International Standard Organization.
ISO 1999. ISO 13407:1999 Human-centred design processes for interactive systems
Geneva: International Organization for Standardization.
ISO 2000a. ISO 11064-1:2000 Ergonomic design of control centres - Part 1: Principles for the design of control centres. Geneve: International Organization for Standardization.
ISO 2000b. ISO 14971:2000 Medical devices - Application of risk management to medical devices. Geneve: International Standard Organization.
ISRAELSKI, E. W. & MUTO, W. H. 2005. Human Factors Risk Management as a Way to Improve Medical Device Safety: A Case Study of the Therac-25 Radiation Therapy System *In:* GOSBEE, J. W. & GOSBEE, L. L. (eds.) *Using Human Factors to Improve Patient Safety.* Oakbrook Terrace: Joint Commission Resources, Inc.
JANHAGER, J. 2005. *User Consideration in Early Stages of Product Development – Theories and Methods.* PhD Thesis, The Royal Institute of Technology.
JOHANNESSON, H., PERSSON, J.-G. & PETTERSSON, D. 2004. *Produktutveckling: effektiva metoder för konstruktion och design,* Stockholm, Liber.
JORDAN, P. W. 1998. *An introduction to usability,* London, Taylor & Francis.
KARLSSON, I. C. M. 1996. *User Requirements Elicitation - A Framework for the Study of the Relation between User and Artefact.* PhD Thesis, Chalmers University of Technology.
KAUFMAN, D. R., PATEL, V. L., HILLIMAN, C., MORIN, P. C., PEVZNER, J., WEINSTOCK, R. S., GOLAND, R., SHEA, S. & STARREN, J. 2003. Usability in the real world: assessing medical information technologies in patients' homes. *Journal of Biomedical Informatics,* 36**,** 45-60.
KAULIO, M., KARLSSON, I. C. M., GRUBB, H. & MELLBY, C. 1999. *PRE: product requirements engineering: kundförståelse i produktutvecklingen* Mölndal, Institutet för verkstadsteknisk forskning (IVF) & Chalmers University of Technology.

KAYE, R. & CROWLEY, J. 2000. *Medical Device Use-Safety: Incorporating Human Factors Engineering into Risk Management,* Rockville, US Food and Drug Administration.

KHAJOUEI, R., HASMAN, A. & JASPERS, M. W. M. 2011. Determination of the effectiveness of two methods for usability evaluation using a CPOE medication ordering system. *International Journal of Medical Informatics,* 80**,** 341-350.

KIM, J. W. & JUNG, W. 2003. A taxonomy of performance influencing factors for human reliability analysis of emergency tasks. *Journal of Loss Prevention in the Process Industries,* 16**,** 479-495.

KIRWAN, B. 1994. *A guide to practical human reliability assessment,* London, Taylor & Francis.

KITAJIMA, M. & POLSON, P. G. Year. A comprehension-based model of exploration. *In*, 1996 Vancouver, BC, Canada. ACM, 324-31.

KOLLURO, R. V. & BROOKS, D. G. 1996. Integrated Risk Assessment and Strategic Management. *In:* KOLLURO, R., BARTELL, S., PITBLADO, R. & STRICOFF, S. (eds.) *Risk assessment and management handbook: for environmental, health, and safety professionals.* London McGraw-Hill.

KOUTSABASIS, P., SPYROU, T. & DARZENTAS, J. 2007. Evaluating usability evaluation methods: Criteria, method and a case study. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).*

KUSHNIRUK, A. W., KAUFMAN, D. R., PATEL, V. L., LÉVESQUE, Y. & LOTTIN, P. 1996. Assessment of a computerized patient record system: a cognitive approach to evaluating medical technology. *M.D. computing: computers in medical practice,* 13**,** 406-415.

LANE, R., STANTON, N. A. & HARRISON, D. 2006. Applying hierarchical task analysis to medication administration errors. *Applied Ergonomics,* 37**,** 669-679.

LAUNIS, M. 2001. Participation and Collaboration in Workplace Design. *In:* KARWOWSKI, W. (ed.) *International Encyclopedia of Ergonomics and Human Factors* London and New York: CRC Press.

LEARNER, D. B. & PHILLIPS, F. Y. 1993. Method and progress in management science. *Socio-Economic Planning Sciences,* 27**,** 9-24.

LEVENTHAL, L. M. & BARNES, J. A. 2008. *Usability engineering: process, products, and examples,* Upper Saddle River, N.J., Pearson/Prentice Hall.

LEVESON, N. G. 1995. *SafeWare: system safety and computers* Reading, Addison-Wesley.

LEWIS, C. & WHARTON, C. 1997. Cognitive Walkthrough. *In:* HELANDER, M., LANDAUER, T. K. & PRABHU, P. (eds.) *Handbook of Human-computer Interaction.* New York: Elsevier Science BV.

LILJEGREN, E. 2006. Usability in a medical technology context assessment of methods for usability evaluation of medical equipment. *International Journal of Industrial Ergonomics,* 36**,** 345-352.

LILJEGREN, E., BLIGÅRD, L.-O. & OSVALDER, A.-L. Year. Developing User-Friendly Interfaces for Medical Devices. *In:* 8th IFAC Symposium on Automated Systems Based on Human Skill, 2003 Göteborg, Sweden.

LILJEGREN, E. & OSVALDER, A.-L. 2004. Cognitive engineering methods as usability evaluation tools for medical equipment. *International Journal of Industrial Ergonomics,* 34**,** 49-62.

LIU, Y. 2004. *Usability evaluation pf interface design for medical devices - exploring methodologies for evaluation prototypes.* Licentiate Thesis, Chalmers Universty of Technology.

LIU, Y., OSVALDER, A.-L. & DAHLMAN, S. 2005. Exploring user background settings in cognitive walkthrough evaluation of medical prototype interfaces: A case study. *International Journal of Industrial Ergonomics,* 35**,** 379-390.

LUNDGREN, L. 2010. Analys av Eson Trauma Transfer: Utvärdering av produktanvändning Halmstad: Hälsoteknikcentrum Halland.

LUNDGREN, L., BLIGÅRD, L.-O., BRORSSON, S. & OSVALDER, A.-L. 2011. Implementation of usability analysis to detect problems in the management of kitesurfing equipment. *5th Asia-Pacific Congress on Sports Technology.* Melbourne.

MAHATODY, T., SAGAR, M. & KOLSKI, C. 2010. State of the art on the cognitive walkthrough method, its variants and evolutions. *International Journal of Human-Computer Interaction,* 26**,** 741-785.

MALTIN, M. W. 2009. *Cognitive Psychology,* Hoboken, Wiley.

MATTES, J., TRERNBLAND, M., ELIASSON, K. & NYMAN, T. 2012. Methods for ergonomic risk assessment at the work site - What´s in use and why? *Poster presented at the 44th International Conference of the Nordic Ergonomic Society.* Saltsjöbaden.

MCATAMNEY, L. & CORLETT, E. N. 1993. RULA: A survey method for the investigation of work-related upper limb disorders. *Applied Ergonomics,* 24**,** 91-99.

MESHKATI, N., HANCOCK, P. A., RAHIMI, M. & DAWES, S. M. 1995. Techniques in Mental Workload Assessment. *In:* WILSON, J. & CORLETT, N. (eds.) *Evaluation of Human Work.* 2 ed. London: CRC Press.

MILITELLO, L. G. & HUTTON, R. J. B. 1998. Applied Cognitive Task Analysis (ACTA: a practitioner's toolkit for understanding cognitive task demands. *Ergonomics,* 41**,** 1618-1641.

MILLER, J. R. & JEFFRIES, R. 1992. Interface-usability evaluation: science of trade-offs. *IEEE Software,* 9**,** 97-8.

MOIRA, F. T. 1990. Biomechanics methods in posture analysis. *In:* WILSON, J. R. & CORLETT, E. N. (eds.) *Evaluation of human work.* Cornwall: Taylor & Francis Inc.

MONÖ, R. G. 1997. *Design for product understanding: the aesthetics of design from a semiotic approach,* Stockholm, Liber.

MORADI, S. & POUR, N. A. 2011. Usability problems and use errors appraisal in public transport website. *In:* DE WAARD, D., GÉRARD, L., WICZOREK, R. & MANZCY, D. (eds.) *Human Centred Automation.* Maastricht: Shaker Pulishing.

MORIC, A., BLIGÅRD, L.-O. & OSVALDER, A.-L. 2004. Usability of Reusable SpO2 Sensors: A Comparison between two Sensor Types. *36th Annual Congress of the Nordic Ergonomics Society Conference (NES).* Kolding, Denmark.

MÅRTENSSON, M. 2010. *User Interface Design of a Cardiac Output Monitor - Redesign with focus on usability and displaying quantitative information.* Master Thesis, Chalmers University of Technology.

NATIONALENCYKLOPEDIN. 2012. *Nationalencyklopedin* [Online]. Stockholm: Nationalencyklopedin Available: www.ne.se [Accessed 2012-02-07 2012].

NIELSEN, J. 1993. *Usability engineering,* Boston, Academic Press.

NIELSEN, J. & MACK, R. L. (eds.) 1994. *Usability inspection methods,* New York: Wiley.

NIELSEN, L. 2004. *Engaging Personas and Narrative Scenarios,* Copenhagen, Samfundslitteratur.

NILSSON, D. & OREDSSON, V. 2010. *Product development and evaluation of movable inserts for the TENA assortment.* Master Thesis, Chalmers University of Technology.

NILSSON, W. & ÅHMAN, M. 2005. *behandlat enligt Human Factors Engineering.* Master Thesis, Chalmers Univeristy of Technology.

NORELL, M. 1992. *Stödmetoder och samverkan i produktutvecklingen.* PhD Thesis, The Royal Institute of Technology.
NORMAN, D. A. 2002. *The Design of Everyday Things,* New York, Basic Books Inc.
NORMAN, D. A. & DRAPER, S. W. 1986. *User centered system design: new perspectives on human-computer interaction,* Hillsdale, N.J., Erlbaum.
NRC 2004. Human Factors Engineering Program Review Model (NUREG-0711, Revision 2) U.S. Nuclear Regulatory Commission.
OBERT, C. & FORSELL, M. 2000. *Fokusgrupp: ett enkelt sätt att mäta kvalitet,* Höganäs, Kommunlitteratur.
OSVALDER, A.-L., PETTERSSON, I. & BLIGÅRD, L.-O. 2005. Work Physiology Evaluation of New Machine Prototype I, Internal Report Gambro Lundia AB, Classified.
OXSTRAND, J. 2006. *Utveckling av säkert och effektivt användargränssnitt – vilka faktorer påverkar risken för mänskligt fel?* Master Thesis, Chalmers University of Technology.
PATTERSON, P. A. & NORTH, R. A. 2006. Fitting Human Factors in the Product Development Process. *Medical Device & Diagnostic Industry Magazine.*
PERSSON, S. 2005. *Toward enhanced interaction between engineering design and industrial design.* PhD Thesis, Chalmers University of Technology.
PETTERSSON, I. & OSVALDER, A.-L. 2005. Ergonomic Evaluation of Child Car Seats - Comfort and Usability.
PHEASANT, S. 1990. Antropometry and the design of workspaces. *In:* WILSON, J. R. & CORLETT, E. N. (eds.) *Evaluation of human work.* Cornwall: Taylor & Francis Inc.
PHEASANT, S. & HASLEGRAVE, C. M. 2006. *Bodyspace: anthropometry, ergonomics and design of work* London Taylor & Francis.
PORTER, J. M., FREER, M., CASE, K. & BONNEY, M. C. 1990. Computer aided ergonomics and workspace design. *In:* WILSON, J. R. & CORLETT, E. N. (eds.) *Evaluation of human work.* Cornwall: Taylor & Francis Inc.
RASMUSSEN, J. 1983. Skills, rules and knowledge; signals, signs and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man and Cybernetics,* SMC-13**,** 257-66.
RASMUSSEN, J. & SVEDUNG, I. 2000. *Proactive risk management in a dynamic society,* Karlstad, Swedish Rescue Services Agency (Statens räddningsverk).
REASON, J. 1990. *Human error,* Cambridge Cambridge Univ. Press, cop.
REASON, J. 1997. *Managing the Risks of Organizational Accidents,* Aldershot, Ashgate.
REASON, P. & BRADBURY, H. 2001. *Handbook of action research: participative inquiry and practice,* London
REISBERG, D. 2001. *Cognition: exploring the science of the mind,* New York, Norton.
REXFELT, O., ALMEFELT, L., ZACKRISSON, D., HALLMAN, T., MALMQVIST, J. & KARLSSON, M. 2011. A proposal for a structured approach for cross-company teamwork: A case study of involving the customer in service innovation. *Research in Engineering Design,* 22**,** 153-171.
RÖNNERMAN, K. 2004. *Aktionsforskning i praktiken - erfarenheter och reflektioner,* Lund, Studentlitteratur.
SANDOM, C. 2004. Safety assessment and human factors. *In:* SANDOM, C. & HARVEY, R. (eds.) *Human Factors for Engineers.* London Institution of Electrical Engineers.
SANDSJÖ, L. 2004. *Ambulatory Monitoring and Analysis of Surface Electromygraphic Signals in Ergonomic Field Studies.* PhD-thesis, Chalmers University of Thechnology.
SHARP, H., ROGERS, Y. & PREECE, J. 2007. *Interaction design: beyond human-computer interaction,* Hoboken, N.J., Wiley.

SKYTTNER, L. 2005. *General systems theory: problems, perspectives, practice,* Singapore, World scientific.
SOHLBERG, P. & SOHLBERG, B.-M. 2009. *Kunskapens former: vetenskapsteori och forskningsmetod,* Malmö, Liber.
SOLSO, R. L. 2001. *Cognitive psychology,* Boston, Allyn and Bacon.
STANTON, N. 2005. *Human factors methods: a practical guide for engineering and design,* Aldershot, Ashgate.
STANTON, N. & BABER, C. 1996. A systems approach to human error identification. *Safety Science,* 22**,** 215-228.
STANTON, N. A. 2006. Hierarchical task analysis: Developments, applications, and extensions. *Applied Ergonomics,* 37**,** 55-79.
STANTON, N. A. & BABER, C. 2002. Error by design: Methods for predicting device usability. *Design Studies,* 23**,** 363-384.
STANTON, N. A. & BABER, C. 2005. Validating task analysis for error identification: Reliability and validity of a human error prediction technique. *Ergonomics,* 48**,** 1097-1113.
STANTON, N. A., HEDGE, A., BROOKHUIS, K., SALAS, E. & HENDRICK, H. (eds.) 2005. *Handbook of human factors and ergonomics methods,* New York: CRC Press.
STANTON, N. A. & STEVENAGE, S. V. 1998. Learning to predict human error: Issues of acceptability, reliability and validity. *Ergonomics,* 41**,** 1737-1756.
STANTON, N. A. & YOUNG, M. S. 1999. *A Guide to Methodology in ergonomics,* London, Taylor & Francis.
STRICOFF, S. 1996. Safety Risk Analysis and Process Safety Management. *In:* KOLLURO, R., BARTELL, S., PITBLADO, R. & STRICOFF, S. (eds.) *Risk assessment and management handbook: for environmental, health, and safety professionals.* London McGraw-Hill.
STRÖMBERG, H. & FREYHALL, T. 2009. *A multifunctional navigation aid for use aboard boats - User centred development of an innovative touch screen interface.* Master Thesis, Chalmers University of Technology.
SUN, J., YU, S., GE, C. & JIANG, C. Year. A collaborative knowledge management system for product design based on folksonomy. *In*, 2010. 1385-1387.
SUOKAS, J. & PYY, P. 1988. Evaluation of the validity of four hazard identification methods with event description. Espoo: Technical Reserach Centre of Finland.
SWAIN, A. D. & GUTTMAN, H. E. 1983. *A Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Application, Nureg/CR-1278,* Washingto DC-2055, USNRC.
TANCREDI, W., WIVÄNGE, Å. & OLOF, T. 2012. Expert Usability Evaluation Of An Electronic Medical Records System – Inspection Methods And Their Appropriateness For Revealing Different Types Of Usability Issues. *To be submitted to International Journal of Medical Informatics*.
TAYLOR, J. R. 1979. A Background to Risk Analysis. Risö: Electric Department, Risö National Laboratory.
TAYLOR, J. R. 1994. *Risk analysis for process plant, pipelines and transport,* London, E & FN Spon.
THUNBERG, A. 2006. *A Cognitve Approach to the Design of Alarm System for Nuclear Power Plant Control rooms.* Licentiate Thesis, Chalmers University of Technology.
ULLMAN, D. G. 2002. *The mechanical design process,* New York, McGraw-Hill.
ULRICH, K. T. & EPPINGER, S. D. 2004. *Product design and development,* Boston, McGraw-Hill.

WESTLING, G. 2002. *Balancing innovation and control: the role of face-to-face meetings in complex product development projects.* PhD Thesis, Economic Research Institute, Stockholm School of Economics.
WHARTON, C., BRADFORD, J., JEFFRIES, R. & FRANZKE, M. 1992. Applying cognitive walkthroughs to more complex user interfaces: experiences, issues, and recommendations. *ACM Conference on Human Factors in Computing Systems - CHI '92***,** 381-388.
WHARTON, C., RIEMAN, J., LEWIS, C. & POLSON, P. G. 1994. The Cognitive Walkthrough Method: A Practitioners´s Guide. *In:* NIELSEN, J. & MACK, R. L. (eds.) *Usability Inspection Methods* New York, UK: John Wiley And Sons Ltd.
WHITTINGHAM, R. B. 2004. *The blame machine: why human error causes accidents,* Amsterdam Elsevier Butterworth-Heinemann.
WICKENS, C. D. & HOLLANDS, J. G. 1999. *Engineering psychology and human performance,* Upper Saddle River, Prentice Hall.
WIDING, M. & BUI, V. 2012. *Usage study of flat saws - Ergonomic evaluation and concept development of interaction device.* Master Thesis, Chalmers University of Technology.
WILLIAMS, J. C. 1986. HEART - A proposed Method for Assessing and Reducing Human Error. *Proceedings - Advances in Reliability Technology Symposium.*
WILSON, J. & CORLETT, N. (eds.) 1995. *Evaluation of human work: practical ergonomics methodology,* London Taylor & Francis.
WINKEL, J. & MATHIASSEN, S. E. 1994. Assessment of physical work load in epidemiologic studies: Concepts, issues and operational considerations. *Ergonomics,* 37**,** 979-988.
WIXON, D. 2003. Evaluating usability methods: why the current literature fails the practitioner. *Interactions.* Association for Computing Machinery.
WOODS, D. & COOK, R. I. 1999. *The New Look at Error, Safety, and Failure: A Primer for Health Care* [Online]. Available: csel.eng.ohio-state.edu/woods/error/NewLook_primer.pdf [Accessed 2006-11-09 2006].
YOUNG, M. S. & STANTON, N. A. 2004. Applying Interviews to Usability Assessment. *In:* STANTON, N., HEDGE, A., BROOKHUIS, K., SALAS, E. & HENDRICK, H. (eds.) *Handbook of Human Factors and Ergonomics Methods.* New York: CRC Press.