DEVELOPMENT OF AN EVALUATION METHOD FOR DRIVER VEHICLE INTERACTION

Maria Gustafsson and Anna Tornvall

DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2018

MASTER THESIS

Development of an Evaluation Method for Driver Vehicle Interaction

Maria Gustafsson and Anna Tornvall

LUND UNIVERSITY

Development of an Evaluation Method for Driver Vehicle Interaction

Published by

Department of Design Sciences
Faculty of Engineering LTH, Lund University
P.O. Box 118, SE-221 00 Lund, Sweden

Subject: Technical Design (MMKM10)
Supervisor: Kirsten Rassmus-Gröhn
Co-supervisor: Per Kristav
Examiner: Johanna Persson

Abstract

The master's thesis was performed in collaboration with the department responsible for the driver vehicle interaction at Scania CV AB. The interaction between the driver and the vehicle contributes to the usability and overall experience of the product as well as to safety. In order to create a good interaction it is important to evaluate it. The purpose of the thesis work was therefore to develop an evaluation method that could be used throughout the design process to evaluate the driver vehicle interaction experience related to trucks and buses. The thesis work focused on expert evaluations rather than user tests. Investigations of *what* factors should be evaluated, *how* the factors should be evaluated and *when* in the design process the evaluation should be carried out were conducted.

Literature studies and interviews with employees at Scania were performed, which resulted in the selection of factors to be evaluated as well as the specific context-of-use. A list of criteria for evaluation was compiled and translated into Key Performance Indicators to be measured. These laid the foundation for the selection of suitable evaluation methods as well as for the development of evaluation frameworks and recommendations for each selected evaluation method. The frameworks were tested on employees at Scania and refined accordingly. In addition, a case study was conducted with the purpose to evaluate an ongoing project at Scania with the new frameworks.

The final result of the thesis work is a selection of nine evaluation methods that should be used to evaluate the driver vehicle interaction. The proposal of the evaluation methods consists of four evaluation frameworks and recommendations regarding five evaluation methods. All evaluation methods are based on the Key Performance Indicators that derive from the selection of factors to be evaluated, namely pleasurability, task focus and flexibility. The majority of the methods should be conducted by experts and the remaining require user tests. The evaluation methods should be executed during different stages of the design process depending on the required product representation.

Keywords: User Experience, usability, evaluation method, driver vehicle interaction, Scania

Sammanfattning

Examensarbetet utfördes i samarbete med avdelningen ansvarig för interaktionen mellan förare och fordon på Scania CV AB. Interaktionen mellan förare och fordon bidrar till användbarheten och den övergripande upplevelsen av produkten såväl som till säkerhet. För att skapa en god interaktion är det viktigt att utvärdera denna. Syftet med examensarbetet var därför att utveckla en utvärderingsmetod som kunde användas genom hela designprocessen för att utvärdera interaktionsupplevelsen mellan förare och fordon avseende lastbilar och bussar. Examensarbetet fokuserade på expertutvärderingar snarare än användartester. Undersökningar av *vilka* faktorer som ska utvärderas, *hur* faktorerna ska utvärderas och *när* i designprocessen utvärderingen ska ske genomfördes.

Litteraturstudier och intervjuer med anställda på Scania utfördes, vilket resulterade i valet av faktorer som ska utvärderas och det specifika användningsområdet. En lista över utvärderingskriterier sammanställdes och översattes till nyckeltal som ska mätas. Dessa lade grunden för valet av passande utvärderingsmetoder samt för utvecklingen av utvärderingsramverk och rekommendationer för respektive vald utvärderingsmetod. Ramverken testades på anställda på Scania och förbättrades därefter. Vidare utfördes en fallstudie med syfte att utvärdera ett pågående projekt på Scania med de nya utvärderingsramverken.

Det slutgiltiga resultatet av examensarbetet är ett val av nio utvärderingsmetoder som ska användas för att utvärdera interaktionen mellan förare och fordon. Förslaget av utvärderingsmetoder består av fyra utvärderingsramverk samt rekommendationer avseende fem utvärderingsmetoder. Alla utvärderingsmetoder är baserade på nyckeltalen som härleds från valet av faktorer som ska mätas, nämligen nöje, fokus på uppgiften och flexibilitet. Majoriteten av metoderna ska genomföras av experter och resterande kräver användartester. Utvärderingsmetoderna ska utföras under olika stadier av designprocessen beroende på vilken produktrepresenation som krävs.

Nyckelord: användarupplevelse, användbarhet, utvärderingsmetod, interaktion mellan förare och fordon, Scania

Acknowledgments

Foremost, we would like to thank Stefan Larsson, our supervisor at Scania, for guiding us through the thesis work.

We would also like to thank the department of Driver Vehicle Interaction for welcoming us to Scania. A special thanks to the employees who have participated in interviews and tests. We are now looking forward to continuing our careers as trainees at Scania.

We would like to thank Kirsten Rassmus-Gröhn, our supervisor at the department of Design Sciences at LTH, for providing us with valuable inputs and guidance throughout the thesis work.

Finally, we would like to thank Johanna Persson, our examiner, and our opponents. We hope you enjoyed the reading.

Lund, June 2018

Maria Gustafsson and Anna Tornvall

Table of contents

List of acronyms and abbreviations

ASWAT	Adapted Subjective Workload Assessment Technique
AUS	Infotainment System
CW	Cognitive Walkthrough
DALI	Driver Activity Load Index
DVI	Driver Vehicle Interaction
ECW	Enhanced Cognitive Walkthrough
GUI	Graphical User Interface
HE	Heuristic Evaluation
HTA	Hierarchical Task Analysis
HMI	Human Machine Interface
ICL	Instrument Cluster
ISO	International Organization of Standardisation
KPI	Key Performance Indicator
LUI	Logical User Interface
MDT	Microsoft Desirability Toolkit
NASA-TLX	NASA Task Load Index
NCG	Next Generation Scania
PUI	Physical User Interface
SART	Situation Awareness Rating Scale
SHERPA	Systematic Human Error Reduction and Prediction Approach
SUS	System Usability Scale
USD	User-Centred Design
UX	User Experience
VR	Virtual Reality

1 Introduction

This chapter covers the background of the chosen subject and gives an introduction to Scania CV AB and their work. Furthermore, the aim and delimitations of the thesis work are described.

1.1 Background

The complexity of interactive systems in the automotive industry has increased during recent years. The drivers are provided with new interactive systems and functions such as infotainment systems, driver assistance systems and different types of comfort systems. These contribute to improved safety and efficiency as well as personal comfort within the driving situation (Kern & Schmidt, 2009). Since these types of products require users to interact with them it is important to work with interaction design. By working with interaction design, negative aspects that users can experience when interacting with a product can be reduced and positive aspects such as pleasure can be improved (Rogers, Sharp, & Preece, 2002, pp. 1-2). In order to ensure that the users are able to use the system and that they like it, it is important to evaluate the system during the design process. Moreover, evaluations allow the designers to make sure that the users have a pleasurable experience when interacting with the system (Rogers, Sharp, & Preece, 2002, pp. 319-323).

1.2 Scania CV AB

Scania CV AB is a major Swedish manufacturer of transport solutions. The company was founded in 1891 and is today part of the German automotive company Volkswagen Truck & Bus. Research and Development activities take place at the headquarters in Södertälje, Sweden with branches in both Brazil and India while production is spread out over Europe, Latin American and Asia.

The transport solutions include trucks (long-haulage, urban applications and construction), buses (city bus, intercity bus and coach) and engines. Scania also offers services including workshop services, tailor-made maintenance, financing and insurance solutions, driver training and coaching as well as services for support and management of customer operations. Scania provides a modular system which enables a tailor-made product. Instead of providing a fixed amount of vehicle models the modular system enables different components and parts of the vehicle to be combined in numerous ways (Scania CV AB, 2018a).

1.2.1 Property-driven development

Since 1997, the organisational structure of the Research and Development department at Scania has been based on product properties, which refer to different areas of responsibility in the vehicle such as Driver Environment or Load Carrying Capacity. This has provided Scania with a greater understanding of the customer needs related to these specific areas. In the last few years the properties have therefore become a central part of the product development process.

The product properties are further divided into 28 vehicle properties responsible for different elements within the vehicle. For example the product property Driver Environment is further divided into Driver Vehicle Interaction, Physical Vehicle Ergonomics, Ride and Vibration Comfort, Climate Comfort, Infotainment and Connectivity as well as Acoustics. The organisational structure of the Research and Development department is represented in Figure 1.1.

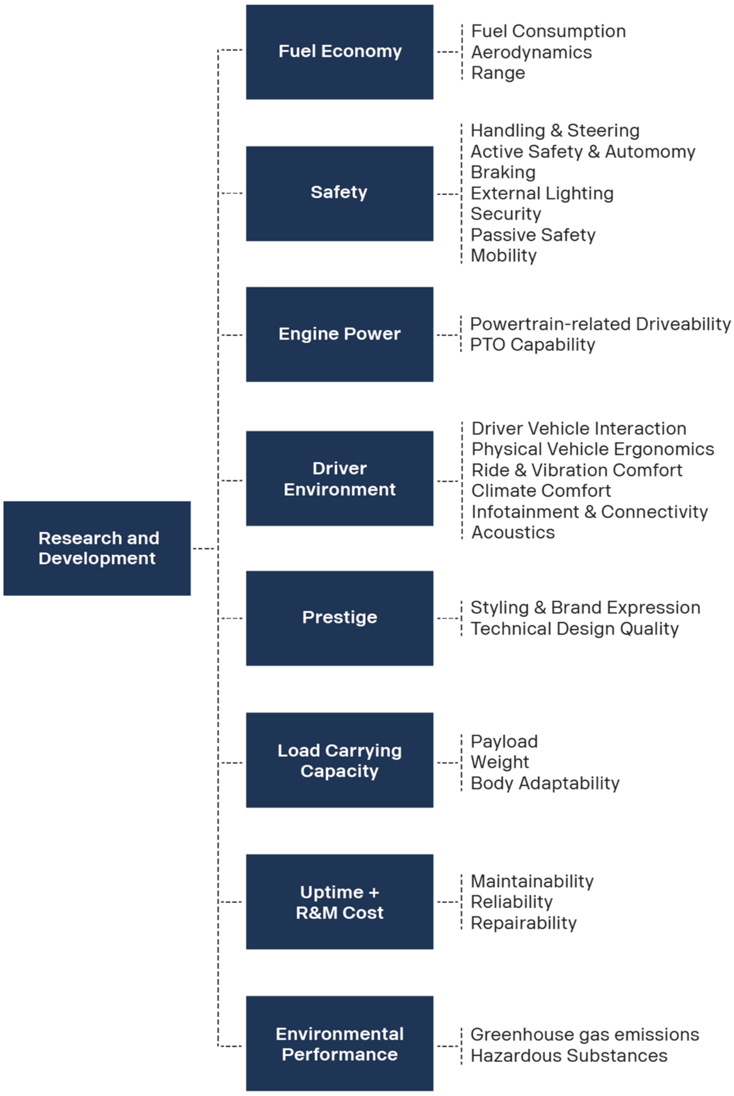


Figure 1.1 Organisational structure of the Research and Development department.

1.2.2 Driver Vehicle Interaction

The thesis work was conducted at the department of the vehicle property Driver Vehicle Interaction (DVI). The department of DVI works with the interaction experience within the truck cab as well as the bus drivers' area. The following definition was established by Scania in order to describe DVI's responsibilities.

"The vehicle property Driver Vehicle Interaction includes all interaction experienced between the driver and the vehicle regarding being able to perceive and understand visual, auditory, and haptic information, and being able to use this information to make appropriate decisions and take correct actions. This includes, for example, perceiving and understanding the visual information contained within menu structures, using this information to make correct decisions, and handling the controls correctly. The interaction experienced between the driver and the truck should be consistent with human factors principles. Physical driver ergonomics, service staff and production ergonomics is not included" (Persson, 2013).

The organisational structure at the department of DVI consists of a long-term, mid-term and short-term team as well as a simulation and evaluation team (shown in Figure 1.2). The long-term team focuses on visions through studying user needs and creating concepts for the future, the mid-term team focuses on strategy and works with interaction concepts for the next generation vehicles and the short-term team focuses on implementation of interaction concepts and updates on current vehicles. The simulation and evaluation team supports research projects and product development for short and long time perspectives. All teams follow a User-Centred Design (UCD) process, which is further described in chapter 2.3.

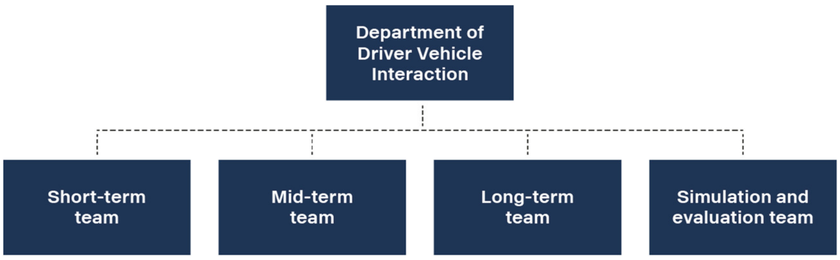


Figure 1.2 Organisational structure of the department of Driver Vehicle Interaction.

1.2.3 Sub-properties of Driver Vehicle Interaction

DVI is further divided into sub-properties with the purpose to establish requirement specifications in the beginning of projects, follow up requirements during projects and execute validations on complete products or systems in the end of projects. The sub-properties are based on the experience of the system. The selected sub-properties are a combination of established definitions in the area of usability and User Experience (UX) and were compiled during a workshop at the department of DVI. The sub-properties are briefly described below and illustrated in Figure 1.3, and further described in chapter 2.4.

Affordance is a sub-property that provides cues on how a Human Machine Interface (HMI) should be used, while *feedback* gives information about what action has been done and what has been accomplished. *Consistency* refers to that similar operations and elements should be used for achieving similar tasks and *flexibility* refers to the transformation of content for diverse user needs. Another sub-property is *responsiveness* which concerns the system's timely response to user input. *Error handling* refers to both error prevention and handling of errors, while *task focus* concerns the ability for users to keep their task in focus. The final sub-property is *pleasurability* which concerns the user's internal state, the characteristics of the designed system and the context within which the interaction occurs.

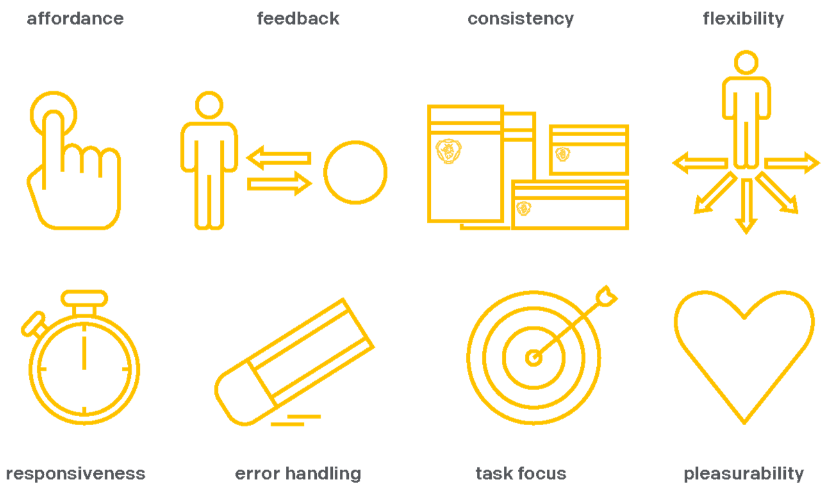


Figure 1.3 Sub-properties of the department of Driver Vehicle Interaction.

1.2.4 Usability evaluation at the department of Driver Vehicle Interaction

The department of DVI performs user tests throughout the design process in order to get user insights and enable re-design. They usually perform tests on low fidelity prototypes such as paper prototypes or high fidelity prototypes such as interactive digital prototypes. The tests are performed on truck or bus drivers that are working at the test drive laboratory at Scania.

The simulation and evaluation team at the department of DVI conducts extensive user tests in order to validate the performance and usability of prototypes in the driving simulator or Virtual Reality (VR). The department of DVI strives to integrate the driver simulator and VR into their design processes to a greater extent. The team uses eye tracking, driving performance and physiological measures in order to evaluate the prototypes. Moreover, they use around 20 questionnaires for evaluating the interaction with prototypes. For example they use the Driver Activity Workload Index (DALI), NASA Task Load Index (NASA-TLX), Situation Awareness Rating Scale (SART), System Usability Scale (SUS) and Adapted Subjective Workload Assessment Technique (ASWAT).

The department of DVI has also outsourced validations of products or systems in the past. In 2015, Semcon did a usability evaluation of the truck that was about to be launched, Next Generation Scania (NCG), in comparison with the former truck. The aim was to discover potential interaction issues and gain usability and user insights. The study was a summative validation, conducted by experts, on the new truck along with comparative user tests on both the former and the new truck. Several evaluation methods such as Hierarchical Task Analysis (HTA), Enhanced Cognitive Walkthrough (ECW), Predictive User Error Analysis (PUEA) and Heuristic Evaluation (HE) were used during the summative validation. The user tests were carried out on 33 participants that were asked to perform several tasks and answer questionnaires. The test leaders also noted task time, completion rates and observed usability issues (Bergmark, Hallila, Karlsson, & Nilsson, 2015).

The department of DVI does not use a standard procedure for all evaluations conducted. This results in that all evaluations are executed differently, depending on the project, and give different results. Standardized evaluation frameworks are therefore needed in order to improve their design process.

1.3 Aim

The purpose of the thesis work was to develop an evaluation method for the driver vehicle interaction experience, which could be used as part of the product development process at the department of DVI. Important criteria of the sub-properties laid the foundation of the development. Aspects such as organisation, resources and current development processes were also taken into account throughout the thesis work. The following research questions were answered.

- *What* factors of Driver Vehicle Interaction should be evaluated?
- *How* should the factors of Driver Vehicle Interaction be evaluated?
- *When* in the design process should the evaluation be carried out?

In addition, a case study was conducted where selected parts of a user interface in an ongoing development project at the department of DVI were evaluated based on the evaluation method.

Primarily, the method will work as an aid to evaluate the driver vehicle interaction experience in ongoing projects in order to improve the internal work at the department of DVI. In the long run the evaluation method can work as a foundation for establishing requirement specifications in the beginning of projects and executing validations on complete products or systems in the end of projects. Secondarily, the result will be used to strengthen the arguments in project status reports towards the management. This is important since other properties tend to be prioritized due to objective arguments, such as reduced fuel consumption or weight.

1.4 Delimitations

The thesis work was delimited to primarily investigate expert evaluation methods rather than user tests when developing the evaluation method. This way the evaluation method can be carried out by the employees at the department of DVI since they are experts within the subject. The department of DVI were interested in investigating expert evaluations since they are less time and resource consuming than user tests. Expert evaluations can work as a complement to the user tests already performed at the department of DVI.

The development of the evaluation method focused on three parts of the instrument panel; the Instrument Cluster (ICL), the Infotainment System (AUS) and the buttons in the steering wheel (see Figure 1.4). The major parts of the ICL are indicator lamps and symbols, measuring instruments and driving related information while the AUS includes media, radio, traffic information, camera, navigation and other services. The buttons in the steering wheel include radio control, ICL navigation buttons, downhill speed control, cruise control and adaptive cruise control. These parts were used as a basis for the development and testing of the evaluation method and were evaluated in the case study. The selected parts are of importance since the key interaction types developed by the DVI department are represented through them and they exist in ongoing development projects. Other interaction types, such as smartphone applications and control rooms, which are also part of ongoing projects are not represented.

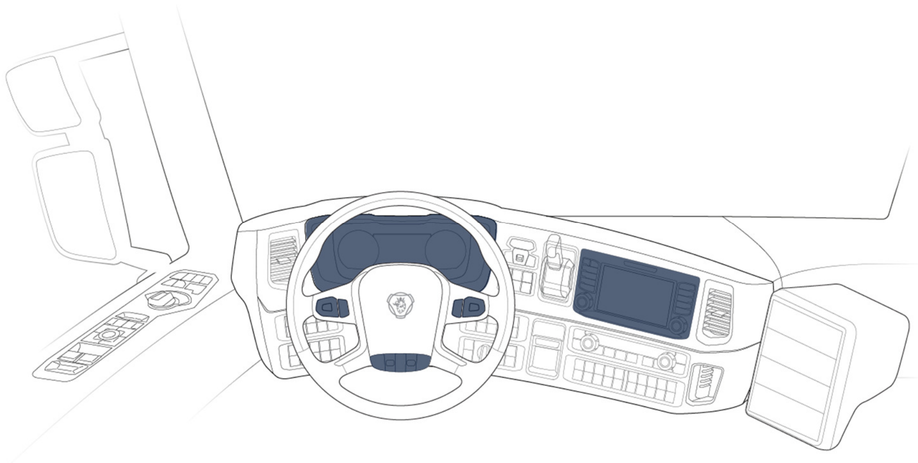


Figure 1.4 Selected parts of the instrument panel.

2 Theory

This chapter covers the theory that laid the foundation for the analysis. The terms usability, User Experience and User-Centred Design as well as the sub-properties are further described. Furthermore, evaluation methods considered relevant for the thesis work are also described.

2.1 Usability

There is no universal definition of usability (Gray & Salzman, 1998) but various authors and institutes have established different important factors to consider when designing for usability. Some significant definitions of usability are mentioned below.

Nigel Bevan, in cooperation with the International Organization of Standardization (ISO) (1998), has defined usability as the "extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use". Effectiveness, efficiency and satisfaction are referred to as usability measures. Effectiveness is the ability to complete a task, while efficiency is the effort required to complete a task (Tullis & Albert, 2013, p. 7). Satisfaction is when the user feels comfortable and has a positive attitude towards the use of a product (International Organization of Standardization (ISO), 1998).

Jacob Nielsen (1993a, pp. 25-26) describes usability as how well users can use a specific functionality within a system in relation to the system performance. The main usability attributes are learnability, efficiency, memorability, errors and satisfaction. The system should be easy to learn, efficient to use, easy to remember, pleasant to use and have a low error rate. He has also set principles for interaction design through ten heuristics (Nielsen, 1995a); visibility of system status; match between system and the real world; user control and freedom; consistency and standards; error prevention; recognition rather than recall; flexibility and efficiency of use; aesthetic and minimalist design; help users recognize, diagnose and recover from errors; help and documentation.

Don Norman (2013, pp. 71-73) has defined seven stages of action through questions that are linked to the use of a product; "What do I want to accomplish? What are the alternative action sequences? What action can I do now? How do I do it? What happened? What does it mean? Is this okay? Have I accomplished my goal?" In order to provide the user with the right information to answer the questions seven fundamental design principles should be considered; discoverability, feedback, conceptual model, affordances, signifiers, mappings and constraints.

Shneiderman (2016) has defined eight user interface principles which he refers to as "The Eight Golden rules of Interaction Design"; strive for consistency, seek universal usability, offer informative feedback, design dialogs to yield closure, prevent errors, permit easy reversal of actions, keep users in control and reduce short-term memory load.

2.2 User Experience

There are several diverse definitions of User Experience (UX). Nevertheless, a study has shown that researchers, educators and practitioners have gained a common agreement on that UX is dynamic, subjective and depends on the context (Roto, Hassenzahl, Vermeeren, Law, & Kort, 2009). Some significant definitions of UX are mentioned below.

According to ISO (2010a) UX is the user's perception and response from using a product and includes parameters such as emotions, beliefs and psychological responses that occur before, after and throughout the use. It is a result of system performance and functionality as well as other aspects such as brand image and presentation. The context of use is also an important parameter. Usability in terms of the user's personal goals includes aspects of UX that can be assessed by defining criteria.

Sascha Mahlke (2007) describes UX as three components; perception of instrumental qualities, perception of non-instrumental qualities and emotional user reaction. Perception of instrumental qualities concern usefulness and usability while perception of non-instrumental qualities concern aesthetic, symbolic and motivational aspects. Emotional user reactions include areas such as subjective feelings, psychological reactions and behavioural tendencies.

In comparison to usability UX looks at the user's entire interaction with the product, including thoughts, feelings and perceptions (Tullis & Albert, 2013, p. 5). Virpi Roto (2007) states that the subjective measures of usability are part of UX, while objective measures such as time to completion are irrelevant. Consequently, only some parts of efficiency are included in UX. He proposes four elements to describe the measurable characteristics of UX; utility, usability, enjoyment and pride.

2.3 User-Centred Design

User-Centred Design (UCD) is a design process aimed for designing usable products through the use of suitable techniques, processes, methods and procedures. The design process should evolve around actual user behaviour rather than desirable user behaviour (Rubin & Chisnell, 2008, pp. 12-13). ISO (1999) states that "Human-centred systems support users and motivate them to learn. The benefits can include increased productivity, enhanced quality of work, reduction in support and training costs, and improve user satisfaction."

According to ISO (2010a) UCD is based on six principles; design should be based on the understanding of the users, tasks and environments; users should be involved throughout the design process; design should be driven by user-centred evaluations; design processes should be iterative; design teams should include different skills and perspectives and design should consider the whole UX. Consequently, there is a relation between UCD and UX and since usability is considered as part of UX there is a link between the three terms.

2.4 Sub-properties of Driver Vehicle Interaction

A clear definition of the sub-properties was not established at the department of DVI. Thus, this was done in the beginning of the thesis work. The definitions of the eight sub-properties, used throughout the thesis work, were based on previous work compiled at the department of DVI as well as different established definitions by significant authors. Relating terms, such as mapping and signifiers as well as effectiveness, efficiency and satisfaction, were also considered and included in the definitions in order to acknowledge all aspects of the driver vehicle experience. The sub-properties are defined below.

2.4.1 **Affordance**

Affordance is the relationship between the properties of a system and the capabilities of a user. It determines how a system can be used through provided cues of how to operate. If the user can figure out what actions are possible without instructions, such as signs or labels, the system has perceived affordance (Norman, 2013, pp. 10-13). An example of an affordance is the multi-function button on the Infotainment System (AUS), which supports the motion of turning and pushing due to its design (see Figure 2.1). When turning the button the volume is adjusted and when pushing it the AUS is turned on or off.

Another important term to consider is *signifier,* which is closely related to affordance. While affordances give cues of possible actions, signifiers are signs of what the function is for (Norman, 2013, pp. 13-19). An example of a signifier is the symbol next to the multi-function button, since it explains that the button turns the AUS on or off (see Figure 2.1).

Figure 2.1 Multi-function button on the Infotainment System (Lindh, 2016a).

Figure 2.2 Navigation buttons in the steering wheel (Lindh, 2016b).

It is also important to look at the relationship between two sets of elements within a system in terms of layout. This is called *mapping* (Norman, 2013, pp. 20-22). An example of mapping is the navigation buttons in the steering wheel that controls the Instrument Cluster (ICL) (see Figure 2.2). When pressing up, down, left or right the ICL menu navigates accordingly.

In order to create affordance in physical controls it is important to adapt the size, shape, material, motion and location to the functionality. Moreover, the force and sensitivity should be proportional to the action (Olsson, 2015). For example, touch-screen buttons should give clickability cues through the use of depth and colour, which consequently give the impression of a real button (Lucaites, Fletcher, & Pyle, 2017). In addition, it is important to follow standard interaction elements that are recognizable to the user (Olsson, 2015).

2.4.2 **Feedback**

Feedback communicates the result of an action done by the user and provides the user with information about the status of the system. The user should not only be informed that something has happened, but also what has happened and how to control it (Norman, 2013, p. 23). Depending on the situation feedback can be informative, alerting or warning. This can be communicated through visual, auditory or tactile feedback (Dehghanpour, 2015). An example of feedback is the green lamp on the right side of the menu page that indicates the location when navigating within the menu structure of the ICL (see Figure 2.3). In addition, feedback provides information regarding the system status. For example, this mainly concerns errors trough different types of warnings and alerts presented in the ICL such as the alert that informs the user that the fuel is running out.

The system should respond immediately to all types of user input and the feedback given should be informative. If these requirements are not fulfilled the feedback can be more disturbing than useful. The feedback can also be overwhelming and it is therefore important that the feedback is proportional to the importance of the information given (Norman, 2013, pp. 23-25).

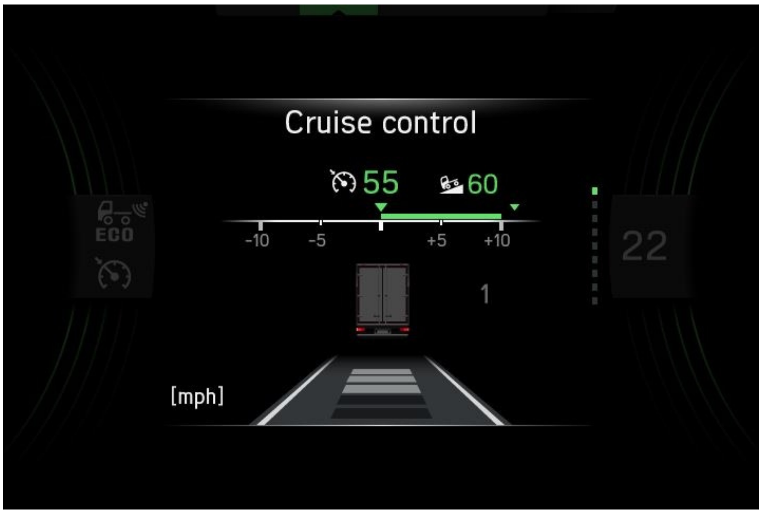


Figure 2.3 Navigation within menu structure (Scania CV AB, 2018b).

2.4.3 **Consistency**

Consistency simplifies the use by standardizing the way information is presented and eliminating varying representations of the same functions (Shneiderman, 2016). A limited amount of appearances and behaviours reduces the cognitive load. There are two types of consistency, internal and external, which both are important. Internal consistency covers both the aesthetics and the functionality of the system. External consistency covers consistency to similar systems, systems within the platform and the real world (Cao, Zieba, Stryjewski, & Ellis, 2015). An example of internal consistency is that the same font is used on the different displays within the vehicle (see Figure 2.4). An example of external consistency is the menu bar in the top of the AUS display, which users are familiar with from other devices such as smartphones (see Figure 2.5). Consistency can be achieved by using for example colours, fonts, font sizes, styles, element sizes and spacing consistently (Medium, 2016).

Figure 2.4 Consistency in displays (Wink, 2016a).

Figure 2.5 Menu bar in Infotainment System (Scania CV AB, 2017).

2.4.4 **Flexibility**

Flexibility simplifies the interaction for both experienced and inexperienced users by offering different paths through the system. *Accelerators* can speed up the interaction for more experienced users (Nielsen, 1995a). An example of flexibility through accelerators is the mute button in the steering wheel (see Figure 2.6).

Flexibility enables *customization* of the system, in terms of content, layout and functions, in order to meet all kinds of user needs (Norman, 2013, pp. 246-247). This contributes to user control of the overall UX (Schade, 2016a). An example of flexibility through customization is the ability to select which function the user wants to see in the programmable fields of the ICL (see Figure 2.7). *Personalization* is when the system identifies the user and delivers user specific content, experiences or functionalities (Schade, 2016b).

Customization functions should be available and visible to the user as well as easy to use. The user should be able to change previous selections. By layering the customization functions the complexity of the customization is reduced. This can be done through progressive disclosure which means that only commonly used options are shown. Unnecessary customization functions that do not serve a purpose should be eliminated (Schade, 2016a).

Figure 2.6 Mute button in the steering wheel (Wink, 2016b).

Figure 2.7 Programmable fields in the Instrument Cluster (Scania CV AB, 2018b).

2.4.5 **Responsiveness**

Responsiveness refers to the system's response to a user input, in terms of time. There are three general time limits for the response time of systems; 0.1 second, 1 second and 10 seconds. The response time of 0.1 second is the limit for making the user feel like the system is reacting instantaneously, 1 second is the limit for the user's flow of thought not to be interrupted and 10 seconds is the limit for keeping the user's attention (Nielsen, 1993b). Specifically for feedback it is important that the system response is immediate as mentioned in chapter 2.4.2.

2.4.6 **Error handling**

Correct *error handling* allows the user to discover and recover from wrong actions. This is created through clear feedback (Norman, 2013, p. 67). There are two types of input errors; slips and mistakes. Slips are when the user has the right goal in mind but accidently performs the wrong action while mistakes occur when the user has the wrong goal in mind and performs the wrong action (Norman, 2013, pp. 171-172). An example of error handling is the possibility to go back to the previous view within the menu structure of the ICL with the backward button in the steering wheel (see Figure 2.8).

As mentioned in chapter 2.4.2, feedback also concerns warnings and alerts which consequently play an important role in error handling. The users must be informed and able to solve the problems caused in the vehicle. An example of error handling is the warning in the ICL that tells the user that there is an airbag functional fault that needs to be checked at next workshop visit (see Figure 2.9).

Errors can be prevented through the use of appropriate affordances, signifiers, mapping and constraints (Norman, 2013, p. 67).

Figure 2.8 Backward button in the steering wheel (Boman, 2017).

Figure 2.9 Warning in the Instrument Cluster (Scania CV AB, 2018b).

2.4.7 **Task focus**

In practice, *task focus* is the ability of users to complete tasks without exposure to irrelevant information. Task focus refers to effectiveness and efficiency (Krupenia, 2015). *Effectiveness* is the ability to complete a task, while *efficiency* is the effort required to complete the task (Tullis & Albert, 2013, p. 7).

2.4.8 **Pleasurability**

Pleasurability is part of the UX resulting from the use of a product or system (Roto, 2007). In order to create pleasurable experiences it is necessary to have ha holistic approach and consider emotional aspects rather than the usability of the product. This is of importance since it makes the product successful (Gkouskos & Chen, 2012). However, the department of DVI considers usability as an important component that contributes to pleasurability. As mentioned in chapter 2.2, several significant authors have stated that usability is part of the UX and consequently pleasurability can be seen as a more holistic term that covers a wide range of aspects of UX.

The department of DVI also believes that the term satisfaction is important when considering pleasurability. *Satisfaction* is when the user feels comfortable and has a positive attitude towards the use of a product (International Organization of Standardization (ISO), 1998).

Garvin (1987) proposes an approach for analysing the quality of a system in order to be competitive on the market and to please the users. The approach consists of eight dimensions; performance, features, reliability, conformance, durability, serviceability, aesthetics and perceived quality.

2.5 Evaluation methods

It is necessary to evaluate the usability in product development processes in order to understand user needs and improve products. It is also important to establish requirements at the beginning of the development process and later validate if the requirements have been fulfilled. The reason for evaluating usability is to achieve acceptable effectiveness, efficiency and satisfaction on a user interface level. By considering the user’s satisfaction as well as pleasure UX can be measured (Bevan, 2008).

The evaluation methods can be divided into analytical and empirical methods. Analytical methods are generally executed in early stages of the design process in order to determine which concepts fulfil the criteria. They can predict objective performance while interacting with the system, such as potential errors. Analytical evaluations can be executed on simulations based on paper prototypes or computer-based prototypes (Stanton & Harvey, 2013, pp. 45-51). The evaluation can be executed by experts making judgements while simulating usage based on design principles (Ko, n.d.). After performing analytical methods the prototype can be assessed by the use of empirical methods that measure actual performance when users interact with prototypes (Stanton & Harvey, 2013, pp. 45-51).

One can also divide evaluation methods into summative and formative methods. Formative methods are used to identify usability or UX problems, understand user needs and to refine requirements. This is achieved through a small number of user or expert evaluations. Consequently, the data is not reliable in terms of user performance and satisfaction and should only be used as design feedback. Formative methods are therefore carried out in the beginning and during the design process. Summative methods are either used to compare different products or systems or validate if requirements established in the beginning of a development process have been fulfilled. Summative methods are therefore carried out in the end of design processes. In order to make accurate conclusions the measures need to be valid and reliable. This is achieved through numerous user tests carried out in a realistic context-of-use. A statistical assessment should also be carried out in order to confirm the validity of the test results (Bevan, 2008).

Moreover, the evaluation output can be divided into either quantitative or qualitative data. Quantitative data provide numerical data which can be put into different categories or units. The data can be represented through rating scales or binary answers (McLeod, 2017). Qualitative data provide insights and understanding of specific problems and can be represented through spoken or written statements (Surbhi, 2016).

Several evaluation methods were investigated but due to time limitations not all evaluation methods could be studied. The methods described below are those considered of interest for the thesis work. These are divided into analytical and empirical approaches. The analytical approaches are also considered as formative since they should be performed during the design process. The empirical methods can be used both during and in the end of a design process and can therefore be considered as both formative and summative. Depending on evaluation method the output can be quantitative or qualitative.

2.5.1 **Analytical approaches**

Hierarchical Task Analysis

The Hierarchical Task Analysis (HTA) is used to describe tasks in terms of a hierarchy of goals, sub-goals, operations and plans. The upper levels of the hierarchy include goals and sub-goals while the bottom level describes what operations needs to be accomplished in order to reach the sub-goals. Furthermore, the plans indicate how the goals are achieved. The method is often required, or can be applied as a basis, for other evaluation methods that evolve around tasks. It provides a great understanding of tasks, in terms of cognition, and is time and resource efficient since it does not require extensive training (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 46-54; Stanton, 2006).

Cognitive Walkthrough

The original Cognitive Walkthrough (CW) is a usability inspection method which focuses on the ease of learning when using a product or system for the first time. Other usability attributes such as functionality are correlated to the ease of learning (Wharton, Rieman, Lewis, & Polson, 1994). More specifically, the method focuses on the user activities and the user's goals and previous knowledge when performing specific tasks. The evaluation is performed by experts who consider a set of tasks of a product or system that the end user needs to accomplish (Magnusson, Rassmus-Gröhn, Tollmar, & Deaner, 2009). This can either be done by a group of experts, such as designers, engineers and representatives from other organizational units, or by the individual designer. The aim of the method is to discover usability issues that enables re-design. The method can be conducted during the design process and can be performed on various types of prototypes (Wharton, Rieman, Lewis, & Polson, 1994).

The CW has been modified several times and the most current one is the third version. In order to perform the third version of the CW the following steps need to be considered; defining inputs, set up evaluators, walking through the action sequence of each task, recording critical information and revising the usability issues in order to re-design. The inputs required when conducting a CW are product representation, task scenarios including the context of use, a sequence of actions that the user needs to accomplish in order to complete the tasks as well as information about the end user. In general, the task scenarios should be based on an analysis of the market, user needs, concept testing and requirements as well as the core functionality of the system. For a simple system, it is possible to evaluate all important task scenarios while for a more complex system it is beneficial to evaluate a task scenario that has proven to be problematic (Wharton, Rieman, Lewis, & Polson, 1994). A HTA can be used as a basis for the evaluation in order to describe the sequence of each task (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, p. 95). The evaluators walk through the action sequence of each task and evaluate it according to the following set of questions; "Will the user try to achieve the right effect? Will the user notice that the correct action is available? Will the user associate the correct action with the effect they are trying to achieve? If the correct action is performed, will the user see that progress is being made toward solution of their task?". The questions should be answered by motivation stories for user failure or success (Wharton, Rieman, Lewis, & Polson, 1994).

The original CW has also been presented through different variants such as the Heuristic Walkthrough, Norman's Cognitive Walkthrough and the Enhanced Cognitive Walkthrough (Mahatody, Sagar, & Kolski, 2010). The Enhanced

Cognitive Walkthrough (ECW) is an improvement of the third version of the CW. The method is used to detect potential usability issues of a system and offers a more extensive presentation of analysis than the original CW since it does not only evaluate the sequence of each task but the tasks and functions in general. This is accomplished through two levels of questions that are related to either the functions provided by the system or the operations performed by the user. The ECW also includes a better description of the usability issues by scoring the tasks as well as the failure and success together with a categorisation of the problem types (Bligård & Osvalder, 2013).

Inspection-based evaluation

Heuristics and checklists together with usability guidance, industry practices and standards can be used during expert evaluations. The overall term for this kind of expert evaluation is Inspection-based evaluation. Inspections are user-centred evaluations, which are part of the UCD process. They are essential elements in order to collect information about user needs, allocate weaknesses and strengths of the concept, assess whether the requirements have been fulfilled and establish baselines or compare design concepts. The main advantage of all types of Inspection-based evaluations is that they are more simple and cost-effective than user tests. In addition, they can cover a wide range of users and tasks. Still, some problems cannot always be discovered by experts. As earlier mentioned, it is therefore important that the inspections are performed by experts with prior knowledge of the users and the ability to step into the role as users (International Organization of Standardization (ISO), 2010a).

Heuristic Evaluation

Heuristic Evaluation (HE) is a method used to assess the usability, errors, mental workload and design quality of a user interface (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 439-443). A small group of evaluators individually examine the interface against a set of usability principles called heuristics. After the inspection the evaluators communicate and gather their findings. The output of a HE is a list of usability problems with detailed comments, made by the evaluators, which describe the specific problems. When different evaluators examine the user interface the evaluation will cover a wide range of problems, since different people encounter different usability problems (Nielsen, 1995b). This requires trained experts with the right experience and knowledge. The main advantages of HE is that it is time and cost effective and can be executed in an early stage of the design process (U.S. Department of Health & Human Services). The method can be applied to various prototypes such as paper-based diagrams, mock-ups and functional devices (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, p. 440).

One of the most renowned list of heuristics, mentioned in chapter 2.1, was developed by Nielsen (1995a). Since then technology has advanced and the heuristics are no longer applicable to all products or systems. In many cases they are too general to examine products where the technology is constantly developing, such as mobile phones. Thus, designers should develop their own set of principles to evaluate their specific designs. It is recommended to have between five and ten principles where Nielsen's set of heuristics can be incorporated into the specific area. After establishing the list of specified heuristics the group of evaluators, that should be experts within the domain, can be selected. The evaluators should also be briefed of the assignment before initiating the first evaluation phase. In this phase the evaluators can use the product or system and get a feel for the interaction as well as choose specific elements to evaluate (Wong, n.d.). Alternatively, a set of tasks based on a HTA can be used. The tasks work as a foundation when evaluating against the defined list of heuristics (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 439-440). During the second evaluation phase the evaluators use the heuristics to evaluate the chosen elements or tasks. The final step of the evaluation is a debriefing session where the evaluators gather their findings and make a list of problems (Wong, n.d.).

Checklists

Checklists are used to evaluate the usability and design of a specific user interface. The method involves evaluators inspecting interface elements against a checklist which consists of a set of predefined criteria. The evaluators should have experience or knowledge of the product or system in order to conduct an accurate evaluation. The method can be conducted at any time during the design process and can be performed on various types of prototypes, such as paper drawings and finished products. The method is generic and can be used in any domain. However, the evaluator must decide what kind of checklist should be considered for the specific domain. An existing checklist can be used if it is appropriate or be adapted to suit the product or system. It is also possible to create a new one if that is necessary (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 436-439)

Heo, Ham, Park, Song and Yoon (2009) propose checklists based on four types of evaluations; task-based evaluation, Logical User Interface (LUI) evaluation, Graphical User Interface (GUI) evaluation and Physical User Interface (PUI) evaluation. The firstly mentioned is task dependent while the latter are task independent. The advantage of this division is that it supports both modular and holistic evaluations of a user interface. When conducting a task dependent checklist at least one of the three interface areas will be covered and no interface area can be completely interpreted without task situations. Examples of elements that can be evaluated without the use of tasks are fonts and icons as well as wording or labelling.

There are several checklists available for evaluating interfaces, both for in-vehicle systems and other application areas such as websites and mobile phones, as well as proposed guidelines for creating checklists, which are describe below.

The Transport Research Laboratory propose a checklist for in-vehicle information systems with the aim to provide a structured approach for assessing the interface design and identifying where development and measurements need to be done. The checklist includes questions regarding installation, information presentation, interaction, system behaviour and information about the system. The questions can be answered true, false or not applicable. Based on these answers the evaluator responds if there is no, minor, serious or not applicable concerns about the design (Stevens & Cynk, 2011). Similarly, the European Commission (2008) recommend a set of principles regarding safety and efficiency for the use of in-vehicle information and communication systems. The Japan Automobile Manufacturers Association (2004) has also developed a list of guidelines for in-vehicle display systems based on basic concepts, scope, installation of display systems, functions of display systems, display system operation while vehicle in motion and the presentation of information to users.

Oztekin, Kong and Uysal (2010) propose a checklist for eLearning systems that evaluates both quality and usability. The checklist consists of questions that can be assessed on a 5-point Likert Scale. Measures of effectiveness, efficiency and satisfaction are considered results of the checklist dimensions. Checklist questions were determined based on an analysis of different checklist approaches. Another checklist, used for the evaluation of mobile phones, consists of questions that are assessed on a 7-point Likert Scale and based on task scenarios. The checklist questions where developed based on user interface elements arranged in a hierarchal structure that are linked to usability principles (Gu Ji, Ho Park, Lee, & Hwan Yun, 2006).

Systematic Human Error Reduction and Prediction Approach

The Systematic Human Error Reduction and Prediction Approach (SHERPA) is a method used to predict potential human or design initiated errors. The method is conducted by an evaluator and can be applied in any domain. A HTA is required as an input since the evaluation evolves around tasks. During the evaluation each bottom level operation of the HTA is analysed. The operations and related errors are classified into task types and error types and are then further described. More specifically, the operations are classified into action, checking, retrieval, communication or selection tasks with related error types. For example, an action task is pressing a button and a related error type can be that the operation is executed on the wrong button. In addition, the evaluator determines and describes the consequences and recovery potentials associated with the errors as well as the probability and criticality. The evaluation results in possible error reduction strategies (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 143-151).

Critical Path Analysis

Critical Path Analysis (CPA) is used to estimate duration of activities that concern several tasks. The first step of the evaluation is to analyse the tasks and divide them into sub-tasks that can be ordered based on time. The sub-tasks can be based on a HTA. These tasks should be assigned a modality; visual, auditory, cognition, manual or speech tasks, which can be represented in a time and modality based diagram. The timings of the tasks should then be established based on research. When this is done the evaluator can calculate the time to perform the whole task. It is also possible to calculate the earliest start time and the latest finish time (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 507-512).

Role-playing and Empathic modelling

Role-playing is a method where the designers take on the role as users in a realistic task scenario. This way the designers can get a sense of empathy as well as understand challenges and experiences. However, this method should be used during concept generation and in early iterations (Martin & Hanington, 2012, p. 148). A similar method is empathic modelling, where designers can experience different types of impairments through using for example sight impairment glasses or a wheelchair. This way the designer, or evaluator, can encounter problems that were not previously discovered (Magnusson, Rassmus-Gröhn, Tollmar, & Deaner, 2009, pp. 27-28).

2.5.2 **Empirical approaches**

Usability tests

Usability tests are carried out by observing users while they walk through given tasks or scenarios. The tasks or scenarios should represent typical and actual end-user goals (Martin & Hanington, 2012, pp. 194-195). There is a wide range of usability test approaches with different objectives and requirements in terms of time and resources. The tests can be divided into formal experiments and informal iterative tests. The latter are quick tests with low complexity and are conducted in order to expose usability problems in industrial product development projects. The informal tests can be conducted relatively early in the process and require specified test objectives, a group of representative users and a representation of environment. The participants are observed and interviewed when using, or reviewing, a prototype in order to collect both qualitative and quantitative performance data. In addition, recommendations of improvements are given (Rubin & Chisnell, 2008, pp. 19-25).

In connection to usability tests the participants can report their experience through various methods such as rating scales, lists of attributes and open-ended questions. Commonly used rating scales are Likert Scales, where the participants rate their agreement to a statement on a five- or seven-point scale, and Semantic Differential Scales, where the participants rate the experience on a scale with opposite adjectives at each end. The reporting can be implemented after each task or after the whole session. The advantage of reporting after each task is that one can gain insights of usability problems related to specific tasks. Post session ratings measure the perceived usability of the overall interaction. Examples of questionnaires that are commonly used after the interaction session are System Usability Scale (SUS) and

Questionnaire for User Interface Satisfaction (QUIS) (Tullis & Albert, 2013, pp. 122-137).

Microsoft Desirability Toolkit

The Microsoft Desirability Toolkit (MDT) is a method used to measure satisfaction in order to uncover user reaction and attitude towards a system. The method includes 118 product reaction cards with words that describe user's personal reactions to a specific system, such as *usable*, *complex* or *reliable*. The method is conducted together with participants in the end of usability tests. The participants are asked to sort through the product reaction cards and pick five that best match their personal reaction. During a post-test interview the participants are then asked to explain why they chose those cards. As an alternative the participants can go through a simpler checklist of adjectives and select as many words as they like. The participants are then asked to circle five words that they find most relevant and then explain why they chose those words during the post-test interview. It is possible to customize the list by changing to more relevant words that better suit the system to be evaluated (Travis, 2009).

NASA Task Load Index

NASA Task Load Index (NASA-TLX) is an evaluation tool that subjectively assesses workload when interacting with systems. The questionnaire is based on six factors; *mental demand, physical demand, temporal demand, performance, effort* and *frustration*. Through a weighted average of the results it is possible to derive a global score of the workload (National Aeronautics and Space Administration, 2017). The subjective rating should be executed during or after users have performed tasks. Since the method requires an operational system it should preferably be used on almost complete systems or by the use of simulations (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, p. 303).

Driving Activity Load Index

The Driving Activity Load Index (DALI) is a subjective method that evaluates six workload factors; *effort of attention, visual demand, auditory demand, temporal demand, interference* and *situational stress*. It is a relative measure where workload in different situations can be compared. The evaluation is based on a rating scale and a weighting procedure in order to get a global score. The method is a revised version of NASA-TLX designed for a driving context. It can be used to evaluate driving workload during the execution of secondary tasks and complex driving situations (Pauzié, 2008).

Physiological measures

Physiological measures can be used to evaluate the user's emotions when interacting with a system (Tullis & Albert, 2013, p. 163). It is also a way to evaluate the mental workload. There are many different types of measures such as eye movements, heart rate variance, blink rate and skin conductance. All methods require special measuring equipment, which can be expensive, and expertise in how to use it. The evaluation methods should be implemented on various participants when performing tasks and it is therefore recommended to use a HTA as a basis (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 314-316).

Performance measures

Performance measures are based on user behaviours while interacting with a product or system. The performance can be measured by *task success*, which is the most commonly used performance metric. It measures the effectiveness based on a set of tasks and is often executed in lab-based usability tests. It is possible to measure the binary success where the users either complete the task or not, or levels of success where the evaluator defines the various levels. Another way of evaluating the performance is by measuring *time on task*. This is considered a good way to measure efficiency. *Efficiency* can also be evaluated by studying the amount of effort put into completing a task by measuring the number of clicks or the lostness (Tullis & Albert, 2013, pp. 63-89). Different types of performance measures can also be used to evaluate the mental workload from using a system. The performance measures used for this type of evaluation should be selected for the specific purpose of the evaluation. An example of performance measures used to evaluate the mental workload when performing primary and secondary tasks in a driving environment are speed, lateral position and headway (Stanton, Salmon, Walker, Baber, & Jenkins, 2005, pp. 301-310).

3 Method

This chapter covers the methodology used for developing an evaluation method for driver vehicle interaction. Furthermore, methods used for data gathering and analysis as well as tests required in the methodology are described.

3.1 Methodology

The design of the methodology used in the thesis work was based on one general approach of how to measure things that have not yet been measured and several specific approaches of how to measure usability in particular projects, both for vehicles and within other areas.

Emardson (2015) proposes a general approach and states that the main steps of measuring new parameters are defining what needs to be measured, breaking it down to the main measurable components, identifying evaluation methods and instruments, confirming the validation, identifying measuring scales and defining a reference.

Stanton and Harvey (2013) propose a methodology of how to conduct a usability evaluation of in-vehicle information systems in a driver context. After a need for development of a product or system has been specified, high level usability factors considered important to evaluate, such as efficiency or flexibility, should be chosen. These can be chosen from usability definitions stated by significant authors, such as those mentioned in chapter 2.1. The high level usability factors are then linked to a context-of-use based on the equipment, the users involved, the environment within which the product or system is used and the tasks performed. This results in contextual factors, which are especially important for vehicles since the context is closely linked to safety. The high level usability factors can then be examined in relation to the contextual factors which results in a list of usability criteria that covers the aspects for the specific product or system. These can later be translated into Key Performance Indicators (KPIs) that describe how the usability criteria should be measured. When the interaction is formed and prototypes are created suitable evaluation methods, covering the identified KPIs, are specified.

A similar methodology, for evaluating the usability of mobile phones, was developed by Heo, Ham, Park, Song and Yoon (2009). They propose a hierarchical model where the usability factors are divided into four abstraction levels. The top level is the overall usability of the product or system while the second level consists of usability indicators that indicate the general usability of any product or system. These can be based on usability definitions stated by significant authors. The third level is the criteria, connected to the specific product or system, which can be measured through the use of different evaluation methods. These are based on usability property, found in the lowest level of the hierarchy, which are observable properties of the specific product or system.

Moreover, this way of approaching the problem was confirmed by Lars-Ola Bligård (2018). He underlined the importance of analysing the context of use and the definitions of the sub-properties in relation to the driver vehicle interaction at Scania.

These four approaches contributed to the development of an adapted methodology used for this thesis work. The main insights from Emardson's approach was that the factors to be measured must be established and divided into smaller measurable components. The main insights from Stanton and Harvey was that the context-of-use needs to be taken into consideration and that existing evaluation methods can be used as basis. The main steps of the adapted methodology used in the thesis work came from Stanton and Harvey. Heo, Ham, Park, Song and Yoon as well as Bligård confirmed the methodology since they propose similar methodologies. According to Stanton and Harvey (2013) as well as Emardson (2015) it is important to adapt the steps in the methodology to the specific objectives of the evaluation. Thus, the methodology used in the thesis work was based on the department of DVI's specific requirements. In contrast to the methodology proposed by Stanton and Harvey the methodology used also included aspects of User Experience (UX).

3.2 Implementation

As mentioned, the different approaches on how to develop an evaluation method contributed to the methodology designed for the thesis work. The steps of the methodology are summarized in Figure 3.1 and described below. The complete implementation of the methodology is presented in chapter 4.

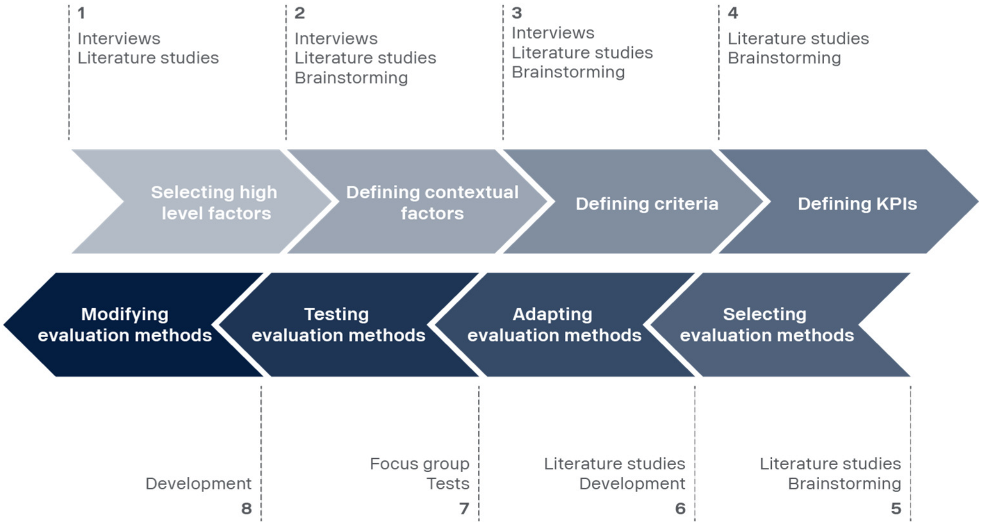


Figure 3.1 Thesis work methodology.

Firstly, literature studies and interviews with employees at the department of Driver Vehicle Interaction (DVI) were conducted to get a deeper understanding of the terminology used within usability, UX and User-Centred Design (UCD). This included research on the sub-properties and other high level factors established by different significant authors. This resulted in the selection of high level factors. In order to understand the context-of-use, within the specific driver vehicle environment, interviews with employees at the DVI department were conducted. In addition, literature studies were carried out to complement the interviews. The findings were analysed during a brainstorming session, which contributed to the definition of contextual factors in relation to the previously selected high level factors.

To set the criteria based on the high level factors for the specific context-of-use interviews with employees were conducted. Brainstorming sessions were included in the interviews in order to connect the high level factors to the contextual factors. Important criteria were defined and analysed, which resulted in a list of criteria. The list, together with research findings, worked as a basis when establishing the list of KPIs. This was done during brainstorming sessions. The KPIs could then be connected to suitable evaluation methods that were found through literature studies. These were adapted to suit Scania and the specific KPIs. The adaption resulted in frameworks and recommendations for each evaluation method. This was done through researching, developing and walking through the evaluation methods iteratively. Moreover, the product representations, evaluation data and evaluators for each method were defined.

Some of the adapted evaluation methods were tested on employees at the department of DVI and a focus group with employees was held in order to develop the test protocol. The tests resulted in further modifications and improvement of the proposed evaluation methods.

The steps of the methodology were carried out in an iterative way since new aspects were considered along the way. For example the criteria and KPIs were revised several times when selecting evaluation methods and the evaluation methods were further adapted after they were tested. In the end of the thesis work a case study was conducted during which the evaluation methods were implemented on an ongoing project at the department of DVI.

An overview of what has been accomplished during the thesis work, including the inputs and the outputs of the thesis work, is described in Figure 3.2.

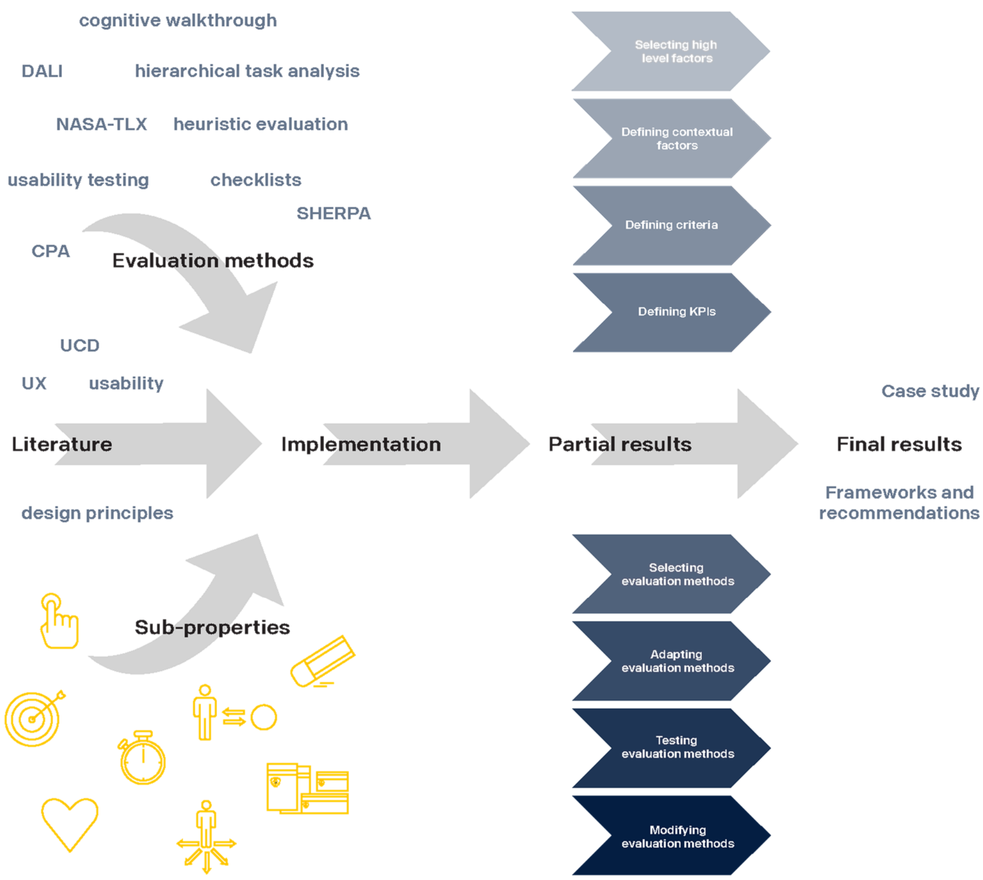


Figure 3.2 Overview of thesis work.

3.3 Planning

The thesis work was carried out over an estimated time frame of 20 weeks. The results were presented to the Department of Design Sciences at the Faculty of Engineering, Lund University, at the end of week 20.

In the beginning of the thesis work a project plan was created with the aim to describe the content, delimitations and methodology of the thesis work. A Gantt chart was also created in order to allocate time and resources of planned activities. This Gantt chart was used as a guidance throughout the project. During the project some activates required more or less time than estimated and in the end of the thesis work an updated Gantt chart was created to describe the actual process. The original and the updated Gantt chart are shown in Figure 3.3 and 3.4.

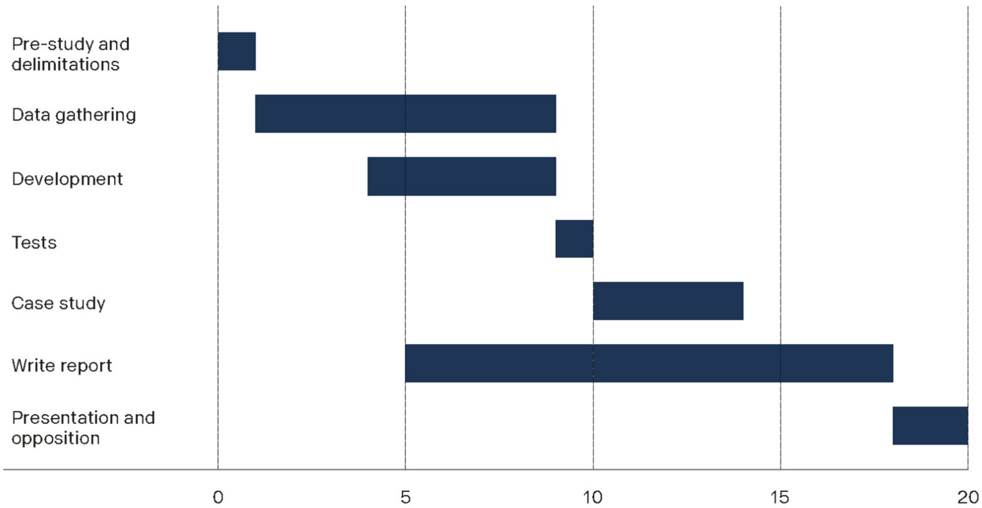


Figure 3.3 Original Gantt chart.

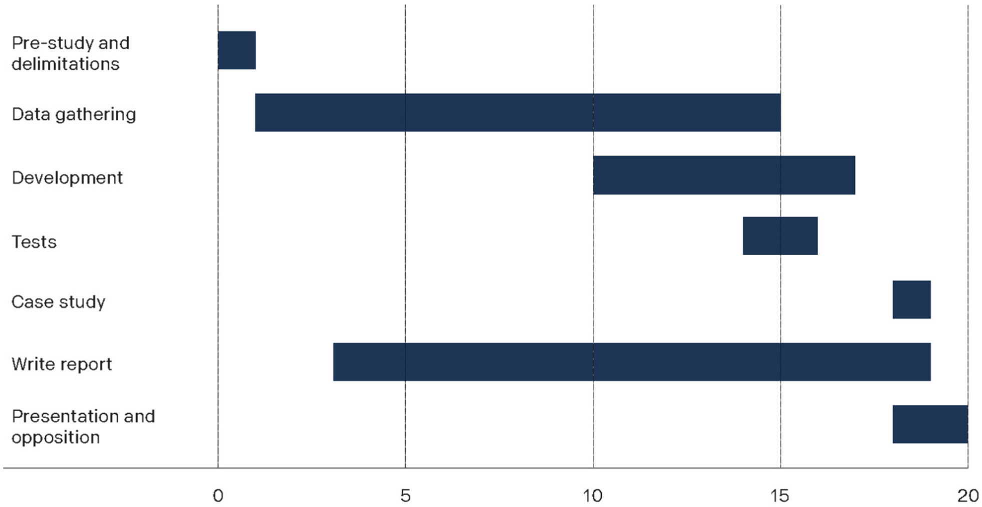


Figure 3.4 Updated Gantt chart.

3.4 Data gathering

3.4.1 **Literature studies**

Literature studies were carried out in the beginning of the thesis work in order to gain insights on the subject and get a deeper understanding of the terminology used within usability, UX and UCD. Literature studies were also carried out in an iterative way during the process when needed, as mentioned in chapter 3.2. Usability, UX, the sub-properties and possible evaluation methods as well as methodology on how to develop an evaluation method were investigated with the purpose to define and analyse the state of art. Furthermore, these insights led to conclusions which laid the foundation of the development of the evaluation methods.

Different types of sources such as books, e-books, articles from papers, reports and webpages related to the subject were studied. Furthermore, standards concerning the subject, mainly from the International Organization of Standardisation (ISO), were studied. The sources were selected based on research and recommendations given by the thesis work supervisors. Books and e-books written by significant authors within the subject, such as Norman (2013), Nielsen (1993a) and Stanton (2005), were chosen. Articles and report were chosen based on their relevance and importance for the subject. These were found through reliable search engines such as LUBsearch and Google Scholar. Reliable and suitable webpages that were established by known authors, such as Nielsen Norman Group, were chosen. Access to standards was provided by Scania.

3.4.2 **Interviews**

Interviews are a qualitative research method to collect direct information of the participant's personal experience, opinions and insights. During interviews it is possible to see and interpret the participant's reactions, expressions and body language (Martin & Hanington, 2012, p. 102). The main advantage of interviews, compared to questionnaires, is that the questions can be adapted to the situation. As a result, one will get a broader view and capture other dimensions and perspectives as well as situations and courses of events (Ahrne & Svensson, 2015, p. 38).

There are two types of interviews; structured and unstructured. Structured interviews follow pre-established questions while unstructured interviews follow a conversational format and allow flexibility. The sessions of unstructured interviews are guided by a researcher and are often based on a set of topics. Since unstructured interviews are conversational they are more comfortable for the participants, which is an advantage compared to structured interviews (Martin & Hanington, 2012, p. 102). Semi-structured interviews combine pre-established questions with topics that arise during the interview. This kind of interview allows the participant to mention important information that was not considered by the researcher (Sweeney & Pritchard, n.d.). Furthermore, the interview questions can either be closed or open. Closed questions result in short or even single-worded predetermined answers while open questions allow more detailed answers (Magnusson, Rassmus-Gröhn, Tollmar, & Deaner, 2009).

In the beginning of the thesis work unstructured interviews with several employees were carried out to get an overview of Scania, the department of DVI, property-driven development and the sub-properties. This also provided an understanding of their professional opinions of the subject. During the process semi-structured interviews were conducted with six employees, two representatives from each team at the department of DVI. Each representative was interviewed once during two steps of the methodology; defining contextual factors and defining criteria. Thus, the participants were involved throughout the project and did not need to be re-informed about the progress which was beneficial for the thesis work. Open questions were used throughout all interviews conducted during the thesis work. The interviews are further described in chapter 4.2 and 4.3 and the interview protocols of each step can be found in Appendix A.

3.4.3 **Focus groups**

The focus group method is an effective way to explore and get an understanding of people's thoughts and perspectives within a specific topic. (Ahrne & Svensson, 2015, pp. 81-82). As it is a peer setting the participants can share experiences, needs and perceptions (Martin & Hanington, 2012, pp. 92-93).

A focus group was held with nine employees from the department of DVI in order to investigate possible task scenarios that could be used when evaluating prototypes during the tests. The focus group is further described in chapter 4.7.2.

3.5 Data analysis and representation

3.5.1 **Brainstorming webs, tree diagrams and flow diagrams**

Brainstorming is traditionally used to generate creative ideas within a group. The method can also be used to graphically organize or visually present knowledge as well as to define new relationships between components or generate alternative solutions. As a result, brainstorming sessions challenge old ways of thinking. There are three common visualization methods used when brainstorming; brainstorming webs, tree diagrams and flow diagrams. Brainstorming webs are used to define and map out characteristics, facts and ideas related to a central concept, tree diagrams can be used to classify and connect ideas in a hierarchy, while flow diagrams represent processes (Martin & Hanington, 2012, pp. 22-23).

Brainstorming was used to generate ideas based on literature and interview findings when defining contextual factors, criteria and KPIs as well as when selecting evaluation methods. Brainstorming webs were used when analysing the data collected during the interviews related to defining contextual factors. The contextual factors were then represented in a tree diagram. Tree diagrams and flow diagrams were also used to illustrate the organizational structures at Scania (see chapter 1.2.1 and 1.2.2) and the methodology of the thesis work (see chapter 3.1).

3.5.2 **Affinity diagrams**

Affinity diagramming can be used as a tool to compile results from research. The captured insights or requirements are written down on individual notes, which are clustered in appropriate themes (Martin & Hanington, 2012, pp. 12-13). Affinity diagrams were conducted in order to analyse and organise the data collected during the interviews related to defining criteria and KPIs (see chapter 4.3 and 4.4).

3.6 Tests

Usability tests can be used to evaluate a product, expose issues and eliminate frustration for users. The method includes observations of participants using the product as well as collection of qualitative and quantitative data. The test participants should be representatives for the actual end users and the tests should be executed in an actual context-of-use (Rubin & Chisnell, 2008, pp. 21-25). Before the actual tests pilot tests should be carried out in order to enable re-design of the test. The test participant should be someone who has prior knowledge of the project but has not been part of the specific design (Magnusson, Rassmus-Gröhn, Tollmar, & Deaner, 2009, p. 9).

Six usability tests were conducted in order find potential errors in the developed evaluation methods. This was done through participants evaluating prototypes based on the developed evaluation methods. The test participants were six employees at the department of DVI, which are the actual end users. During the tests the participants were asked to answer questions about the evaluation methods. The answers together with observations made during the tests resulted in qualitative data which contributed to the development of the final evaluation methods. The tests were spread out over two weeks and each test had a duration of two hours. One pilot test was conducted before the final tests. The pilot test participant was the thesis work supervisor at Scania. He had been involved in the project but not in the specific design of the evaluation methods, which are the requirements for pilot test participants. The tests are further described in chapter 4.7 and the test protocols can be found in Appendix B.

4 Implementation

This chapter covers the implementation of the methodology and describes partial results of each of the steps; selecting high level factors, defining contextual factors, defining criteria, defining Key Performance Indicators, selecting evaluation methods, adapting evaluation methods and modifying evaluation methods. Furthermore, the chapter includes a case study during which the evaluation methods were implemented on an ongoing project at the department of Driver Vehicle Interaction.

4.1 Selecting high level factors

As mentioned in chapter 3.1, high level factors that are considered important to evaluate should be chosen from usability definitions stated by significant authors (Stanton & Harvey, 2013, p. 28). The analysis was based on research findings and consultation with employees at the department of DVI.

Pleasurability, task focus as well as *flexibility* were selected as a high level factors. All the eight sub-properties established by the Driver-Vehicle Interaction (DVI) department, mentioned in chapter 1.2.3 and 2.4, were analysed when selecting the high level factors and their relevance for the specific area was studied. A majority of the employees mentioned that the sub-properties are closely related and dependent. This resulted in a hierarchy, shown in Figure 4.1, where only the top level sub-properties were chosen as high level factors. The reason was that they include aspects of the other sub-properties. Furthermore, other definitions stated by significant authors, such as learnability, were discussed, but since they were not considered by the DVI department they were not selected as high level factors.

Pleasurability was chosen as a high level factor since it is a part of the overall UX, as mentioned in chapter 2.4.8. The other sub-properties affect the usability of a product or system and since usability is part of UX, as mentioned in chapter 2.2, they contribute to pleasurability.

Error handling, affordance, feedback, responsiveness and consistency as well as flexibility contribute to the task focus. Hence, task focus was selected as a high level factor. Error handling was considered as contribution criteria to task focus since when users are able to discover and recover from wrong actions they are able to perform tasks more effectively. This can be achieved by the appropriate use of design principles such as affordances, signifiers, mapping and feedback. Affordances and feedback are design principles that provide information needed to successfully use a product or system (Norman, 2013, pp. 67, 71-73). Without these principles the user will not be able to complete a task. For the same reason responsiveness, which is a user interface design rule (Nielsen, 2010), and consistency, which is an interface principle (Shneiderman, 2016), are important aspects of task focus. Flexibility in terms of accelerators are closely linked to efficiency of use (Nielsen, 1995a) and was therefore also considered as part of the task focus.

However, flexibility was chosen as a high level factor since customization and personalization as part of flexibility include aspects that do not necessarily contribute to task focus. Instead, these aspects evoke emotional reactions as part of the UX, as mentioned in chapter 2.4.4.

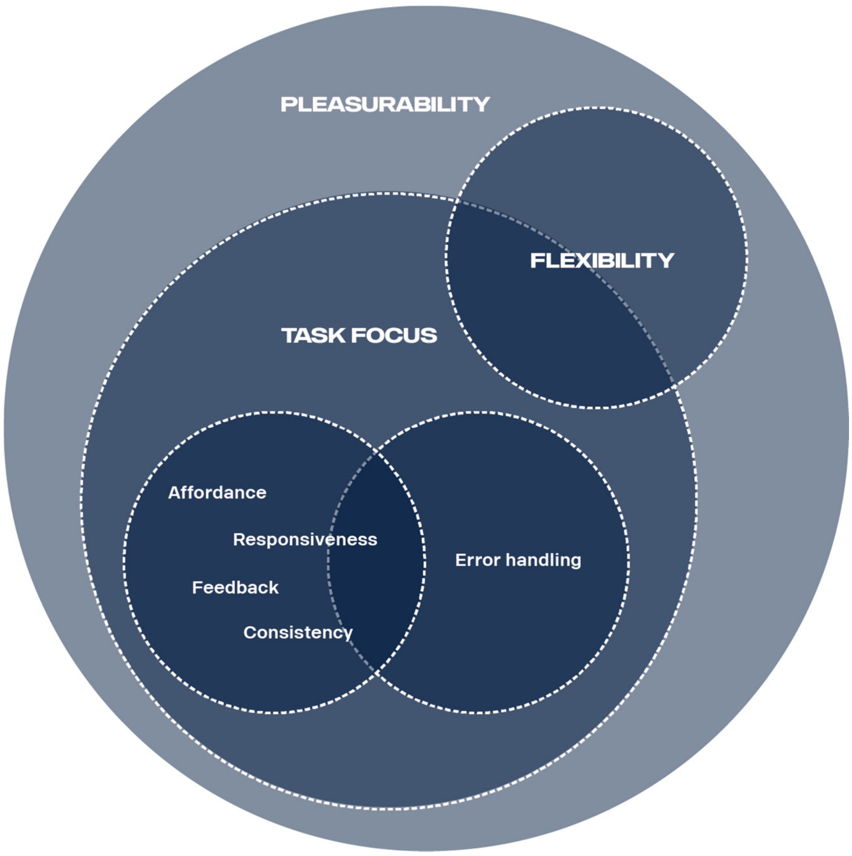


Figure 4.1 Sub-property hierarchy.

4.2 Defining contextual factors

As mentioned in chapter 3.1, it is important to take the context into consideration in order to execute a reliable usability evaluation (Thomas & Bevan, 1996). The contextual factors can be defined through a context of use analysis. The analysis should consist of detailed information about the users, the physical and social environments within which a product or system is used, the tasks and the equipment that is used (International Organization of Standardization (ISO), 1998). Data needed for defining contextual factors can be gathered through interviews, workshops, surveys, site visits, artefact analysis, focus groups, observational studies, and contextual inquiry (Context of Use Analysis, 2009).

The analysis was based on the data gathered through semi-structured interviews with six employees, two representatives from each team at the department of DVI. Three of the participants were interaction designers at Scania while the other three were UX design consultants currently working at Scania. Their expertise within the use of the product was considered as reliable input for the definition of the contextual factors. The contextual factors mentioned by Stanton and Harvey (2013, pp. 34-35) were also considered when defining the contextual factors due to the fact that they regard a similar context of use.

The interviews were conducted gradually during a time period of one week, where each interview had a duration of one hour. During the interviews the participants were first asked to describe their perception of the context-of-use for driver vehicle interaction in general. Later on the participants were asked to describe the context-of-use related to tasks, equipment and users as well as physical and social environment. The participants were asked to map out their answers on a paper to visualise their thoughts. The questions allowed open and detailed answers. The interview protocol can be found in Appendix A.1.

The participants discussed similar topics such as different types of users and users with different levels of experience. Furthermore, the participants mentioned that varying driving conditions and tasks should be considered and that tasks are often executed in a dual task environment. They mentioned that there is a conflict between the primary task of driving and the secondary tasks of administrative work and providing personal comfort. Stanton and Harvey (2013, pp. 34-35) also underline the importance of considering a dual task environment. Some participants also mentioned that the vehicle is a unique environment since it does not only concern driving but also living and working. Brainstorming webs were used when analysing the data gathered through the interviews. The results from the interviews contributed to the contextual factors, which are summarized in Figure 4.2 and described below. The contextual factors were revised several times through consulting employees.

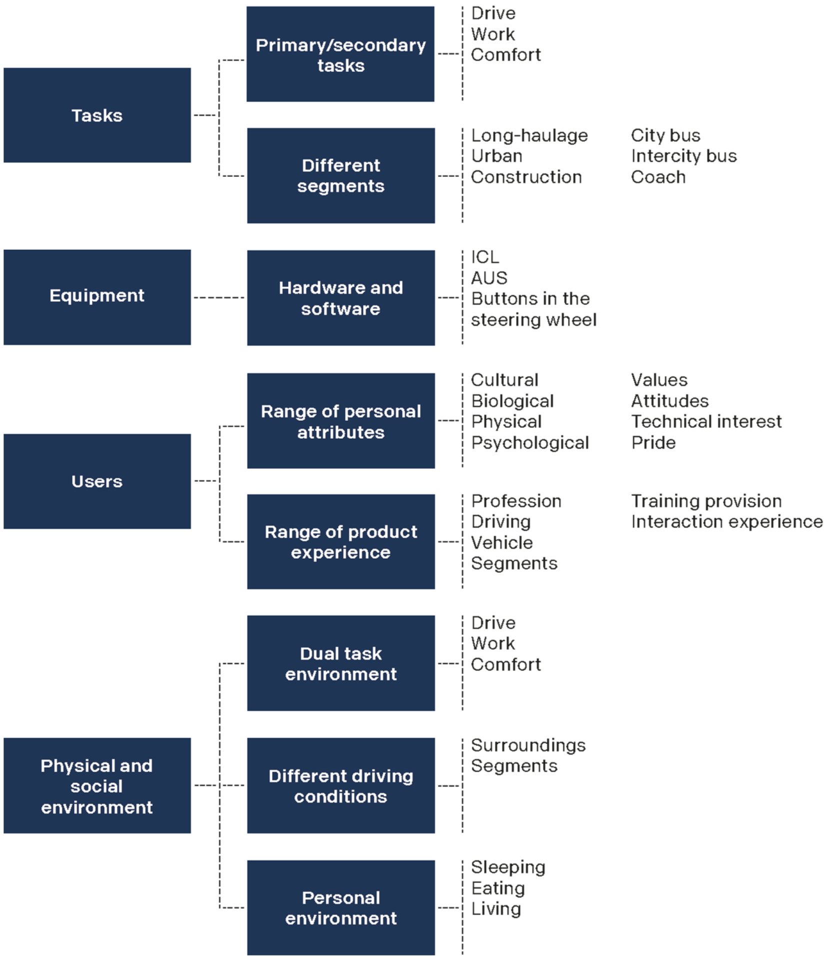


Figure 4.2 Contextual factors.

4.2.1 Tasks

Primary and secondary tasks

The interaction consists of two different types of tasks; primary and secondary tasks. Primary tasks are related to driving while secondary tasks refer to working and providing personal comfort. The use of the Instrument Cluster (ICL) and related buttons in the steering wheel (ICL navigation buttons, speed control and cruise control) are primary tasks, while the use of the Infotainment System (AUS) and related radio control buttons are linked to secondary tasks. It is important to acknowledge that some tasks, such as interacting with the speedometer, tachometer or the fuel gauge, are more important than other tasks. The importance can vary depending on the user and the segment.

Different segments

The truck segments are long-haulage, urban applications and construction, while bus can be divided into city bus, intercity bus and coach. These can be further divided into applications. Examples of urban applications are distribution trucks and dust carts. The tasks vary a lot depending on the segment and the specific application area. Consequently, the different segments require different functions and the interaction is affected. For example, distribution truck drivers exit the vehicle frequently and interact with customers. Furthermore, the vehicles within individual segments can vary due to modularization.

4.2.2 **Equipment**

Hardware and software

The equipment in terms of the ICL, AUS and buttons in the steering wheel consist of a combination of hardware and software such as buttons, knobs, switches, displays and touchscreens.

4.2.3 **Users**

Range of personal attributes

The interaction is affected by the range of personal attributes and it is therefore important to take this context into consideration in order to design for all kinds of users. The users have a variety of cultural, biological, physical and psychological characteristics. This includes aspects such as geography, religion, physical and psychological disabilities, age and gender. This means that the users have different preferences and requirements regarding for example colour, symbols and layout. In addition the users have different personal values, preferences, attitudes and technical interests toward the product. This affects the experience of the product or system. Another important aspect is pride, regarding both the vehicle and the profession, since driving these kinds of vehicles can be part of a personal identity or lifestyle.

Range of product experience

Another important context regarding users is their range of interaction experience of the product and its specific systems. Some users have been driving Scania vehicles professionally for several years and are consequently familiar with the product, while others have used other brands. Another aspect that affects the interaction is the range of training provision. Some road carriers provide driver training, while other users do not even have access to the driver's manual. Since a majority of the users are professional drivers they are experts in terms of driving and their specific segment. Moreover, the users have expert knowledge of the vehicle and its capabilities.

4.2.4 **Physical and social environments**

Dual task environment

Within the vehicle, tasks are often executed in a dual task environment. This is a unique context that must be taken into consideration since it concerns safety. When interacting with the system there is a conflict between the primary task of driving and the secondary tasks of administrative work and providing personal comfort. Working includes for example paper work and interacting with road carriers or customers through the AUS, while providing personal comfort includes for example listening to music and talking on the phone. Consequently, the use of the ICL and related buttons in the steering wheel (ICL navigation buttons, speed control and cruise control) are connected to primary tasks, while the use of the AUS and related buttons are secondary tasks. The tasks can either be executed in sequence or in parallel.

The negative impact of the secondary tasks is confirmed by Stanton and Harvey (2013, pp. 29-31). They state that the usability of the secondary function is of high importance since it might affect the driving performance and consequently the safety in the vehicle. The reason is that depending on how much effort is put into the secondary task the effort put into the primary task decreases. Consequently, the effectiveness of the primary task is negatively affected.

Different driving conditions

The interaction is affected by the surroundings in terms of weather conditions, road conditions and traffic environment along with other factors such as time of day, location and traffic regulations. Furthermore, the surroundings related to the different segments affect the interaction. Depending on if it is a long-haulage, construction or distribution truck, or city bus, intercity bus or coach, the conditions can differ. Construction trucks can for example be affected by disturbing noise. The interaction is also affected by the cab layout which can vary a lot due to modularization. Combinations of the mentioned conditions can cause difficult driving situations that highly affect the interaction within the cab.

Personal environment

Another area specific context is that some users spend long hours or live in the vehicle. The interaction experience is affected since there are other user expectations for this type of environment than an ordinary work or driving environment. The personal environment includes eating and sleeping as well as other personal activities that are executed off-duty, both outside and inside the vehicle.

4.3 Defining criteria

The definition of criteria contributes to the development of an evaluation for a specific product or system. To define the criteria for a product or system it is essential to consider the contextual factors in detail. The high level factors, described in chapter 4.1, can be matched to the contextual factors, described in 4.2, in order to define criteria within a specific context (Stanton & Harvey, 2013, pp. 34-36).

To define the criteria semi-structured interviews with the same participants as when defining contextual factors were conducted. The usability criteria mentioned by Stanton and Harvey (2013, pp. 34-35) were also studied since they were considered relevant.

The interviews were conducted gradually during a time period of one week, where each interview had a duration of one hour. During the interviews the participants were first asked to describe which high level factors they believed had the greatest impact on each specific contextual factor. When the participants had matched the high level factors with the contextual factors they were asked to describe important criteria related to these. The participants were asked to map out their answers on a paper to visualise their thoughts. The questions allowed open and detailed answers. The interview protocol can be found in Appendix A.2.

The participants mentioned that all the high level factors; pleasurability, task focus and flexibility, affect the different contextual factors to some extent. The data gathered through the interviews was analysed through affinity diagramming and translated into criteria. The results from the interviews contributed to the criteria, which are described below and summarized in Table 4.1.

Regarding pleasurability, it is important that a wide range of users feel satisfied both during short and long term usage. They should have a positive experience in the personal environment, which means while spending time in the vehicle when not working, as well as through all channels that concern the product. Moreover, the users should experience personal fulfilment through the use of the product or system. In order to maintain task focus, the users should be efficient and able to complete tasks regardless of personal attributes and product experience. This should be applicable in varying driving conditions and in a dual task environment. The users should be able to choose their own course of action depending on their personal attributes and product experience. In addition, customization depending on the segment as well as personal attributes and product experience should be possible. The product or system should also support personalization.

Table 4.1 Criteria.

criteria
pleasurability
Short and long term satisfaction
Satisfaction for users with different personal attributes
Positive experience in personal environment
Positive experience through all channels that concern the product
Personal fulfillment through use of system
task focus
Efficiency regardless of product experience and personal attributes
Ability to complete tasks regardless of product experience and personal atttributes
Efficiency in varying driving conditions
Ability to complete tasks regardless of driving condition
Efficiency in a dual task environment
Ability to complete tasks in a dual task environment
flexibility
Ability to choose your course of action in relation to personal attributes and product experience
Ability to customize for different segments
Ability to customize for a wide range of personal attributes and product experience
Personalization for dual task environment and varying driving conditions
Peronalization for personal environment, user attributes and product experience

4.4 Defining Key Performance Indicators

Key Performance Indicators (KPIs) are measurable values, either objective or subjective, that describe the criteria in detail and the relation to the evaluation needed for measuring them (Stanton & Harvey, 2013, p. 35). The criteria was analysed through affinity diagramming and brainstorming was executed, which resulted in 23 KPIs. The KPIs in relation to the criteria are described below and summarized in Table 4.2, 4.3 and 4.4.

4.4.1 **KPIs for pleasurability**

When using a system for the first time the user should have a positive attitude, which must remain throughout the entire product lifecycle. Moreover, the system quality should match the user's expectations regarding technology, material, etc. A positive attitude towards the system should apply to a wide range of users with different personal attributes such as cultural, biological, physical and psychological characteristics as well as different values, preferences, attitudes and technical interests toward the product.

The user should feel emotionally attached to the product even when not driving or working, both outside the vehicle and when spending time in the vehicle. Furthermore, the user should maintain a product liking through all channels that concern the product. For example, through commercials, websites, stores and maintenance services. This KPI is marked grey in Table 4.2 since this is an indicator that cannot entirely be effected by the work at the department of DVI. It would be misleading to evaluate this on account for DVI. Nevertheless, this is an important indicator that affects the UX. In the future this indicator might be taken into account since the area of responsibility of the department of DVI is constantly evolving and emerging. For example, smartphone applications are now part of the work at the department of DVI. In the future DVI might be responsible for more channels that can contribute to the overall UX. Furthermore, the user should feel like an expert and in control of the system as well as safe when using the system, which will contribute to the UX.

Table 4.2 KPIs for pleasurability.

criteria	key performance indicators
Short and long term satisfaction	Users should have a positive attitude towards the use of the system, on initial use and throughout the usage System quality should match user expectations throughout the usage
Satisfaction for users with different personal attributes	All types of users should have a positive attitude towards the use of the system
Positive experience in personal environment	Users should have an emotional attachment to the product while not working
Positive experience through all channels that concern the product	Users should maintain a product liking through all channels such as marketing, sales, use of product and maintenance
Personal fulfillment through use of system	Users should feel like experts and in control of the system Users should feel safe when using the system

4.4.2 **KPIs for task focus**

The system should assist the user so that all kinds of users can maintain a low mental workload when performing task. Moreover, all users, regardless of product experience and personal attributes, should be able to interact with the system. It is important that the users understand the functions and their impact on the vehicle. Experienced users should understand all functions while novice users should understand the most essential functions, especially those that regard safety. This is important since all users do not have access to the Driver's Manual or training. Furthermore, all users should be able to recover from wrong actions.

Task times should be minimized regardless of the driving conditions and the task structure and functions must support safety in difficult driving conditions. Tasks that affect safety should be prioritized and only important information should be shown in difficult driving conditions. It is important that users can distinguish and localise functions in varying driving conditions. In a dual task environment the secondary task times should be minimized and the users should maintain a low mental workload. Moreover, the performance of the primary task should not be compromised by secondary tasks. For example, the driving should not be negatively affected by the use of the AUS.

Table 4.3 KPIs for task focus.

criteria	key performance indicators
Efficiency regardless of product experience and personal attributes	All types of users should maintain a low mental workload while performing tasks
Ability to complete tasks regardless of product experience and personal atttributes	All types of users should be able to interact with the system as well as understand functions and their impact Novice users should be able to handle essential functions Users should be able to recover from wrong actions
Efficiency in varying driving conditions	Task times should be minimized regardless of driving conditions
Ability to complete tasks regardless of driving condition	Task structure and functions should support safety in difficult driving conditions Users should be able to distinguish and localise functions in varying driving conditions
Efficiency in a dual task environment	Secondary task times should be minimized Users should maintain a low mental workload while performing secondary tasks
Ability to complete tasks in a dual task environment	The performance of primary tasks should not be compromised by secondary tasks

4.4.3 **KPIs for flexibility**

Different paths through the system should be provided in order to match personal attributes and product experience. This means that experienced users can use an accelerator while novice users can use the ordinary path to complete the same task. The user should be able to customize the system before use, which means that the functions can be customized for specific segments. Moreover, all types of users, with different personal attributes and product experience, should be able to customize the system according to their own preferences, regarding settings in the vehicle. The tasks should not be negatively affected by customization.

The system should allow personalization through recognizing the user's needs in difficult driving situations. Moreover, all kinds of personal needs, for users with different personal attributes and product experience, should be recognized by the system. These KPIs are marked grey in Table 4.4 since they have not been fully implemented at Scania and can therefore not be evaluated. Still, this is something that the department of DVI strives for and might be possible to evaluate in the future.

Table 4.4 KPIs for flexibility.

criteria	key performance indicators
Ability to choose your course of action in relation to personal attributes and product experience	Various paths through the system should be provided to the user that match personal attributes and product experience
Ability to customize for different segments	Users should be able to customize the system for their specific segment tasks before use
Ability to customize for a wide range of personal attributes and product experience	All types of users should be able to customize functions during use Tasks should not be negatively affected by customization
Personalization for dual task environment and varying driving conditions	The system should recognize user needs in difficult driving situations
Personalization for personal environment, user attributes and product experience	The system should recognize personal needs for all types of users

4.5 Selecting evaluation methods

According to the International Organization of Standardisation (ISO) (1998) there is no general approach on how usability evaluation methods should be selected or combined. However, the evaluation methods should align with the purpose of the evaluation. As proposed by Stanton and Harvey (2013, p. 55) following aspects were considered when selecting applicable evaluation methods; context-of-use, criteria and KPI as well as time-scale of the project, resource constraints and access of people. Other evaluations, such as the evaluation proposal for in-vehicle information systems made by Stanton and Harvey (2013) as well as the usability evaluation conducted by Semcon (2015), also affected the selection. Many different evaluation methods were investigated, but due to time limitation not all evaluation methods could be considered. Among the investigated ones, those that are of interest for the thesis work are described in chapter 2.5. The selected evaluation methods for the KPIs are described below and shown in Table 4.5. Moreover, product representation, evaluation output and the evaluators are described.

4.5.1 Evaluation methods

As mentioned in the delimitations, the thesis work focused on analytical methods. The main reason is that the department of DVI has limited resources and consequently prefers analytical methods when evaluating prototypes during the design process. Another reason is that the simulation and evaluation team at the department of DVI already conducts extensive user tests in order to validate the performance and usability of prototypes.

Evaluation methods for pleasurability

As mentioned in chapter 2.2, UX is the user's perceptions and feelings towards the use of a system. Consequently, users should test the system in order to evaluate UX. The test scenarios can be based on a Hierarchical Task Analysis (HTA). It is important to acknowledge that the use of questionnaires in connection to usability tests can be misleading. This is due to the fact that many questionnaires only include positively and not negatively phrased questions. People are more likely to agree than disagree with a given statement which results in questionnaires being biased towards positive responses. In comparison to questionnaires, the Microsoft Desirability Toolkit (MDT) elicits negative comments and permits users to stay critical (Travis, 2009). As a result of this reasoning it is considered suitable to evaluate the KPIs related to pleasurability by the use of MDT, which was therefore selected as evaluation method. The simulation and evaluation team at the department of DVI already use this method to evaluate the satisfaction during user tests and they have made a customized version that suits Scania, which can be used. It is important to acknowledge that other empirical methods related to user tests, which are not discussed in the thesis work, might be required in order to capture all aspects of pleasurability.

Inspection-based evaluations together with a HTA were selected to evaluate the KPIs regarding pleasurability in terms of system quality. This is based on the fact that Inspection-based evaluations are part of the User-Centred Design (UCD) process, which includes UX. The fact that expert evaluations can evaluate UX is confirmed by Bevan (2008) who states that the UX expert evaluations should be based on guidelines or heuristics, or through simply letting the experts walk through tasks. Moreover, Vermeeren, Cremers, Kort and Fokker (2008) suggests that UX can be evaluated through expert reviews where the prototypes are assessed based on the evaluator's opinions or through letting experts imagine what the real end users would answer. As it is considered important for the experts to imagine what the real end users would experience it can be useful to combine the expert assessment with Role-playing or Empathic modelling.

Evaluation methods for task focus

Since Inspection-based evaluation is a good way to evaluate the usability and design quality of prototypes this evaluation method can be applied to many KPIs. In order to better understand the cognitive tasks, the evaluations should be based on a HTA. The method was selected for evaluating the KPI concerning the range of user attributes in relation to the interaction with the product or system and understanding of functions. Inspection-based evaluations were also selected for the KPI that concerns task structure and functions in relation to safety in difficult driving conditions. The same line of reasoning was applied for the KPIs that state that users should be able to distinguish and localise functions in varying driving conditions, that secondary task functions should be easy to localise and understand and that the performance of primary tasks should not be compromised by the use of secondary tasks.

The Inspection-based evaluation can also be used to evaluate errors. It was therefore selected to evaluate the KPI regarding the ability to recover from wrong actions. As a complement to the Inspection-based evaluation an error evaluation method such as Systematic Human Error Reduction and Prediction Approach (SHERPA) was selected. This is an analytical evaluation method that can predict both human and design initiated errors and is therefore suitable for the specific KPI. In order to conduct a SHERPA it is necessary to base it on a HTA.

KPIs related to mental workload can be evaluated through several evaluation methods. In order to make a reliable evaluation empirical methods are preferred over Inspection-based evaluations. Mental workload questionnaires should be executed in connection to user tests. In this case, Driver Activity Load Index (DALI) was selected since the questionnaire is specifically adapted to drivers and driving tasks. Consequently, it is considered to give more accurate results than for example NASA Task Load Index (NASA-TLX). As earlier mentioned, different types of mental workload questionnaires are already conducted by the simulation and evaluation team and it is therefore not necessary to change their procedures. However, DALI should be complemented by the use of performance measures, such as speed. This was therefore selected. The advantage of this is that the evaluation can be based on users performing both primary and secondary tasks. Moreover, physiological measures, such as skin conductivity and heart variance, together with a HTA should be implemented to evaluate the mental workload. Thus, this was also selected.

The KPI regarding the ability for novice users to handle essential functions can be evaluated by the use of Cognitive Walkthrough (CW) and was therefore selected. The reason is that CW evaluates the ease of learning, which is relevant for novice users. Learnability will therefore be evaluated to some extent even though it was not considered as a high level factor, as mentioned in 4.1. As earlier mentioned, CW is based on tasks and it is therefore necessary to conduct a HTA along with the CW.

The KPIs that concern task time should be evaluated through empirical methods based on users performing tasks. The reason is that experts that have prior knowledge of the system cannot perform tasks in a realistic manner. The task times should be evaluated through measuring the time to completion for different types of users and in difficult driving conditions. It is also possible to predict task time through analytical methods. Task time can be evaluated through evaluation methods such as Critical Path Analysis (CPA) and this was therefore selected. A major advantage of CPA is that it takes into account that tasks can be done in parallel.

Evaluation methods for flexibility

Regarding flexibility, all KPIs can be evaluated through the use of Inspection-based evaluations and this was therefore selected. The majority of the KPIs can easily be evaluated by experts within the domain. It is considered beneficial to base the evaluation on tasks and it is therefore necessary to conduct a HTA as well.

Table 4.5 KPI matrix.

KPI	analytical methods					empirical methods			
	Hierarchical Task Analysis	Cognitive Walkthrough	Inspection-based evaluation	Systematic Human Error Reduction & Prediction Approach	Critical Path Analysis	Microsoft Desirability Toolkit	Driving Activity Load Index	Physiological measures	Performance measures
Users should have a positive attitude towards the use of the system, on initial use and throughout the usage	✓					✓			
System performance should match user expectations throughout the usage	✓		✓						
All types of users should have a positive attitude towards the use of the system	✓					✓			
Users should have an emotional attachment to the product while not working	✓					✓			
Users should maintain a product liking through all channels such as marketing, sales, use of product and maintenance	✓					✓			
Users should feel like experts and in control of the system	✓					✓			
Users should feel safe when using the system	✓					✓			
All types of users should maintain a low mental workload while performing tasks	✓						✓	✓	✓
All types of users should be able to interact with the system as well as understand functions and their impact	✓		✓						
Novice users should be able to handle essential functions	✓	✓							

Users should be able to recover from wrong actions	✓		✓	✓					
Task times should be minimized regardless of driving conditions	✓								✓
Task structure and functions should support safety in difficult driving conditions	✓		✓						
Users should be able to distinguish and localise functions in varying driving conditions	✓		✓						
Secondary task times should be minimized	✓				✓				✓
Users should maintain a low mental workload while performing secondary tasks	✓						✓	✓	✓
The performance of primary tasks should not be compromised by secondary tasks	✓		✓						
Various paths through the system should be provided to the user that match user attributes and product experience	✓		✓						
Users should be able to customize the system for their specific segment tasks before use	✓		✓						
All types of users should be able to customize functions during use	✓		✓						
Tasks should not be negatively affected by customization	✓		✓						
The system should recognize user needs in difficult driving situations	✓		✓						
The system should recognize personal needs for all types of users	✓		✓						

4.5.2 **Product representation**

As mentioned in chapter 2.5, many of the selected evaluation methods can be conducted throughout the whole design process by the use of different types of product representations. This is an advantage since it is important to evaluate the design concepts in early stages of the project to better understand user needs. Moreover, the changes in this phase can be relatively inexpensive compared to later in the design process (International Organization of Standardization (ISO), 2010a).

As the analytical evaluation methods can be executed early in the design process, on various types of prototypes, only low fidelity prototypes are required when evaluating related KPIs. Still, it might be favourable to evaluate more advanced prototypes such as interactive digital prototypes. Furthermore, the department of DVI strives to better integrate the driver simulator and Virtual Reality (VR) into their design process, and these product representations should therefore also be used for evaluation.

Regarding empirical evaluation methods prototypes that a user can interact with should be used. This might require more advanced prototypes such as interactive digital prototypes represented in the driver simulator or VR.

The product representation also depend on the specific KPI to be evaluated. Some KPIs might not require to be evaluated in the driving simulator due to the fact that they are driving independent while other KPIs require evaluation in VR or the driving simulator. KPIs related to the understanding of the system do not necessarily need to be evaluated in VR or the driving simulator. However, tests in a more realistic context can uncover important aspects that affect the interaction and understanding of a system. An example would be that vibrations might affect the ability to press buttons. Consequently, this is important to take into consideration when evaluating these KPIs out of the real context. KPIs related to the dual task environment and different driving conditions are therefore suitable to evaluate in VR or the driving simulator.

4.5.3 **Evaluation output**

CW and SHERPA provide qualitative data since the outputs are written statements with extensive explanations. CPA, DALI, physiological and performance measures provide quantitative data since the outputs are numerical data. Inspection-based evaluation and MDT provide both qualitative and quantitative data. Furthermore, HTA is considered as basis for other evaluation methods and does not provide outputs other than a structure of the task to be evaluated.

4.5.4 **Evaluators**

The analytical methods, HTA, CW, Inspection-based evaluation, SHERPA and CPA, require experts. For DVI, expertise means experience and knowledge of UX, usability and human factors as well as the prototype to be evaluated. In this case the experts are the employees at the department of DVI. The main advantage of having employees with experience of DVI at Scania as evaluators is that they are familiar with Scania's previous work, guidelines and brand specific principles. Moreover, many of the employees have a truck driver's licence, which means that they have hands-on experience of the product. This is advantageous since this allows them to better understand the users. The evaluation should preferably be conducted in a group or by several evaluators separately. The reason is that different evaluators capture different aspects.

For all analytical methods, which include experts, it can be useful to consider role-playing and empathic modelling methods. Moreover, it can be valuable to use the Six Thinking Hats method, which provides separate and analytical thinking which can lead to product improvements. This is done through letting evaluators mentally wear hats with different mind sets; the white hat focuses on facts, the yellow hat is optimistic, the black hat judges, the red hat is full of emotions, the green hat is creative and the blue hat controls the activity (The de Bono Group, n.d.).

Since MDT, DALI, performance and physiological measures are empirical methods they require users to evaluate prototypes. The users can for example be professional truck or bus drivers.

4.6 Adapting evaluation methods

In order to successfully use the selected evaluation methods they needed to be revised in order to fit Scania and the specific KPIs. The evaluation methods mentioned in chapter 4.5 were therefore adapted and further developed. This was done in an iterative way by researching, developing and walking through the evaluation methods several times. When walking through the evaluation methods interactive prototypes that represent the third version of the ICL were used as a basis. They were created over time during the development of the NCG truck and represent different stages of the design process. Examples of the prototypes are shown in Figure 4.3, 4.4 and 4.5.

Figure 4.3 Digital interactive prototype of Instrument Cluster from 2011.

Figure 4.4 Digital interactive prototype of Instrument Cluster from 2013.

Figure 4.5 Digital interactive prototype of Instrument Cluster from 2015.

Evaluation frameworks of how to conduct the evaluation methods were developed for HTA, Inspection-based evaluation, SHERPA and CW. These are described in chapter 4.6.1. The main focus was on the Inspection-based evaluation since it covers many KPIs. All frameworks include a description of the aim and instructions on how to use the framework. Several frameworks also include examples that can be used as guidance when evaluating prototypes as well as templates to fill out during the evaluation. The graphic layout of the frameworks was based on guidelines given by Scania (Scania CV AB, 2016).

Evaluation frameworks of the other evaluation methods were not conducted due to the delimitations of the thesis work and time limitation. Nevertheless, recommendations for these evaluation methods were given. These are described in chapter 4.6.2. The adaption of each evaluation method are further described below and divided into evaluation frameworks and recommendations.

4.6.1 **Evaluation frameworks**

Hierarchical Task Analysis

The existing framework of HTA proposed by Stanton (2006), mentioned in chapter 2.5.1, was studied in order to develop an adapted evaluation framework of HTA. The HTA should be applied as a basis for other evaluation methods that evolve around tasks. The adapted HTA framework answers to related KPIs shown in Table 4.5.

The existing framework was revised to improve the understanding. For example, the steps of the instructions on how to conduct the method were clarified. Moreover, an example was created to complement the instructions. When combining the adapted HTA framework with other evaluation methods it was clear that depending on complexity of the HTA, different levels of the HTA (goals, sub-goals or operations) should be evaluated. Which level to be evaluated was set to be decided by the evaluator.

It is important to select tasks that are relevant to the specific KPIs and the evaluation methods. For example, when evaluating KPIs regarding dual tasks it is necessary to select tasks that are often executed while driving. The evaluators should also select tasks that are of high relevance or tasks that users execute frequently. For example, users often pair their phone with the AUS through Bluetooth. Since the task analysis can be time consuming the department of DVI can use existing task scenarios made for user tests when conducting HTAs. Consequently, HTAs can easily be implemented in their current work.

Cognitive Walkthrough

The existing framework proposed by Wharton, Rieman, Lewis and Polson (1994), mentioned in chapter 2.5.1, was studied in order to develop an adapted evaluation framework of CW. The adapted CW framework answers to related KPIs shown in Table 4.5.

The existing questions to be answered while the evaluators walk through the action sequence of each task to be evaluated were revised. Both the third version of the original CW proposed by Wharton, Rieman, Lewis and Polson (1994) and revised versions by other significant authors (Bligård & Osvalder, 2013; Magnusson, Rassmus-Gröhn, Tollmar, & Deaner, 2009) were analysed in order to determine suitable phrasing of the questions to be answered. The questions used are the following; "Will the user know what needs to be achieved? Will the user notice that the function is available? Will the user associate the cues with the function? Will the user get feedback when using the function? Will the user get feedback to understand that the task has been performed?"

Inspection-based evaluation

An adapted evaluation framework of Inspection based-evaluation was conducted through studying several existing checklists as suggested by Stanton, Salmon, Walker, Baber & Jenkins (2005, pp. 436-439) in chapter 2.5.1. Furthermore, the existing framework of HE, also mentioned in chapter 2.5.1, was considered. The Inspection-based evaluation answers to related KPIs shown in Table 4.5.

In order to structure the adapted evaluation framework it was divided into different parts and evaluation areas. The evaluation areas included different evaluation items expressed through statements to be analysed, scored and commented on by the evaluators. An example was added to the instructions, which includes the item; "The system layout is presented in a consistent manner".

Evaluation items from existing checklists for both in-vehicle systems as well as other application areas, mentioned in chapter 2.5.1, were included and adapted to better suit Scania as well as to improve the understanding. Other items suitable for the specific KPIs were also included in the adapted evaluation framework. When walking through the evaluation method items that were considered missing were added and unnecessary and repetitive items were removed. Moreover, numerous items were further revised since it was found that some items were difficult to interpret. During the development it was discovered that some evaluation items were impossible to score without supportive information. This type of information was therefore added in order to improve the framework. The supportive information was based on information provided through the existing checklists as well the principles of the sub-properties mentioned in chapter 2.4. An example of supportive information for the item "The system layout is presented in a consistent manner" is "Information (colour, text, icons and symbols) of related functions should be arranged similarly".

Some items needed more specific measures, which were found in the existing checklists and in internal or external standards. The need of specific measures for KPIs connected to task focus mainly concerned language, font, icons and symbols, colour and volume as well as safety aspects for driver vehicle interaction. The measures are described below.

The language should follow recommendations given by the department of linguistics at Scania and font type should be set to the defined Scania font. A font study has recently been conducted by the department of DVI and the font is continuously updated in order in order to improve the readability and usability. The font size should be set according to ISO (2017a). The recommended font size was also confirmed by Arbetarskyddsstyrelsen (1998) and Dobres, Reimer, Parikhal, Wean, and Chahine (n.d.) who recommend similar sizes. Icons and symbols should conform to standard road vehicle icons and symbols according to ISO (2010b) and the size should be set according to recommendations given by the department of DVI. The DVI department established these recommendations through research within the area. Colours to express danger should be set according to ISO (2017b) and colour combinations should conform to ISO (2017a). The volume of auditory information should follow recommendations given by the department of DVI. Numbers of inputs and the total task duration for safety should follow recommendations made by Stevens and Cynk (2011) and Japan Automobile Manufacturers Association (2004). Response time should follow guidelines given by Nielsen (1993b), which are also mentioned in chapter 2.4.5.

Notes were added to specific items suggesting the evaluator to consult others during the analysis since these items were considered difficult to analyse objectively. Harley (2016) proposes that in order to investigate if the meaning of the icons and symbols is clear one can test the icon recognisability through asking people to guess what the icons symbolize out of context. This was considered useful for understanding icons and symbols as well as text, voice messages, auditory and tactile signals. Second opinions can be given by co-workers at the department of DVI. To analyse voice messages and auditory signals tests in listening studio might be required. Furthermore, language might require an opinion from the department of linguistics. Notes were also added where the evaluator needs to execute an additional procedure. For example, in order to ensure that the colours chosen support vision impairments colour blindness tests should be executed by the use of Adobe Illustrator.

The variety of assessment scales used in existing checklists was also analysed when developing the adapted Inspection-based evaluation. During the development it was discovered that assessment through Likert scales was preferred instead of binary answers. The items of the Inspection-based evaluation were therefore set to be assessed on a 5 point Likert Scale with the following scores; *strongly disagree, disagree, neutral, agree* and *strongly agree.* The reason for choosing a Likert Scale is that some items concern several aspects to be evaluated. Moreover, Likert Scales are simple to conduct, easy to interpret by the evaluators and are considered as reliable scales. The main disadvantage is that they have a central tendency bias, which must be taken into consideration. Another disadvantage is that users show tendencies to agree with statements in order to please the researcher (Bertram, n.d.).

However, the latter disadvantage mentioned was not considered relevant for this type of evaluation.

Moreover, it was considered beneficial to assess the importance of the items since some items are more or less critical or even not applicable to the specific prototype to be evaluated. This way the final score is not biased towards non critical aspects. Rating the design issues in such way was suggested by Arvola (2014, p. 138). The importance scale was set to; *low, medium* and *high*. In addition, the evaluator should comment on design issues and improvement when needed. This way the score will result in quantitative data while the comments will result in qualitative data. The final comments conform to the existing framework of HE since the output of HE is also a list of detailed comments.

As proposed by Heo, Ham, Park, Song and Yoon (2009) the adapted Inspection-based evaluation was also divided into task dependent and task independent items. This means that some items should be scored when performing tasks, while other items should be scored through interacting freely with the prototype. Markings next to the items were added to describe if it they should be evaluated as task dependent or independent. A HTA should preferable be used as basis when analysing task dependent items.

Systematic Human Error Reduction and Prediction Analysis

The existing framework of SHERPA proposed by Stanton, Salmon, Walker, Barber and Jenkins (2005, pp. 143-151), mentioned in chapter 2.5.1, was studied in order to develop an adapted evaluation framework of SHERPA. The adapted SHERPA framework answers to the related KPI shown in Table 4.5.

The instructions of the existing framework were revised in order to improve the understanding. Moreover, the classification of task types and error types was adapted to better suit the department of DVI and the specific KPI as well as to clarify the options. An example of a clarification of an error type is that “Operation omitted” was changed to “The user does not complete the action”. In a similar way the language of the other error types as well as task types was improved. A template for filling out the answers as well as an example was created to complement the instructions. Furthermore, SHERPA was set to be evaluated on the each step of a suitable level of a HTA.

4.6.2 **Recommendations**

Critical Path Analysis

The evaluation framework of CPA is recommended to be adapted based on the existing framework proposed by Stanton, Salmon, Walker, Baber and Jenkins (2005, pp. 507-512), mentioned in chapter 2.5.1. This is recommended for related KPI shown in Table 4.5. The evaluation method was not further investigated due to time limitations but it is considered useful to predict task times through the use of analytical methods and not only measure the task times through user tests.

Microsoft Desirability Toolkit

The existing framework for MDT proposed by Travis (2009), mentioned in chapter 2.5.2, is recommended to be used in connection to user tests. Moreover, the framework used by the simulation and evaluation team at the department of DVI should work as a complement to this. Since MDT is an empirical method is was not further investigated in the thesis work due to the delimitations. It is recommended to adapt the framework in terms of a list of words that better suit Scania's visions as well as the related KPIs, shown in Table 4.5.

Driver Activity Load Index

The existing framework for DALI proposed by Pauzié (2008), mentioned in chapter 2.5.2, and the one used by the evaluation and simulation team at the department of DVI, is recommended to be used. DALI as well as other mental workload questionnaires are already conducted by the simulation and evaluation team and can therefore easily be integrated into the evaluation of the driver vehicle interaction. Since DALI is a questionnaire that should be conducted in connection to user tests, which is an empirical method, it was not further investigated in the thesis work due to the delimitations. The existing framework for DALI should be used to evaluate the related KPIs shown in Table 4.5.

Physiological measures

The existing frameworks for physiological measures such as eye movements, heart rate variance, blink rate and skin conductance, mentioned in chapter 2.5.2 should be used in order to evaluate mental workload in connection to user tests. It is advantageous that the simulation and evaluation team already executes these types of measurements since the evaluations require certain equipment and expertise within the domain. Consequently, the work that the simulation and evaluation team does should continue. Since the physiological measures are empirical methods they were not further investigated in the thesis work due to the delimitations However, it is important to acknowledge that all different types of physiological measures should be studied in order to ensure that the chosen methods capture all aspects of the related KPIs shown in Table 4.5.

Performance measures

It is recommended to use existing frameworks of performance measures, mentioned in chapter 2.5.2, in connection to user tests. The performance measures that are not already conducted by the simulation and evaluation team at the department of DVI can likely be implemented without major costs and efforts. It would therefore be easy to integrate the procedures and frameworks into the evaluation of driver vehicle interaction. This is recommended for related KPIs shown in Table 4.5.

Regarding the KPIs that concern task times the time to completion should be calculated during user tests. It is important to set predefined goals or to compare the results with other prototypes, products or systems. It should be taken into account that there are rules on how long drivers can take their eyes off road. The total off road time for a task should not exceed eight seconds (Japan Automobile Manufacturers Association, 2004). The KPIs that regard mental workload should also be evaluated through the use of performance measures. As mentioned in chapter 2.5.2 the mental workload of in-vehicle systems can be evaluated by the use of speed, lateral position and headway measures. Still, this should be further investigated to ensure that the measures cover the KPIs. It is important to select appropriate tasks as well as users in order to cover the performance of secondary tasks and a wide range of users.

Since the performance measures are empirical methods they were not further investigated in the thesis work due to the delimitations.

4.7 Testing evaluation methods

Six tests were carried out in order to find potential errors in the developed evaluation frameworks. The tests were conducted gradually during a time period of two weeks, where each test had a duration of two hours. Not all evaluation frameworks could be tested due to time limitations. The evaluation frameworks tested were one part of the Inspection-based evaluation framework and the framework of SHERPA. Some participants were asked to start with the Inspection-based evaluation while other participants started with SHERPA to neglect the influence of learning. The participants were asked to evaluate prototypes based on the evaluation methods and questions about the evaluation frameworks were asked both during and after the tests. The answers together with observations made during the tests resulted in qualitative data which was analysed through affinity diagramming. This contributed to the development of the final evaluation frameworks. Even though the tests were conducted in order to evaluate the evaluation frameworks the participants found design issues in the prototype used during the evaluation. The issues found were similar, for example many participants found it difficult to find specific functions due to the lack of symbol perception and text.

The test participants were six employees at the department of DVI, which are the actual end users. Three of the participants were interaction designers at Scania while the other three were UX design consultants currently working at Scania. Moreover, the Scania employees had truck driver’s licences, which means that they have hands-on experience of the product.

One pilot test was conducted with the thesis work supervisor at Scania before the final tests in order to improve the test protocols. This also resulted in improvements that contributed to the development of the final evaluation frameworks.

4.7.1 **Test protocols**

The test protocols included information that was used to describe the test to the participants and explain its purpose. An advantage of this was that all participants received the same instructions. Moreover, the protocols included the material needed and the questions to be answered by the participants during the test. During the evaluation questions about the understanding and possible improvements were asked. This concerned the instructions and the scoring system as well as all steps of SHERPA and the items of the Inspection-based evaluation. In the end of the test questions regarding missing parts, overall improvement suggestions, overall impression of the evaluation framework and the layout were included. The test results are described in chapter 4.7.3 and the protocols can be found in Appendix B.

4.7.2 **Test material**

One of the digital interactive prototypes that were used during the adaption of the evaluation methods were set to be evaluated by the participants during the tests. The prototype represents the third version of the ICL developed in 2011 (se Figure 4.3). This was created during the development of the NCG truck and represent one stage of the design process. The participants were given hard copies of the evaluation frameworks to be tested. During the tests the participants were asked to fill these out as if it would have been an actual evaluation of the prototype.

Both the Inspection-based evaluation and SHERPA are based on tasks and should preferably be conducted together with a HTA. A focus group was held in order to investigate important tasks related to the prototypes to be evaluated. Example of tasks that where discussed were pairing mobile devices, navigation re-routing and programmable fields. Based on these findings a HTA was developed that the participants were able to use as basis when evaluating the prototype during the tests. The selected task was to set the upper right programmable field in the ICL to cruise control information, the HTA is shown in Figure 4.6. This was one of the tasks that could be performed on the selected prototype. In practice the evaluators will have to conduct the HTA by themselves but in order to optimize the tests and minimize the test duration this was excluded.

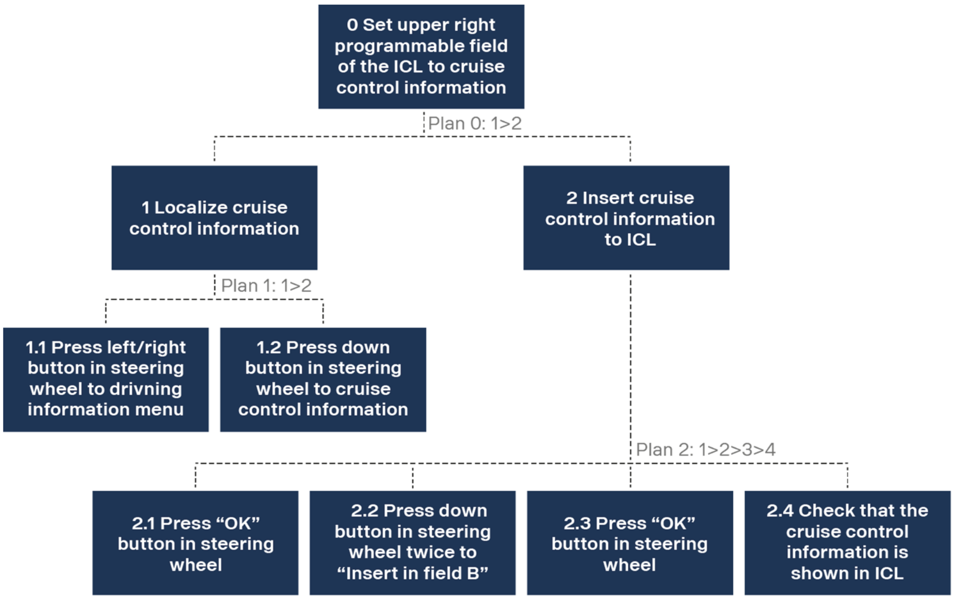


Figure 4.6 Hierarchical Task Analysis used as basis during tests.

4.7.3 Test results

Inspection-based evaluation

In general the participants were positive towards the Inspection-based evaluation. They thought that the evaluation method was elaborative and that many important parameters were taken into consideration. However, one participant mentioned that the method might only focus on details and that the holistic perspective might be left out. Many of the participants thought that the evaluation was difficult in the beginning and that it would be useful to get training in beforehand. This would also be advantageous since this would result in the evaluators interpreting the evaluation items similarly. The participants believed that the evaluation would be much easier to conduct when they have gained experience of using the method.

Many of the participants mentioned that the evaluation could be biased when evaluating their own projects and that someone else at the department of DVI should execute the evaluation. They also mentioned that the evaluation might be suitable to do as a group within cross functional teams.

The majority of the participants found the instructions extensive and too long. Nevertheless, they found the content relevant in order to start the evaluation. Many participants mentioned that the example helped clarifying the instructions. The participants also asked for further description of some parts of the instructions such as what HTA, KPI and Likert Scale are and what should be included in the comments section. One participant mentioned that it would be useful to fill out who has done the evaluation and which task goal has been assessed. The participants commented on the layout of the instructions and suggested changing the disposition of titles and removing abbreviations.

Some participants mentioned that it was suitable that the items were divided into task dependent and task independent and that it was clear how to conduct the evaluation according to this. At the same time other participants found it hard to distinguish between the two terms. They mentioned that better instructions and more distinct marking next to the items were needed. Some participants mentioned that it was very helpful that the items were marked with supportive information on how they should be analysed. This way the evaluators do not need to interpret the items on their own. In general the majority of the participants confirmed the reasoning that the items should be scored based on universal standards and not on Scania's guidelines. Consequently, the results will be more trustworthy. Some participants mentioned that they would have liked to score different parts of the supportive information differently. One participant suggested that this could be solved by adding instructions on how to score items with diverse supportive information. Some participants commented on the layout of the content and suggested improvements such as changing the numbering of items and adding headers to each page including current part and evaluation area.

The majority of the participants found the items relevant but commented on specific items that needed to be revised or added. Some participants thought that the items should be updated continuously by the department of DVI. One participant asked for further references to standards included in specific items. Items marked with *note* were questioned by some participants. They could not see the purpose of consulting someone else for advice in order to analyse items and thought that this could be done by themselves.

The participants found the assessment scales suitable. Still, one participant underlined that the neutral option should be removed in order to force the evaluator to either agree or disagree. In general, they were very positive to the importance scale since it captured important aspects that were not considered in the assessment scale. One participant suggested that the department of DVI should decide the importance for each item in beforehand and that this should be used as basis for all evaluations. Many participants were positive towards the fact that the evaluation method included subjective responses in the comments section. The participants appreciated the possibility to capture design issues and possible improvements. Moreover, the participants had general comments on the layout of the scoring system that needed to be revised to make it more understandable.

Several participants proposed that the evaluation frameworks should be digitalised in order to compile the results more effectively.

SHERPA

In general the participants had difficulties using SHERPA and found it time consuming. However, participants mentioned that the output of the method could be very useful and that it should be used. One participant mentioned that by conducting the evaluation on bottom level operations as suggested possible errors are captured that might not have been noticed otherwise. On the other hand, other participants mentioned that it could be difficult to capture all possible errors made by actual users.

The participants found the instructions difficult and needed to go through them several times before they could start the evaluation. Many participants mentioned that the example helped clarifying the instructions but that the example should be simpler. One participant mentioned that it would be beneficial to try out the method before the actual evaluation. Some participants mentioned that it would be useful to get training in beforehand and believed that when they have gained experience of using the method the evaluation would be much easier to conduct.

The majority of the participants mentioned that the different classifications of task types and error types needed to be revised since they were difficult to interpret. One participant mentioned that a description on how to select error types should be added. Furthermore, the possibility to select several error types for each task step and consider several aspects when describing the error types should be described. One participant mentioned that it is important to point out that the consequences refer to the consequences that affect the specific task and not the surroundings, such as vehicle safety.

Some participants commented on the template to be filled out by the evaluator and suggested that some titles should be added and removed as well as rearranged. They also said that more space is needed in the template in order to be able to fill it out properly. However, this would not be a problem if the evaluation framework was digitalised.

HTA

Even though the HTA itself was not tested the participants commented on it since it was used as a basis during the tests. The majority of the participants were familiar with the concept of HTA and were positive towards the method. However, one participant found it difficult to interpret the HTA since it was not developed by the participant itself.

The participants mentioned that it would be good to add extra information to the instructions of the HTA. They should include a description of how the tasks to be evaluated should be selected. Preferably tasks with higher criticality or frequency of use should be considered. Moreover, the interaction elements that should be evaluated should be included in the HTA descriptions.

4.8 Modifying evaluation methods

After the tests the evaluation frameworks were modified according to the test results. The majority of the results were implemented and those insights commented by several participants were given higher priority. Test results that were irrelevant and in some cases contradictory or could not be implemented due to time limitations were not considered.

The instructions as well as the content of both the Inspection-based evaluation framework and the SHERPA framework were revised and extended. Furthermore, the layouts were changed according to the test results. Some of the notes in the Inspection-based evaluation framework were removed. Many participants mentioned the importance of training in beforehand which is something that needs to be considered by the department of DVI. Furthermore, the evaluation should be conducted by someone outside the specific project and preferably by several evaluator either individually or as a group.

The assessment scale of the Inspection-based evaluation framework was not changed as suggested since the majority of the participants found the scale suitable. Several participants proposed that the evaluation frameworks should be digitalized but due to time limitation this was not considered.

Even though the Inspection-based evaluation and SHERPA were the only evaluation frameworks tested the test results could to some extent also be implemented to HTA and CW. Similar modifications regarding instructions and layout were implemented to these evaluation methods. The final evaluation frameworks can be found in Appendix C.

4.9 Case study

A case study was conducted in order to evaluate an ongoing project at the department of DVI concerning the development of new features of the ICL and AUS. The aim of the case study was to compare two different design proposals and give recommendations on design improvements.

The evaluation was based on the developed evaluation frameworks of Inspection-based evaluation, part one and five, and SHERPA. The case study was executed in the driver simulator at Scania and a HTA was developed for a task that was significant for the project and which will be executed frequently. For SHERPA and part one of the Inspection-based evaluation the task was executed when standing still, while part five of the Inspection-based evaluation was evaluated while driving since this part concerns a dual task environment.

The evaluation was conducted during six hours and the evaluators were the thesis workers together with one employee from the department of DVI who is familiar with the project and an expert within human factors.

During the case study two different prototypes, representing the different design proposals, were evaluated. In general, major issues were found in both designs. However, one of the designs was preferred since it was considered more easy to use, both when standing still and while driving. The evaluation resulted in a list of recommendations on design improvements for both designs. These results were later presented to the department of DVI. For example, it was recommended to use a more clear colour coding since the main colour used in the prototypes represented different things, which was considered confusing. The font type and font size should also be more consistent by having the same font for text with similar purpose. Moreover, idea proposals regarding sizes and placements of functions and digital controls for each design were given. The proposals would result in a more intuitive system as well as a more efficient interaction while driving. Concerning the dual task environment both designs were considered suitable. However, minor concerns about the ability to keep focus on the primary task of driving while performing the task scenario was discovered for one of them. The reason for this was that it was difficult to use some controls when driving. The results are not further presented in the thesis work due to confidentiality.

In summary, the case study showed that the evaluation frameworks are possible to implement on real projects and that they will result in useful recommendations that can make a difference for specific projects. Nevertheless, as predicted during the tests the methods were found to be time consuming. Since the case study only covered two of the proposed evaluation frameworks and were conducted during six hours, one can predict that conducting all of the evaluation methods will take much more time.

5 Results and conclusion

This chapter addresses several results found throughout the thesis work. Furthermore, the conclusion includes the answers to the research questions; What factors of Driver Vehicle Interaction should be evaluated? How should the factors of Driver Vehicle Interaction be evaluated? When in the design process should the evaluation be carried out?

5.1 Results

The thesis work has led to many results, which are all important for the evaluation of driver vehicle interaction. If the final results are not implemented directly the partial results will still be interesting for the future work at the department of Driver Vehicle Interaction (DVI).

The partial results were found during the main steps of the thesis work methodology. Literature studies resulted in the selection of pleasurability, task focus and flexibility as proposed *high level factors* to be evaluated. Affordance, feedback, consistency, responsiveness and error handling are represented through these. It is important to take the context-of-use into consideration in order to develop the evaluation method and important *contextual factors* are proposed as a result from interviews. Contextual factors that need to be taken into consideration are primary and secondary tasks, different segments within trucks and buses, the specific hardware and software used, range of personal attributes, range of product experience, dual task environment, different driving conditions and personal environment. Interviews also resulted in 16 *criteria* related to the high level factors in combination to the contextual factors. Examples of criteria are; short and long term satisfaction, efficiency in varying driving conditions and ability to customize for different segments (see Table 4.1 for the full list of criteria). An analysis of the criteria resulted in 23 *Key Performance Indicators* (KPIs) to be measured through evaluation. Examples of KPIs are; system quality should match user expectations throughout the usage, task times should be minimized regardless of driving conditions and users should be able to customize the system for their specific segment tasks before use (see Table 4.2, 4.3 and 4.4 for the full list of KPIs). Literature studies then resulted in nine *evaluation methods* suitable for measuring these KPIs. Literature studies also showed that these can be used during different *stages of the design process*, and consequently by the use of different *prototypes*, depending on the tasks and if they are formative or summative methods.

The final result is a proposal of an evaluation method consisting of evaluation frameworks of four evaluation methods and recommendations regarding five evaluation methods. All evaluation methods suggested for pleasurability, task focus and flexibility are summarized in Table 5.1. The evaluation frameworks for Hierarchical Task Analysis (HTA), Inspection-based evaluation, Systematic Human Error Reduction and Prediction Approach (SHERPA) and Cognitive Walkthrough (CW) are described below. The recommendations for the other evaluation methods can be found in chapter 4.6.2.

The analytical methods can be conducted during the entire design process, which is beneficial. Consequently, it is possible to use different types of prototypes, which are developed during the process. The different product representations are paper-based prototypes, digital interactive prototypes, Virtual Reality (VR) and driving simulator. The time frame is illustrated through the use of prototypes in Figure 5.1

Table 5.1 Suggested evaluation methods and related high performance indicators.

	Hierarchical Task Analysis	Cognitive Walkthrough	Inspection-based evaluation	Systematic Human Error Reduction & Prediction Approach	Critical Path Analysis	Microsoft Desirability Toolkit	Driving Activity Load Index	Physiological measures	Performance measures
	✓		✓			✓			
	✓	✓	✓	✓	✓		✓	✓	✓
	✓		✓						
	analytical methods					empirical methods			

Figure 5.1 Product representations.

The HTA framework consists of six steps to be conducted by the evaluators. The aim of the method is to describe tasks in terms of hierarchy of goals, sub-goals, operations and plans. The upper levels of the hierarchy should include goals and sub-goals while the bottom level should describe what operations needs to be accomplished in order to reach the sub-goals. Furthermore, the plans should indicate how the goals are achieved. The method is often required as a basis for other evaluation methods that evolve around tasks. The HTA framework can be found in Appendix C.

The CW framework consists of five questions that should be answered when walking through each task step of a HTA. The answers, together with improvement strategies, should be filled out in a template. The evaluation covers one Key Performance Indicator connected to task focus. The aim of the evaluation method is to discover usability issues related to first time usage of systems. The output of the evaluation method is a qualitative improvement strategy. The CW framework can be found in Appendix C.

The Inspection-based evaluation framework consists of 72 evaluation items with additional information on how to interpret them. The items are divided into eight parts, which cover ten different KPIs connected to pleasurability, task focus and flexibility. Some parts are further divided into evaluation areas. The aim of the evaluation method is to provide a structured approach for evaluating the usability and design of a specific interface. The output is a quantitative score which can be used in the overall evaluation of the driver vehicle interaction as well as qualitative comments on design improvements. The Inspection-based evaluation framework can be found in Appendix C.

The SHERPA framework consists of eight steps that should be filled out in a template for each task step of a HTA. For some steps there are alternative options while others require written answers. The evaluation method covers one Key Performance Indicator connected to task focus. The aim of the evaluation method is to predict potential human or design initiated errors and related recovery potential. The output of the evaluation method is a qualitative improvement strategy. The SHERPA framework can be found in Appendix C.

In general, the Inspection-based evaluation framework and the SHERPA framework received good results during tests and the participants were positive towards the use of such evaluation methods. Still, the tests showed that the evaluations are time consuming and difficult to execute without training in beforehand. The results showed that the evaluations should be conducted by a group of evaluators with knowledge within User Experience (UX), usability and human factors. This was also confirmed by literature, which also suggested that it can be beneficial to use methods such as role-playing and empathic modelling to make the evaluators better understand the users. The tests showed that the Inspection-based evaluation is elaborative and that it takes many important aspects to be evaluated into consideration. SHERPA did not receive as good results since it was considered difficult and unnecessarily detailed. The tests led to modifications and improvement of the evaluation frameworks. All of the test results and the modification can be found in chapter 4.7.3 and 4.8.

The case study showed that it is possible to implement the proposed evaluation frameworks on real projects. The evaluation conducted during the case study resulted in recommendations on possible design improvements of the designs that were evaluated. Furthermore, the case study showed that the evaluations are time consuming.

5.2 Conclusion

5.2.1 **What factors of Driver Vehicle Interaction should be evaluated?**

The factors of driver vehicle interaction that should be evaluated are the sub-properties; pleasurability, task focus and flexibility and related KPIs, shown in Table 4.2, 4.3 and 4.4, which are measurable indicators that should be achieved. These factors were considered most relevant and were therefore selected as high level factors that laid foundation for the development of the evaluation methods. The main reason was that the other sub-properties contribute to the three terms.

5.2.2 **How should the factors of Driver Vehicle Interaction be evaluated?**

The factors should be evaluated through the use of a combination of analytical and empirical approaches. The analytical methods proposed are HTA, Inspection-based evaluation, SHERPA, CW and Critical Path Analysis (CPA). Moreover, the empirical methods proposed are Microsoft Desirability Toolkit (MDT), Driver Activity Workload Index (DALI), physiological and performance measures. As one of the delimitations of the thesis work was to mainly focus on analytical approaches only frameworks for these types of methods were developed. Unfortunately, a framework of CPA was not developed due to time limitations. The final frameworks are attached in Appendix C. For the other evaluation methods recommendations presented in chapter 4.6.2 should be followed. Time limitations also resulted in that the tests only covered parts of the Inspection-based evaluation and SHERPA and consequently only these frameworks were modified based on feedback from the participants.

The analytical methods should be conducted in a group or by several evaluators separately. This way the different evaluators can complement each other. The evaluators should be employees at the department of DVI that have experience and knowledge of UX, usability and human factors as well as the system to be evaluated. It can be useful to execute Role-playing or Empathic modelling in order to understand the users when conducting analytical methods. In addition, training in beforehand should be implemented in order to ensure that the evaluation frameworks are interpreted similarly by all employees.

5.2.3 When in the design process should the evaluation be carried out?

The factors should be evaluated during different stages of the design process since the proposed evaluation methods are a combination of formative and summative methods. The analytical approaches can be evaluated on different types of product representations during the design process. Thus, the recommendation is to evaluate paper prototypes, digital interactive prototypes, VR prototypes and simulations. The empirical approaches can be evaluated on different types of product representation during the design process or on the actual product in the end of the design process. Depending on the Key Performance Indicator suitable tasks should be chosen, which also affect the choice of product representation.

6 Discussion

This chapter discusses the methodology used throughout the thesis work and the proposed evaluation frameworks and recommendations as well as possible further work that needs to be taken into consideration.

6.1 Methodology

The methodology used in the thesis work was based on methods for developing evaluation methods, both for vehicles and within other areas, found in literature. Consequently, a relatively broad perspective was taken into consideration when developing the methodology for the specific project. However, other methodologies could have been used, which might have given other results. For example, it would have been possible to follow a general design or development process since the thesis work actually develops a product. An example of a design process is the double diamond design process proposed by the Design Council (2018). The main steps of the methodology is to discover, define, develop and deliver. Another method is the Concept Development Process proposed by Ulrich and Eppinger (2012). The main steps of the methodology is planning, concept development, system-level design, detail design, testing and refinement and production. The use of these types of methodologies might have resulted in more creative thinking and the evaluation method could have been developed from scratch rather than based on evaluation methods found through literature.

6.2 Proposed evaluation frameworks and recommendations

The evaluation frameworks were mainly based on literature studies. The sources are considered credible since they were written by significant authors within User Experience (UX), usability and the vehicle domain or similar. This indicates the validity of the proposed evaluation frameworks. Moreover, interviews and tests with employees at the department of Driver Vehicle Interaction (DVI), which are considered experts within UX, usability and human factors as well as the specific context, contributed to the evaluation frameworks. This way the frameworks are also based on expert knowledge, which can be seen as an indication of credibility.

Both during the development of the evaluation frameworks as well as during the tests similar design issues were found. The issues found during the tests of the third version of the Instrument Cluster (ICL) prototype from 2011 have been changed in the design of the ICL used today in the truck Next Generation Scania (NCG), which received a high score in the validation made by Semcon. Thus, one can draw a conclusion that the evaluation methods do capture relevant aspects, which need to be taken into consideration, in an effective and systematic way. This was also confirmed by the test participant's general comments on the evaluation frameworks.

When developing and testing the Inspection-based evaluation it was also clear that the Key Performance Indicators (KPIs) for task focus capture aspects of affordance, feedback, consistency, responsiveness and error handling. All evaluation items can to some extent be connected to one of the mentioned sub-properties. Moreover, it was clear that Systematic Human Error Reduction and Prediction Approach (SHERPA) covers important aspects of error handling. The same line of reasoning was done for pleasurability since it was found that for example Microsoft Desirability Toolkit (MDT) includes many aspects of task focus, and consequently affordance, feedback, consistency, responsiveness and error handling, as well as flexibility. This confirmed the selection of high level factors and the conclusion that pleasurability, task focus and flexibility are the most important factors of the driver vehicle interaction.

The majority of the evaluation methods gives a subjective result and not only an objective rating of each sub-property. All analytical evaluation methods require additional subjective comments on positive and negative aspects of the prototype to be evaluated as well as improvement proposals. This was considered more beneficial for the department of DVI as well as for the depth of the thesis work. However, this output can be difficult to use to strengthen the arguments in project status reports towards the management since they would like to see figures of the development process. Nevertheless, it can be possible to derive an objective score through the use of some of the evaluation methods. For example, the output of the Inspection-based evaluation includes a score from one to five.

It is important to underline that the proposed evaluation methods cannot work as a substitute to user tests. The reason is that experts will not have the same experience as the real end users and consequently not assess the system in the same way. Instead, the analytical methods can work as a complement in order to evaluate the driver vehicle interaction in early stages and to avoid unnecessary faults in the design. The proposed empirical methods can easily be integrated into the existing procedures of user tests conducted by the simulation and evaluation team at the department of DVI.

6.3 Further work

As earlier mentioned, not all evaluation methods proposed were developed or tested due to time limitations and the delimitation to focus on analytical methods. Consequently, further work needs to be done. Before implementing the evaluation methods it is recommended to further test the evaluation frameworks in an iterative way, primarily those method that have not yet been tested. General tests should be executed in order to ensure reliability and validity. Moreover, it can be beneficial to test the evaluation frameworks on participants that have used the evaluation a few times and by letting a group of evaluators test the methods. The recommendations for empirical methods, given in chapter 4.6.2, should be developed, further investigated and tested. Moreover, it is recommended to question the proposals and to investigate other evaluation methods as well as the content of the proposed evaluation methods. For example, Inspection-based evaluation can likely cover more KPIs related to pleasurability than discussed in the thesis work. The reason is that it is shown that pleasurability can be evaluated through the use of such methods.

Since one of the delimitations was to develop the evaluation method based on the ICL, the Infotainment System (AUS) and the buttons in the steering wheel it is not confirmed that the evaluation works for other parts of the truck. In order to ensure that the methods suit other areas it should be further developed and analysed. However, the frameworks were made relatively general in order to enable future evaluation of all types of systems within the vehicle and it might therefore be possible to implement in other areas.

The delimitation to focus on the specific parts of the instrument panel also resulted in that the majority of the KPIs and consequently the evaluation frameworks are adapted for vehicles and driving situations. The reason is that the literature underlines the importance of basing evaluations on the context-of-use. The proposed evaluation methods are therefore not general methods for evaluating interaction, which must be taken into consideration if the department of DVI wants to evaluate other systems such as smartphone applications or control rooms. Thus, the overall methodology requires changes if the methods should cover other contexts. Moreover, the evaluation frameworks do not cover interaction elements that has not yet been developed or is missing. This should also be taken into consideration when implementing the evaluation.

How to summarize and interpret the outputs, in terms of qualitative and quantitative data, of the suggested evaluation methods have not been investigated in the thesis work. Thus, it is important to acknowledge that this needs to be done in the future in order to implement the evaluation methods correctly. This would involve further investigation on statistics and how one can derive a total score based on the different evaluation methods and outputs. The benefit of this would be to show the project status internally at Scania. Moreover, it is important to extend the evaluation frameworks so that it is possible to set requirements in the beginning of projects.

This could be for example be done through defining the level of importance for each item of the Inspection-based evaluation as mentioned during the tests. It can be beneficial to base the requirements and evaluations on comparisons with competitors or older versions.

As suggested during the tests it would be beneficial to have a digitalized version of the evaluation frameworks. This would make it easier to summarize the results and the evaluation process would be more convenient. Due to time limitations this was not developed and it is therefore recommended to implement digitalized versions in the future.

References

Ahrne, G., & Svensson, P. (2015). *Handbok i kvalitativa metoder* (2nd ed.). Stockholm, Sweden: Liber AB.

Arbetarskyddsstyrelsen. (1998). *Arbete vid bildskärm.* Retrieved May 25, 2018, from https://www.av.se/globalassets/filer/publikationer/foreskrifter/arbete-vid-bildskarm-foreskrifter-afs1998-5.pdf

Arvola, M. (2014). *Interaktionsdesign och UX.* Lund, Sweden: Studentlitteratur AB.

Bergmark, M., Hallila, R., Karlsson, A., & Nilsson, J. (2015). *2015 Scania NCG HMI - Validation report.* Gothenburg, Sweden: Semcon.

Bertram, D. (n.d.). *Likert Scales.* Retrieved May 23, 2018, from http://poincare.matf.bg.ac.rs/~kristina/topic-dane-likert.pdf

Bevan, N. (2008). Classifying and Selecting UX and Usability Measures. (E.-C. Law, N. Bevan, G. Christou, M. Springett, & M. Lárusdóttir, Eds.) *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement*, 13-18.

Bligård, L.-O., & Osvalder, A.-L. (2013). Enhanced Cognitive Walkthorugh: Development of the Cognitive Walkthrough Method to Better Predict, Identify, and Present Usability Problems. *Advances in Human-Computer Interaction*, 17.

Bligård, L-O., researcher at the division of Design and Human Factors, Chalmers University of Technology, Gothenburg, Sweden. Telephone communications. (2018, February 5).

Boman, D. (2017). *17070-311*. Retrieved May 23, 2018, from Scania media provider: https://media.scania.com/fotoweb/Preview.fwx?&position=20&archiveType=ImageFolder&archiveId=5106&albumId=5106&sorting=ModifiedTimeAsc&search=interior&fileId=C7BC39EB0244DBA34CF2FADA92307FF130607FBCB0030271854EF17105CF8A405D760AB5A2C8FDE94EA745C98C309131D2

Cao, J., Zieba, K., Stryjewski, K., & Ellis, M. (2015). *Consistency in UI Design - Creativity Without Confusion.* UXPin Inc.

Context of Use Analysis. (2009). Retrieved February 14, 2018, from http://www.usabilitybok.org/context-of-use-analysis

Dehghanpour, A. (2015). *Feedback.* Södertälje, Sweden: Driver Vehicle Interaction, Scania CV AB.

Design Council. (2018). *The Design Process: What is the Double Diamond?* Retrieved May 19, 2018, from Design Council: https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond

Dobres, J., Reimer, B., Parikhal, L., Wean, E., & Chahine, N. (n.d.). *The incredible shrinking letter: how font size affects the legibility of text viewed in brief glances.*

Emardson, R. (2015, September 3). Ragne Emardson - Om att mäta hållbarhet [video file]. Retrieved February 12, 2018, from https://www.youtube.com/watch?v=fadyzFl2o3A

European Commission . (2008). Commission Recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human-machine interface. *Official Journal of the European Union*.

Garvin, D. A. (1987). Competing on the eight dimensions of quality. *Harvard Business Review, 65*(6), 101-109.

Gkouskos, D., & Chen, F. (2012). *The use of affective interaction design in car user interfaces.* Gothenburg, Sweden: Department of Interaction Design, Chalmers University of Technology.

Gray, W., & Salzman, M. (1998). Damanged Merchandise? A Review of Experiments That Compare Usability Evaluation Methods. *Human-Computer Interaction, 13*(3), 203-261. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15327051hci1303_2

Gu Ji, Y., Ho Park, J., Lee, C., & Hwan Yun, M. (2006). A Usability Checklist for the Usability Evaluation of Mobile Phone User Interface. *International Journal of Human-Computer Interaction*, 207-231.

Harley, A. (2016). *Usability Testing of Icons.* Retrieved April 13, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/icon-testing/

Heo, J., Ham, D.-H., Park, S., Song, C., & Yoon, W. (2009). A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors. *Interacting with Computers, 21*(4), 263-275. Retrieved March 12, 2018, from http://resolver.ebscohost.com.ludwig.lub.lu.se/openurl?sid=EBSCO%3aedselp&genre=article&issn=09535438&isbn=&volume=21&issue=4&date=20090101&spage=263&pages=263-

275&title=Interacting+with+Computers&atitle=A+framework+for+evaluating+the+usability+of+mobile+

International Organization of Standardization (ISO). (1998). *Ergonomic requirements for office work with visual display terminals (VDTs) Part 11: Guidance on usability (ISO 9241-11)*. Geneva, Switzerland: International Organization for Standardization (ISO).

International Organization of Standardization (ISO). (1999). *Human-centered design processes for interactive systems (ISO 13407)*. Geneva, Switzerland: International Organization for Standardization (ISO).

International Organization of Standardization (ISO). (2010a). *Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems (ISO 9241-210:2010)*. Geneva, Switzerland: International Organization of Standardization (ISO).

International Organization of Standardization (ISO). (2010b). *Road vehicles - Symbols for controls, indicators, and tell-tales (ISO 2575:2010, IDT)*. Geneva, Switzerland: International Organization of Standardization (ISO).

International Organization of Standardization (ISO). (2017b). *Societal security - Emergency management - Guidelines for colour coded alerts (ISO 22324:2015, IDT)*. Geneva, Switzerland: International Organization of Standardization (ISO).

International Organization of Standarization (ISO). (2017a). *Road vehicles - Ergonomic aspects of transport information and control systems - Specifications and test procedures for in-vehicle visual presentation (ISO 15008:2017)*. Geneva, Switzerland: International Organization of Standarization (ISO).

Japan Automobile Manufacturers Association. (2004). *Guidelines for In-vechicle Display Systems - Version 3.0.*

Kern, D., & Schmidt, A. (2009). Design Space for Driver-based Automotive User Interfaces. *First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2009).* Essen, Germany.

Ko, A. (n.d.). *How to evaluate analytically*. Retrieved April 19, 2018, from Design Methods: https://faculty.washington.edu/ajko/books/design-methods/how-to-evaluate-analytically.html

Krupenia, S. (2015). *Task Focus -v3*. Södertälje, Sweden: Driver Vehicle Interaction, Scania CV AB.

Lindh, G. (2016a). *16122-200*. Retrieved May 23, 2018, from Scania media provider: https://media.scania.com/fotoweb/Preview.fwx?&position=28&archiveTy

pe=ImageFolder&archiveId=5106&albumId=5106&sorting=ModifiedTimeAsc&search=interior&fileId=C7BC39EB0244DBA34CF2FADA92307FF130607FBCB0030271854EF17105CF8A405D760AB5A2C8FDE94EA745C98C309131D2

Lindh, G. (2016b). *16122-201*. Retrieved May 23, 2018, from Scania media provider: https://media.scania.com/fotoweb/Preview.fwx?&position=29&archiveType=ImageFolder&archiveId=5106&albumId=5106&sorting=ModifiedTimeAsc&search=interior&fileId=C7BC39EB0244DBA34CF2FADA92307FF130607FBCB0030271854EF17105CF8A405D760AB5A2C8FDE94EA745C98C309131D2

Lucaites, K., Fletcher, B., & Pyle, A. (2017). Measuring the Impact of Affordance-Based Clickability Cues. *ACM Conference*, (p. 7). Washington DC, USA.

Magnusson, C., Rassmus-Gröhn, K., Tollmar, K., & Deaner, E. (2009). *User Study Guideline.* Lund, Sweden: Seventh Framework Programme.

Mahatody, T., Sagar, M., & Kolski, C. (2010). State of the Art on the Cognitive Walkthrough Method, Its Variants and Evolutions. *Journal of Human-Computer Interaction, 26*(8), 741-785.

Mahlke, S. (2007). User experience: usability aesthetics and emotions in human-technology interaction. (E. Law, A. Vermeeren, M. Hassenzahl, & M. Blythe, Eds.) *Towards a UX Manifesto*, 26-30.

Martin, B., & Hanington, B. (2012). *Universal Methods of Design.* Beverly, USA: Rockport Publishers.

McLeod, S. (2017). *Qualitative vs. Quantitative*. Retrieved April 19, 2018, from Simply Psychology: https://www.simplypsychology.org/qualitative-quantitative.html

Medium. (2016). *The Usability Checklist for Visual Design !* Retrieved April 26, 2018, from Medium: https://medium.com/the-mvp/the-usability-checklist-for-visual-design-9ca1ea44dc83

National Aeronautics and Space Administration. (2017). *NASA TLX: Task Load Index.* Retrieved March 29, 2018, from National Aeronautics and Space Administration: https://humansystems.arc.nasa.gov/groups/tlx/

Nielsen, J. (1993a). *Usability Engineering.* San Diego, USA: Academic Press.

Nielsen, J. (1993b). *Response Times: The 3 Important Limits*. Retrieved April 24, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/response-times-3-important-limits/

Nielsen, J. (1995a). *10 Usability Heuristics for User Interface Design*. Retrieved February 6, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/ten-usability-heuristics/

Nielsen, J. (1995b). *How to Conduct a Heuristic Evaluation.* Retrieved March 19, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation/

Nielsen, J. (2010). *Website Response Times*. Retrieved March 2, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/website-response-times/

Norman, D. (2013). *The Design of Everyday Things.* New York, USA: Basic Books.

Olsson, E. (2015). *Driver Vehicle Interaction Sub-properties - Affordance.* Södetälje, Sweden: Driver Vehicle Interaction, Scania CV AB.

Oztekin, A., James Kong, Z., & Uysal, O. (2010). UseLearn: A novel checklist and usability evaluation method for eLearning systems by criticality metric analysis. *International Journal of Industrial Ergonomics*, 455-469.

Pauzié, A. (2008). A method to assess the driver mental workload: The driving activity load index (DALI). *IET Intelligent Transport Systems, 2*(4), 315-322. Retrieved May 25, 2018, from http://ieeexplore.ieee.org.ludwig.lub.lu.se/stamp/stamp.jsp?tp=&arnumber=4706518

Persson, D. (2013). *Technical Product Data - Instruction, Vehicle property definitions.* Södertälje, Sweden: Scania CV AB.

Rogers, Y., Sharp, H., & Preece, J. (2002). *Interaction Design: Beyond Human - Computer Interaction.* West Sussex, United Kingdom: John Wiley and Sons Ltd.

Roto, V. (2007). User Experience from Product Creation Perspective. (E. Law, A. Vermereen, M. Hassenzahl, & M. Blythe, Eds.) *Towards a UX Manifesto*, 31-34.

Roto, V., Hassenzahl, M., Vermeeren, A., Law, E.-C., & Kort, J. (2009). *Understanding, Scoping and Defining User eXperience: A Survey Approach.* Boston, USA: CHI.

Rubin, J., & Chisnell, D. (2008). *Handbook of Usability Testing* (2nd ed.). Indianapolis, USA: Wiley Publishing, Inc.

Scania CV AB. (2016). *Scania Identity Manual.* Södertälje, Sweden: Scania CV AB.

Scania CV AB. (2017). *Radio manuals*. Retrieved May 16, 2018, from Scania Global: https://www.scania.com/global/en/home/products-and-services/articles/RadioManuals.html

Scania CV AB. (2018a). *About Scania*. Retrieved February 21, 2018, from https://www.scania.com/group/en/section/about-scania/

Scania CV AB. (2018b). *ICL3 Utilities*. Retrieved May 23, 2018, from Inline: https://inline.scania.com

Schade, A. (2016a). *7 Tips for Successful Customization*. Retrieved March 6, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/customization/?lm=first-principles-interaction-design&pt=article

Schade, A. (2016b). *Customization vs. Personalization in the User Experience.* Retrieved March 12, 2018, from Nielsen Norman Group: https://www.nngroup.com/articles/customization-personalization/

Shneiderman, B. (2016, May). *The Eight Golden Rules of Interface Design*. Retrieved February 5, 2018, from University of Maryland: https://www.cs.umd.edu/users/ben/goldenrules.html

Stanton, N. (2006). Hierarchical task analysis: Developments, applications, and extensions. *Applied Ergonomics, 37*, 55-79.

Stanton, N., & Harvey, C. (2013). *Usability Evaluation for In-Vechicle Systems.* Boca Raton, USA: Tailor & Francis Group, LLC.

Stanton, N., Salmon, P., Walker, G., Baber, C., & Jenkins, D. (2005). *Human Factors Methods: A Practical Guide for Engineering and Design.* Hampshire, United Kingdom: Ashgate Publishing Limited.

Stevens, A., & Cynk, S. (2011). *Checklist for the assessment of in-Vehicle information systems.* Transport Research Laboratory.

Surbhi, S. (2016). *Difference Between Qualitative and Quantitative Data*. Retrieved April 19, 2018, from Key Differences: https://keydifferences.com/difference-between-qualitative-and-quantitative-data.html

Sweeney, D., & Pritchard, M. (n.d.). *Semi-structured Interview*. Retrieved March 9, 2018, from Evaluation toolbox: https://www.evaluationtoolbox.net.au/index.php?option=com_content&view=article&id=31&Itemid=137

The de Bono Group. (n.d.). *Six Thinking Hats*. Retrieved April 24, 2018, from The de Bono Group: http://www.debonogroup.com/six_thinking_hats.php

Thomas, C., & Bevan, N. (1996). *Usability Context Analysis: A Practical Guide.* Teddington, United Kingdom: National Physical Laboratory.

Travis, D. (2009). *Measuring satisfaction: Beyond the usability questionnaire.* Retrieved March 29, 2018, from Userfocus: https://www.userfocus.co.uk/articles/satisfaction.html

Tullis, T., & Albert, B. (2013). *Measuring the User Experience - Collecting, Analyzing, and Presenting Usability Metrics* (2nd ed.). Waltham, USA: Elsevier Inc.

U.S. Department of Health & Human Services. (n.d.). *Heuristic Evaluations and Expert Reviews.* Retrieved March 20, 2018, from usability.gov: https://www.usability.gov/how-to-and-tools/methods/heuristic-evaluation.html

Ulrich, K., & Eppinger, S. (2012). *Product Design and Development* (5th ed.). London, United Kingdom: McGraw-Hill International Education.

Vermeeren, A., Cremers, A., Kort, J., & Fokker, J. (2008). Comparing UX Measurements, a case study. (E.-C. Law, N. Bevan, G. Christou, M. Springett, & M. Lárusdóttir, Eds.) *Proceedings of the International Workshop on Meaningful Measures: Valid Useful User Experience Measurement*, 72-78.

Wharton, C., Rieman, J., Lewis, C., & Polson, P. (1994). *The Cognitive Walkthrough Method: A Practitioner's Guide.* Boulder, Colorado: Institute of Cognitive Science, University of Colorado.

Wink, G. (2016a). *16049-049*. Retrieved May 23, 2018, from Scania media provider: https://media.scania.com/fotoweb/Preview.fwx?&position=30&archiveType=ImageFolder&archiveId=5106&albumId=5106&sorting=ModifiedTimeAsc&search=interior&fileId=C7BC39EB0244DBA34CF2FADA92307FF130607FBCB0030271854EF17105CF8A405D760AB5A2C8FDE94EA745C98C309131D2

Wink, G. (2016b). *16049-006*. Retrieved May 23, 2018, from Scania media provider: https://media.scania.com/fotoweb/Preview.fwx?&position=36&archiveType=ImageFolder&archiveId=5106&albumId=5106&sorting=ModifiedTimeAsc&search=interior&fileId=C7BC39EB0244DBA34CF2FADA92307FF130607FBCB0030271854EF17105CF8A405D760AB5A2C8FDE94EA745C98C309131D2

Wong, E. (n.d.). *Heuristic Evaluation: How to Conduct a Heuristic Evaluation.* Retrieved March 20, 2018, from Interaction Design Foundation: https://www.interaction-design.org/literature/article/heuristic-evaluation-how-to-conduct-a-heuristic-evaluation

Appendix A Interview protocols

This Appendix includes the frameworks used during interviews for defining contextual factors and criteria.

A.1 Defining contextual factors

Interview: Defining contextual factors

Participants

The participants are six employees, two from each of the teams; short-term, mid-term and long-term at the department of DVI. Three of the participants are interaction designers at Scania while the other three are UX design consultants currently working at Scania.

Description

Interviews will be conducted continuously throughout the thesis work. In order to evaluate the driver vehicle interaction experience the sub properties need to be broken down in to smaller elements. This will be done through the steps; *Defining contextual factors* and *Defining criteria.*

This interview will focus on the step *Defining contextual factors*. The aim of the interview is to understand which contexts of use that are most important to the driver vehicle interaction. The context of use is important in order to define what needs to be evaluated.

Material

- Mind map template containing the headlines; *Users, Physical and social environment*, *Tasks* and *Equipment*

Introduction

The interviewer describes the methodology of the thesis work and how the interviews will be carried out.

Questions

1. How would you describe the context of use for driver vehicle interaction in general? Which circumstances will affect the driver?
2. How would you describe the context of use for driver vehicle interaction in relation to the areas; Users, Tasks, Equipment and Physical and social environment? (Please write down and draw your answers on the mind map)

A.2 Defining criteria

Interview: Defining criteria

Participants

The participants are six employees, two from each of the teams; short-term, mid-term and long-term at the department of DVI. Three of the participants are interaction designers at Scania while the other three are UX design consultants currently working at Scania.

Description

Interviews will be conducted continuously throughout the thesis work. In order to evaluate the driver vehicle interaction experience the sub properties need to be broken down in to smaller elements. This will be done through the steps; *Defining contextual factors* and *Defining criteria.*

This interview will focus on the step *Defining criteria*. The aim of the interview is to match high level factors to contextual factors and then decide related criteria by analysing each contextual factors in detail.

Material

- Mind map template containing the hierarchy of contextual factors; *Users (Range of personal attributes* and *Range of product experience), Physical and social environment (Dual task environment, Different driving conditions* and *Personal environment)*, *Tasks (Primary/secondary tasks* and *Different segments)* and *Equipment (Hardware and software)*
- Definition of high level factors (*Pleasurability, Task focus* and *Flexibility*)

Introduction

The interviewer describes the methodology of the thesis work and how the interviews will be carried out. The interviewer then describes the results from the interview regarding contextual factors; the chosen contextual factors and their meaning. The interviewer also describes the definition of the high level factors.

Questions

1. Do you think something is missing regarding the contextual factors?
2. Which high level factors do you believe have the greatest impact on each specific contextual factor? Describe how you think. (Please write down and draw your answers on the mind map)
3. Which are the criteria for each contextual factor and related high level factor? (Please write down and draw your answers on the mind map)

Appendix B Test protocols

This Appendix includes the frameworks used during the testing of Inspection-based evaluation and Systematic Human Error Reduction and Prediction Approach as well as the Hierarchical Task Analysis used as basis during the tests.

B.1 Inspection-based evaluation

Test: Inspection-based evaluation

Participants

The participants are six employees from the department of DVI. Three of the participants are interaction designers at Scania while the other three are UX design consultants currently working at Scania.

Description

This test is carried out in order to evaluate an evaluation method for driver vehicle interaction. The evaluation method to be tested is an Inspection-based evaluation, which is one of the developed evaluation methods and does not cover all aspects to be evaluated. Furthermore, only some part of the framework will be tested.

The aim of the test is to assess the evaluation items in order to improve the framework. The main focus will be the understanding of instructions, scoring system and evaluation items as well as the content.

Material

- Inspection-based evaluation framework
- ICL3 prototype
- Hierarchical Task Analysis: Programmable fields

Introduction

Start by reading the instructions and then score and comment on the evaluation items for the prototype based on the given HTA. For each item you will answer test questions about the Inspection-based evaluation.

Questions

1. Do you understand the instructions? Describe how you interpret them. Should the instructions be changed somehow?
2. Do you understand how you should score the evaluation items? How did it feel to score the items? Should the scoring system be changed somehow?
3. Do you understand the evaluation item? Describe how you interpret it. Should the item be changed somehow? (Repeat for each item)
4. Do you believe that any important parts are missing?
5. Do you have any recommendations for improvement?
6. What is your overall impression of the Inspection-based evaluation?
7. What do you think about the layout?

B.2 SHERPA

Test: SHERPA

Participants

The participants are six employees from the department of DVI. Three of the participants are interaction designers at Scania while the other three are UX design consultants currently working at Scania.

Description

This test is carried out in order to evaluate an evaluation method for driver vehicle interaction. The evaluation method to be tested is called SHERPA (Systematic Human Error Reduction and Prediction Approach, which is one of the developed evaluation methods and does not cover all aspects to be evaluated.

The aim of the test is to assess the SHERPA framework in order to improve the evaluation method. The main focus will be the understanding of instructions and filling out the SHERPA template.

Material

- SHERPA framework
- ICL3 prototype
- Hierarchical Task Analysis: Programmable fields

Introduction

Start by reading the instructions and the steps of the method as well as the example. Then go through all steps of the method based on the given HTA. For each step you will answer test questions about the SHERPA.

Questions

1. Do you understand the instructions? Describe how you interpret them. Should the instructions be changed somehow?
2. Do you understand the steps of the method? Describe how you interpret them.
3. Do you understand how to categorize/answer this? Describe how you interpret them. (repeat for each step)
4. Do you believe that any important aspects are missing?
5. Do you have any recommendations for improvement?
6. What is your overall impression of the SHERPA?
7. What do you think about the layout?

Appendix C Evaluation frameworks

This Appendix includes the developed evaluation frameworks for Hierarchical Task Analysis, Inspection-based evaluation, Systematic Human Error Reduction and Prediction Approach as well as Cognitive Walkthrough.

HTA

(HIERARCHICAL TASK ANALYSIS)

SCANIA

Evaluator(s):

Task goal:

AIM

The aim of this method is to describe tasks in terms of hierarchy of goals, sub-goals, operations and plans. The upper levels of the hierarchy include goals and sub-goals while the bottom level describes what operations needs to be accomplished in order to reach the sub-goals. Furthermore, the plans indicate how the goals are achieved.

This method is often required, or can be applied as a basis, for other evaluation methods that evolve around tasks.

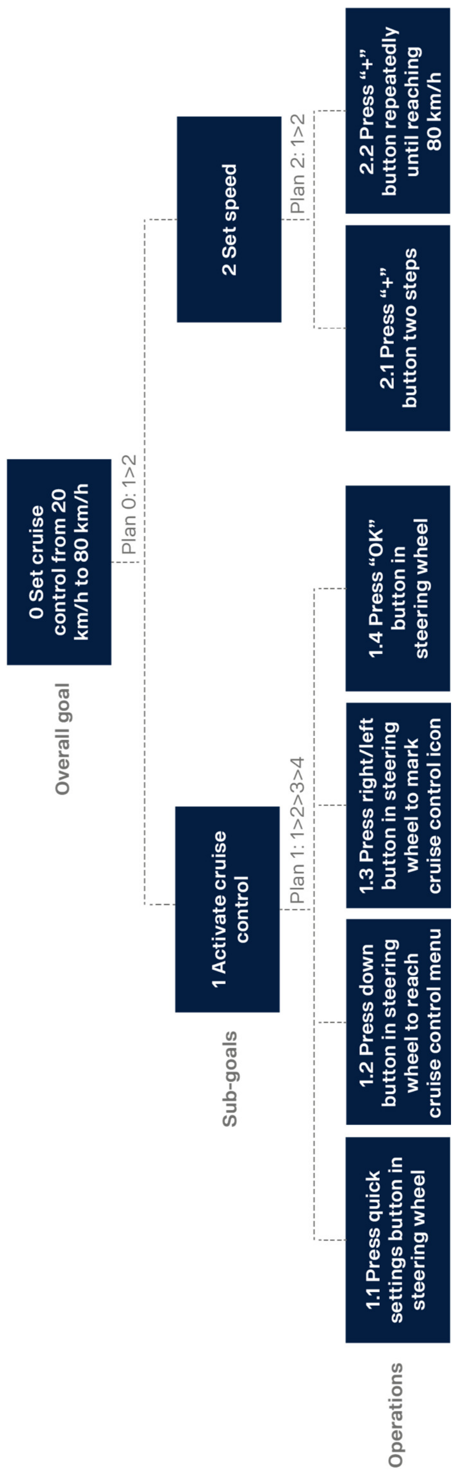
INSTRUCTIONS

Conduct the HTA according to the steps below and illustrate it as a tree diagram. The attached example can be used as guidance when conducting the HTA.

The task to be analyzed should be relevant to the specific Key Performance Indicator* and prototype to be evaluated. Preferably tasks with higher criticality and/or frequency of use should be considered. The interaction elements (physical and/or digital) included in the prototype to be evaluated should be determined before the evaluation can start.

* The Key Performance Indicators are based on the sub-properties and can be found in the thesis work (Development of an Evaluation Method for Driver Vehicle Interaction), chapter 4.4.

EXAMPLE



Overall goal
0 Set cruise control from 20 km/h to 80 km/h
Plan 0: 1>2
Sub-goals
1 Activate cruise control
2 Set speed
Plan 1: 1>2>3>4
Plan 2: 1>2
Operations
1.1 Press quick settings button in steering wheel
1.2 Press down button in steering wheel to reach cruise control menu
1.3 Press right/left button in steering wheel to mark cruise control icon
1.4 Press "OK" button in steering wheel
2.1 Press "+" button two steps
2.2 Press "+" button repeatedly until reaching 80 km/h

METHOD

1. Define the task to be analyzed.

2. Collect data about the task to be analyzed (through interviews, observations, questionnaires, actual usage of product).

3. Determine the overall goal of the task (upper level of the hierarchy).

4. Break down the overall goal into sub-goals (second level of the hierarchy).

5. Break down the sub-goals into operations which are the actions needed to achieve the goal/sub-goals (lower level of the hierarchy).

6. Make a plan of how the goals should be achieved through determining the order of which the sub-goals and operations should be executed.

    a. Linear plan: Do the steps in a sequence (1>2>3>4)
    b. Non-linear plan: Do the steps in any order (1/2/3/4)
    c. Simultaneous plan: Do the steps at the same time (1+2+3+4)
    d. Branching plan: Do the steps when required (X? Y>2 N>3)
    e. Repetitious plan: Repeat the steps (1>2>3>4>1…)
    f. Selection plan: Choose one of the following steps (1:2:3:4)

CW

(COGNITIVE WALKTHROUGH)

SCANIA

Evaluator(s):

Concept:

Task goal:

AIM

The aim of this evaluation method is to discover usability issues related to first time usage of systems. It should be conducted by either one or a group of evaluators with experience and knowledge of UX, usability and human factors as well as the system to be evaluated. The method can be used at any time during the design process and on various prototypes. The output of the evaluation method is a suggested improvement strategy.

This evaluation method covers the Key Performance Indicator*; "Novice users should be able to handle essential features and functions".

The evaluator(s) should consider first time users with no previous experience of the system to be evaluated.

INSTRUCTIONS

A Hierarchical Task Analysis (HTA) related to the prototype to be evaluated must be conducted before the evaluation can start.

Walk through the HTA and answer the related questions according to the method below. The attached example can be used as guidance when conducting the evaluation. It is recommended to record the evaluation session so that it is possible to go back and further analyze the discussions.

* The Key Performance Indicators are based on the sub-properties and can be found in the thesis work (Development of an Evaluation Method for Driver Vehicle Interaction), chapter 4.4.

METHOD

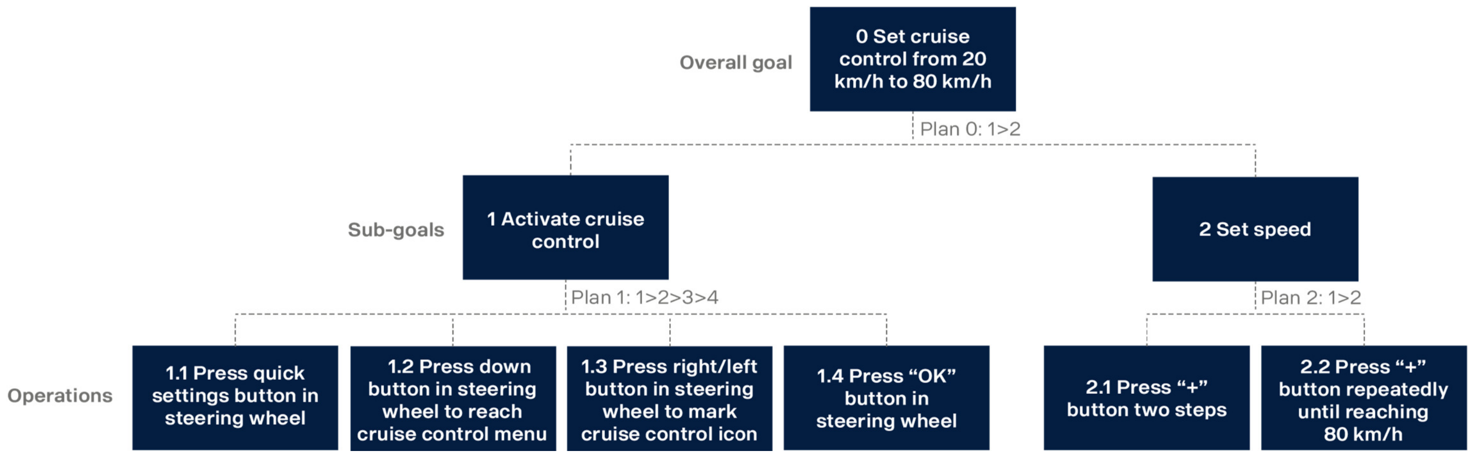
1. Walk through the action sequences of each task and answer the following questions for each task step (sub-goals or operations of the HTA). The questions should be answered with a credible description which considers the user's success or failure of each question.

    a. Will the user know what needs to be achieved?
    b. Will the user notice that the function is available?
    c. Will the user associate the cues with the function?
    d. Will the user get feedback when using the function?
    e. Will the user get feedback to understand that the task has been performed?

2. Propose improvement strategies.

EXAMPLE

Task step	Will the user know what needs to be achieved?	Will the user notice that the function is available?	Will the user associate the cues with the function?	Will the user get feedback when using the function?	Will the user get feedback to understand that the task has been performed?	Improvement strategy
1.1	Failure: The user will not know that the quick settings button is associated with activating the cruise control	Success: If the user associates the quick settings button with activating cruise control, the user will notice the settings button since it is located in the steering wheel	Success: The user will recognize the settings symbol since this symbol is commonly used in other applications	Success: The menu window appears on the ICL when pressing the settings button	Success: The user will get visual feedback on the ICL when pressing the settings button	The user should be informed that the quick settings button is associated with activating cruise control
1.2	Failure: The user will not know that he/she needs to press down since there is no indication that the cruise control menu is below the “Performance Mode” menu	Success: If the user knows that the cruise control menu is below the “Performance Mode” menu, he/she will notice it when pressing down button (the user will notice the down button since it is located in the steering wheel)	Success: The user will associate the cruise control symbols in the ICL and the down button with the function since they are standard symbols	Success: The “ Cruise Control” menu bar is shown when the user presses the down button	Success: The “ Cruise Control” menu bar is shown when the user presses the down button	The “Cruise Control” menu bar should be visible from the beginning
1.3	Success: The user will understand that he/she needs to press right to mark the cruise control symbol	Success: The user will notice the cruise control symbol	Success: The user will associate the cruise control symbol with the function since it is a standard symbol	Success: The cruise control symbol lights up in green when pressing the right button	Success: The cruise control symbol lights up in green when pressing the right button	
1.4	Success: The user will understand that he/she needs to press “OK”	Success: The user will notice the “OK” button since it is located in the steering wheel	Success: The user will associate “OK” with accepting the choice since it is commonly used in other applications	Success: The “Cruise Control” menu bar dissapears when pressing “OK”	Failure: The “Cruise Control” menu bar dissapears when pressing “OK” but the user will not get further confirma-tion that the function is activated	The user should get further confirmation that the cruise control is activated
2.1	Success: The user will understand that he/she needs to press the “+” button	Success: The user will notice the “+” button since it is located in the steering wheel	Success: The user will associate “+” next to the cruise control symbol in the steering wheel with increasing the speed	Success: The cruise control speed indicator and the speedometer will increase when pressing the “+” button	Success: The cruise control speed indicator and the speedometer will increase when pressing the “+” button:	
2.2	Success: The user will understand that he/she needs to press the “+” button repeatedly to reach the desired speed	Success: The user will notice the “+” button since it is located in the steering wheel	Success: The user will associate “+” next to the cruise control symbol in the steering wheel with increasing the speed	Success: The cruise control speed indicator and the speedometer will increase when pressing the “+” button	Success: The cruise control speed indicator and the speedometer will increase when pressing the “+” button:	


Overall goal
0 Set cruise control from 20 km/h to 80 km/h
Plan 0: 1>2
Sub-goals
1 Activate cruise control
2 Set speed
Plan 1: 1>2>3>4
Plan 2: 1>2
Operations
1.1 Press quick settings button in steering wheel
1.2 Press down button in steering wheel to reach cruise control menu
1.3 Press right/left button in steering wheel to mark cruise control icon
1.4 Press “OK” button in steering wheel
2.1 Press “+” button two steps
2.2 Press “+” button repeatedly until reaching 80 km/h

TEMPLATE

Task step	Will the user know what needs to be achieved?	Will the user notice that the function is available?	Will the user associate the cues with the function?	Will the user get feedback when using the function?	Will the user get feedback to understand that the task has been performed?	Improvement strategy

INSPECTION-BASED EVALUATION

SCANIA

Evaluator(s):

Concept:

Task goal:

AIM

The aim of this evaluation method is to provide a structured approach for evaluating the usability and design of a specific interface. It should be conducted by evaluators with experience and knowledge of UX, usability and human factors as well as the system to be evaluated. It should preferably be conducted in a group or by several evaluators separately. The method can be used at any time during the design process and on various prototypes. The output is a score which can be used in the overall evaluation of the driver vehicle interaction as well as comments on design improvements.

INSTRUCTIONS

A Hierarchical Task Analysis (HTA) related to the prototype to be evaluated must be conducted before the evaluation can start.

The evaluation is divided in to parts, which refer to different Key Performance Indicators* to be evaluated. Each part is further divided into different evaluation areas with series of items to be analysed.

Items marked task dependent (D) should be analysed when walking through the HTA. Questions marked task independent (I) should be analysed by interacting freely with the prototype.

Supportive information (marked i, ii, iii, etc.) on how to analyse each item should be considered. Items considering several aspects should be given an average score.

Each item should be scored according to a 5 point scale. As a compliment the importance of the items should be specified according to an importance scale. For the questions that cannot be answered one can choose not applicable (N/A). Additionally, positive and/or negative comments on the design and improvements should be added to items when needed.

EXAMPLE

The system layout is presented in a consistent manner. (**I**)

i. Information (colour, text, icons and symbols) of related functions should be arranged similarly

ii. The same amount of information (colour, text, icons and symbols) of related functions should be given

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

* The Key Performance Indicators are based on the sub-properties and can be found in the thesis work (Development of an Evaluation Method for Driver Vehicle Interaction), chapter 4.4.

PART I

This part of the evaluation covers the Key Performance Indicator; "All types of users should be able to interact with the system as well as understand functions and their impact".

The evaluator should consider users with different cultural and biological backgrounds as well as physical and psychological differences when analysing the items.

EVALUATION AREAS

1 System

System refers to the functions, layout and menu structure in general.

2 Visual information

Visual information refers to text, icons and symbols and colour within the system.

3 Auditory information

Auditory information refers to voice messages and auditory signals within the system.

4 Tactile information

Tactile information refers to haptic signals within the system.

5 Controls

Controls refer to both physical (buttons, knobs, switches, etc.) and digital (sliders, checkboxes, etc.) controls that enable user input within the system.

1 SYSTEM

Functions are visible. (**D**)

i. The right amount of information (colour, text, icons and symbols) should be presented to make the functions visible

ii. Colour, text, icons and symbols should be distinctive to make functions visible

iii. The layout should be distinctive to make functions visible

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Menu navigation is easy to use. (**D**)

i. It should be easy to navigate through menus, sub-menus and options

ii. Feedback should be provided that tells the users where in the menu structure they are

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The system provides clear feedback (visual, auditory and haptic information) to user input. (**D**)

i. The feedback should be understandable and informative

ii. It should be possible to read information before it disappears

iii. It should be possible to continue an interrupted action after being distracted

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The system provides appropriate feedback (visual, auditory and haptic information) to user input. (**D**)

i. The feedback should be suitable for the information given and proportional to the user input (avoid unnecessary amount of feedback)

ii. The feedback should not be disturbing

iii. The feedback response time should be appropriate

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The system layout is presented in a consistent manner. (**I**)

i. Similar functions should be presented with the same amount of information (colour, text, icons and symbols)

ii. Information (colour, text, icons and symbols) of similar functions should be arranged similarly

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

The mapping of the system is appropriate. (**I**)

i. The layout of functions should correspond to a natural and suitable arrangement

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

2 VISUAL INFORMATION

2.1 Text

The meaning of text is clear. (**D**)

i. The information given should be concise and understandable

ii. The information should enable users to proceed to the following action

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Words and abbreviations are appropriate. (**I**)

i. Words and abbreviations should be used correctly

ii. Note: Requires final approval by Scania's linguistic team

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Words and abbreviations conform to international/national standards. (**I**)

i. Words and abbreviations should match the selected language

ii. Words and abbreviations should conform to the area of application

iii. Note: Requires final approval by Scania's linguistic team

Score:

Strongly disagree	Disagree	Neutral	Agree	Strongly agree	N/A
○	○	○	○	○	○

Importance:

Low	Medium	High	N/A
○	○	○	○

Comment:

Numbers and units conform to international/national standards. (**I**)

i. Numbers and units should match the selected language (i.e. metric system, imperial system, etc.)

Score:

Strongly disagree	Disagree	Neutral	Agree	Strongly agree	N/A
○	○	○	○	○	○

Importance:

Low	Medium	High	N/A
○	○	○	○

Comment:

Text ensures readability. (**I**)

i. An easily readable font should be used (a humanist type is recommended)

ii. The appropriate font size should be used according to ISO 15008[1]

The appropriate letter height is calculated through $x = d \cdot \alpha_R$,
where $\alpha_R = 6.98 \cdot 10^{-3}$ rad (recommended) or $\alpha_R = 4.36 \cdot 10^{-3}$ rad (minimum)
(x = letter height [mm], d = distance between eye and letter [mm], α_R = angular dimension [rad])

AaBbCcDd
1234567890.,–+!()[]

Recommended angular dimension for d = 1000 mm

AaBbCcDd
1234567890.,–+!()[]

Minimum angular dimension for d = 1000 mm

iii. Font size should be appropriate to screen size and resolution

iv. The text should be clear without the use of colour coding and/or auditory information

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Text is presented in a consistent manner. (**I**)

i. Same font type should be used for information with similar purpose

ii. Same font size should be used for information with similar purpose

iii. If the same information is presented several times throughout the system, the same words and abbreviations, in terms of grammar and vocabulary as well as sentence construction, should be used

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

[1] Road vehicles – Ergonomic aspects of transport information and control systems – Specifications and test procedures for in-vehicle presentation (ISO 15008:2017)

2.2 Icons and symbols

The meaning of icons and symbols is clear. (**D**)

i. The information given should be understandable

ii. The information should enable users to proceed to the following action

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Icons and symbols conform to international/national standards and norms. (**I**)

i. Icons and symbols that are presented in ISO 2575[2] should be used when possible

ii. Icons and symbols should conform to the area of application

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

[2] Road vehicles – Symbols for controls, indicators and tell-tales (ISO 2575:2010, IDT)

It is possible to perceive symbols. (**I**)

i. Smallest icon size used on screens should not be lower than 32x32 px and smallest icon size used on physical controls should not be lower than 10x10 mm

ii. Icon size should be appropriate to screen size and resolution

iii. Icons and symbols should be clear without the use of colour coding and/or auditory information

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Icons and symbols are presented in a consistent manner. (**I**)

i. If the same information is presented several times throughout the system, the same icon or symbol should be used

ii. Physical and digital icons and symbols should correspond throughout the system

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

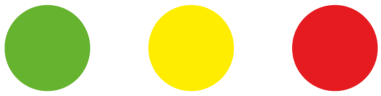
Comment:

2.3 Colour

The colour coding is clear. (**I**)

i. Colours should represent particular meanings

ii. Colours to express dangers should be used according to ISO 22324[3]



Safe Caution Danger

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The colour layout is clear. (**I**)

i. Colour combinations should be conform to ISO 15008[4] (best colour combinations are white and black, white and blue as well as yellow and black)

White text on black | Black text on white | White text on blue | Blue text on white | Black text on yellow | Yellow text on black

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

[3] Societal security – Emergency management – Guidelines for colour coded alerts (ISO 22324:2015, IDT)

[4] Road vehicles – Ergonomic aspects of transport information and control systems – Specifications and test procedures for in-vehicle presentation (ISO 15008:2017)

Colour supports vision impairments. (**I**)

i. Colour coding and layout should not be affected by colour blindness

Insert pictures of the prototype interface to Adobe Illustrator. Convert to RGB colour mode. Choose View > Proof Setup > Colour Blindness > Protanopia-type or Deuteranopia-type.

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

Colour is limited to fixed and distinguished sets. (**I**)

i. Colours should not be similar to each other

ii. No more than five different colours should be used at the same time

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

Colour is presented in a consistent manner. (**I**)

i. If the same information is presented several times throughout the system, the same colour should be used

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

3 AUDITORY INFORMATION

3.1 Voice messages

The meaning of the voice messages is clear. (**D**)

i. The information given should be concise and understandable

ii. The message should enable users to precede to the following action

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

It is possible to perceive the voice messages. (**D**)

i. The information should be presented at an appropriate pitch and speed

ii. The volume should be 15 dB higher than the background noise

Note: Might require tests in listening studio

iii. The language and dialect of voice messages should correspond to user preferences

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Voice messages are presented in a consistent manner. (**I**)

i. If the same information is presented several times throughout the system, the same words and abbreviations, in terms of grammar and vocabulary as well as sentence construction, should be used

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

3.2 Auditory signals

The meaning of the auditory signals is clear. (**D**)

i. Signals should relate to the information given and/or be appropriate for the specific action

ii. The signal should enable users to precede to the following action

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

It is possible to perceive the auditory signals. (**D**)

i. The information should be presented at an appropriate speed

ii. The volume should be 15dB higher than the background noise (especially important for warnings)

Note: Might require tests in listening studio

iii. The direction of the signal should be clear

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Auditory signals are presented in a consistent manner. (**I**)

i. If the same information is presented several times throughout the system, the same auditory signal should be used

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

4 TACTILE INFORMATION

The meaning of the tactile signals is clear. (**D**)

i. Tactile signals should relate to the information given and/or be appropriate for the specific action

ii. Tactile signals should enable users to precede to the following action

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Tactile signals are presented in a consistent manner. (**I**)

i. If the same information is presented several times throughout the system, the same related tactile signal should be used

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

5 CONTROLS

The intended use of the controls is clear. (**D**)

i. Size, shape, material and motion of physical controls should be adapted to match functionality (push, turn, flip, etc.)

ii. Digital controls should give clickability cues through depth and/or colour

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The control actions conform to international/national norms. (**I**)

i. Buttons should be pressed, checkboxes should be marked, sliders should be dragged, switches should be flipped, knobs should be turned, etc.

ii. The control actions should be suitable for the function (e.g. sliders for increasing/decreasing volume)

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The layout of controls is clear. (**I**)

i. Controls should be located close to related functions

ii. The layout of controls should correspond to the user's sequence of action

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Controls are presented in a consistent manner. (**I**)

i. Controls with similar purpose should correspond throughout the system

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

PART II

This part of the evaluation covers the Key Performance Indicator; "Users should be able to recover from wrong actions".

EVALUATION AREAS

1 Error prevention

Error prevention refers to the system's ability to prevent errors from occurring.

2 Input error

Input error refers to an error made by the user and how the system handles it.

3 Change of input

Change of input refers to the possibility to reverse or cancel an input made by the user.

1 ERROR PREVENTION

The system prevents possible errors from occurring. (**D**)

i. Error or warning messages should be shown before a potential input error is made

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The system provides reconfirmation before a potential input error is made. (**D**)

i. The system should ask the user if it wants to continue with the action

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

2 INPUT ERROR

Appropriate feedback (visual, auditory, haptic information) is provided when an input error is made. (**D**)

i. The feedback should be proportional to the seriousness of the error

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Immediate feedback (visual, auditory, haptic information) is provided when an input error is made. (**D**)

i. Response time should not be lower than 0.1 second

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The system provides advice on how to recover from an input error. (**D**)

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

3 CHANGE OF INPUT

It is possible to reverse an input. (**D**)

i. It should be possible to easily go back to the previous menu page

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

It is possible to cancel an input. (**D**)

i. It should be easily recognizable how to cancel an input

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

PART III

This part of the evaluation covers the Key Performance Indicator; “Task structure and functions should support safety in difficult driving conditions”.

The evaluator should consider difficult driving conditions that may occur due to surroundings in terms of weather conditions, road conditions and traffic environment along with other factors such as time of day, location and traffic regulations when analysing the items. Furthermore, different segments should be considered. Depending on if it a long-haulage, construction or distribution truck, or city bus or coach, the conditions can differ.

The system does not require attention that affect the driving. (**D**)

i. Visual information should not force the user to look away from the road
ii. Auditory information should not cancel out important sounds from outside the vehicle
iii. Tactile information does not cancel out important vibrations from outside the vehicle
iv. The system should not require time critical responses

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The system does not require the driver to complete an action. (**D**)

i. The system should not require long and uninterruptible sequences of actions
ii. It should be possible to resume an interrupted action

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Information is provided in an appropriate way. (**D**)

i. Information should be given at an appropriate point in time

ii. Information that is likely to distract the driving should be disabled

iii. Information with higher safety relevance should be prioritized

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

PART IV

This part of the evaluation covers the Key Performance Indicator; "Users should be able to distinguish and localise functions in varying driving conditions".

The evaluator should consider different surroundings in terms of weather conditions, road conditions and traffic environment along with other factors such as time of day, location and traffic regulations when analysing the items. Furthermore, different segments should be considered. Depending on if it a long-haulage, construction or distribution truck, or city bus or coach, the conditions can differ.

EVALUATION AREAS

1 Visual surroundings

Visual surroundings refers to how the surroundings might affect visual information.

2 Auditory surroundings

Auditory surroundings refers to how the surroundings might affect auditory information.

3 Tactile surroundings

Tactile surroundings refers to how the surroundings might affect tactile information.

1 VISUAL SURROUNDINGS

Displays are free from reflection and glare during all lightning conditions. (**I**)

i. It should be possible to interact with the displays in different weather conditions and geographical areas (urban or countryside)

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The display conditions (luminous intensity, contrast and colours) do not affect the vision during night. (**I**)

i. The lights should not be dazzling

ii. The lights should not be distracting

iii. The lights do not cause visual after effects

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Controls (digital and physical) are visible during both day and night. (**I**)

i. It should be possible to distinguish the controls when it is either light or dark

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Symbols and icons are visible during both day and night. (**I**)

i. It should be possible to interpret symbols and icons when it is either light or dark

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Text is visible during both day and night. (**I**)

i. It should be possible to read text when it is light or dark

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Colour coding is not affected by day and night conditions. (**I**)

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

2 AUDITORY SURROUNDINGS

It is possible to perceive auditory information regardless of surrounding noise. (**I**)

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

3 TACTILE SURROUNDINGS

It is possible to distinguish tactile information regardless of surrounding vibrations. (**I**)

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

It is possible to use controls (physical and digital) regardless of surrounding vibrations. (**I**)

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

PART V

This part of the evaluation covers the Key Performance Indicator; “The performance of primary tasks should not be compromised by secondary tasks”.

The evaluator should consider that there is a dual task environment where the primary tasks are related to driving and secondary tasks are related to comfort and administrative work when analysing the items.

The system does not force the user to perform secondary tasks when performing primary tasks. (**D**)

i. Secondary tasks should not be required when performing primary tasks

ii. Functions not intended to be used while driving should be impossible to interact with during driving

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

Secondary tasks are easy to use. (**D**)

i. Secondary information should only require a few glances

ii. The total task duration should not exceed 8 seconds

iii. Tasks should preferably require a maximum of three inputs

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The system does not require use of both hands simultaneously. (**D**)

i. When performing secondary tasks the driver should be able to keep at least one hand on the steering wheel

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

Secondary functions do not distract the driver. (**D**)

i. Dynamic or unpredictable information (visual, auditory and tactile) that can distract the driver should be disabled

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The layout of secondary functions is appropriate. (**I**)

i. Secondary functions should not obstruct the driver's visual field

ii. Secondary functions should not interfere with primary functions

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

PART VI

This part of the evaluation covers the Key Performance Indicator; “Various paths through the system should be provided to the user that match user attributes and product experience”.

The evaluator should consider users with different values, preferences, attitudes and technical interests toward the product as well as users with a range of product experience when analysing the items.

There are several ways to accomplish the same task. (**D**)

i. Alternative interaction paths should be taken into account (accelerators, shortcuts, etc.)

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The alternative interaction paths are appropriate. (**D**)

i. The interactions steps should be understandable

ii. Alternative interaction paths should only be available where needed

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The layout of alternative interaction paths is clear. (**D**)

i. Alternative interaction paths should be visible within the system

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

PART VII

This part of the evaluation covers the Key Performance Indicators; "Users should be able to customize the system for their specific segment tasks before use", "All types of users should be able to customize functions during use" and "Tasks should not be negatively affected by customization"

The evaluator should consider users with different values, preferences, attitudes and technical interests toward the product as well as users within different segments and with a range of product experience when analysing the items.

EVALUATION AREAS

1 Customization before use

Customization before use refers to the ability to customize which functions should be included in the system according to user preferences.

2 Customization during use

Customization during use refers to the ability to customize functions within the system.

1 CUSTOMIZATION BEFORE USE

It is possible to customize the system for specific segment needs. (**I**)

i. It should be possible to select which functions should be included in the system in order to meet specific user preferences

ii. It should be possible to change the layout of functions within the system in order to simplify the interaction of specific user preferences

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

The system works properly when using customized functions. (**I**)

Score:

Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A ○

Importance:

Low ○ Medium ○ High ○ N/A ○

Comment:

2 CUSTOMIZATION DURING USE

The customization options are visible. (**D**)

i. They should be positioned near the content they relate to

ii. The customization options should be well-named

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The customization functions are easy to use. (**D**)

i. The functions should not be time consuming

ii. The functions should not be complex

Score:

○ Strongly disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly agree ○ N/A

Importance:

○ Low ○ Medium ○ High ○ N/A

Comment:

The customization contributes to the interaction. (**D**)

i. There should be a purpose for customization function

ii. The customization function should help the user

Score:

Strongly disagree — Disagree — Neutral — Agree — Strongly agree — N/A

Importance:

Low — Medium — High — N/A

Comment:

It is easy to change previously customized settings. (**D**)

i. It should be clear how to edit customized settings

Score:

Strongly disagree — Disagree — Neutral — Agree — Strongly agree — N/A

Importance:

Low — Medium — High — N/A

Comment:

It is possible to perform tasks without customization. (**D**)

i. All functions should be possible to use without customization

Score:

Strongly disagree — Disagree — Neutral — Agree — Strongly agree — N/A

Importance:

Low — Medium — High — N/A

Comment:

PART VIII

This part of the evaluation covers the Key Performance Indicators; “System quality should match user expectations throughout the usage”.

System performance is high. (**D**)

i. The response time should not be higher than 0.1 seconds
ii. The resolution, colours, etc. of displays should be suitable
iii. Sound quality should be suitable
iv. Tactile quality should be suitable
v. All functions and controls should work as intended

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The reliability of the system is high. (**D**)

i. The probability of system malfunction should be low
ii. The durability of the system should be long, in terms of product life

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

System characteristics conform to established system standards. (**I**)

i. The system should follow technology development and innovation in the branch of industry as well as other industries

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The serviceability is high. (**D**)

i. It should be easy to repair the system

ii. It should be clear when the system needs service

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

The aesthetics of the system is appealing. (**D**)

i. The design should follow Scania's vision and niche

ii. The aesthetics should be appealing to all types of users

Score:

Strongly disagree | Disagree | Neutral | Agree | Strongly agree | N/A

Importance:

Low | Medium | High | N/A

Comment:

SHERPA

(SYSTEMATIC HUMAN ERROR REDUCTION AND PREDICTION APPROACH)

SCANIA

Evaluator(s):

Concept:

Task goal:

AIM

The aim of this evaluation method is to predict potential human or design initiated errors and related recovery potential. It should be conducted by evaluators with experience and knowledge of UX, usability and human factors as well as the system to be evaluated. The method can be used at any time during the design process and on various prototypes. The output of the evaluation method is a suggested improvement strategy.

This evaluation method covers the Key Performance Indicator*; "Users should be able to recover from wrong actions".

INSTRUCTIONS

A Hierarchical Task Analysis (HTA) related to the prototype to be evaluated must be conducted before the evaluation can start.

Walk through the HTA and identify the possible erroneous actions that the user is likely to perform for each task step (bottom level operation of HTA). Classify these according to the method below and fill out the template. The attached example can be used as guidance when conducting the evaluation.

* The Key Performance Indicators are based on the sub-properties and can be found in the thesis work (Development of an Evaluation Method for Driver Vehicle Interaction), chapter 4.4.

METHOD

1. Classify each task step into the following task types (several task types can be selected for each task step).

A	Action task: the user interacts with physical and/or digital controls through for example pressing button, turning knob, swiping and scrolling
R	Retrieval task: the user obtains and interprets information from the system
C	Checking task: the user does a procedural check such as checking system status
S	Selection task: the user chooses one alternative over another
I	Information communication task: the user communicates with the system through for example talking and entering numbers or letters

2. Identify error types that could occur for each task step into the following categories (several error types can be selected for each task type).

   a. Action errors: errors that can occur when the user aims to execute actions through for example pressing button, turning knob, swiping and scrolling

A1	The user executes the action during a too long/short time
A2	The user executes the action too little/much
A3	The user mistimes the action
A4	The user executes the action in the wrong direction
A5	The user misaligns the action
A6	The user executes the right action on a wrong object
A7	The user executes a wrong action on the right object
A8	The user executes a wrong action on a wrong object
A9	The user does not complete the action
A10	The user does not execute the action

   b. Retrieval errors: errors that can occur when the user should obtain and interpret information from the system

R1	The user does not obtain the information
R2	The user obtains the wrong information
R3	The user does not obtain the complete information

   c. Checking errors: errors that can occur when the user aims to conduct a procedural check such as checking system status

C1	The user mistimes the check
C2	The user executes the right check on a wrong object
C3	The user executes a wrong check on the right object
C4	The user executes a wrong check on a wrong object
C5	The user does not complete the check

C6 The user does not execute the check

d. Selection errors: errors that can occur when the user aims to choose one alternative over another

S1 The user makes a wrong selection

S2 The user does not execute the selection

e. Information communication errors: errors that can occur when the user aims to communicate with the system through for example talking and entering numbers or letters

I1 The user does not communicate the information

I2 The user communicates wrong information

I3 The user does not complete the communication

3. Describe the identified errors.

4. Describe the consequences that can be associated with the identified errors, which will affect the completion of the task.

5. Determine the recovery possibility for the identified errors.

6. Rate the probability of the error types occurring based on the following scale.

   a. Low probability (L)
   b. Medium probability (M)
   c. High probability (H)

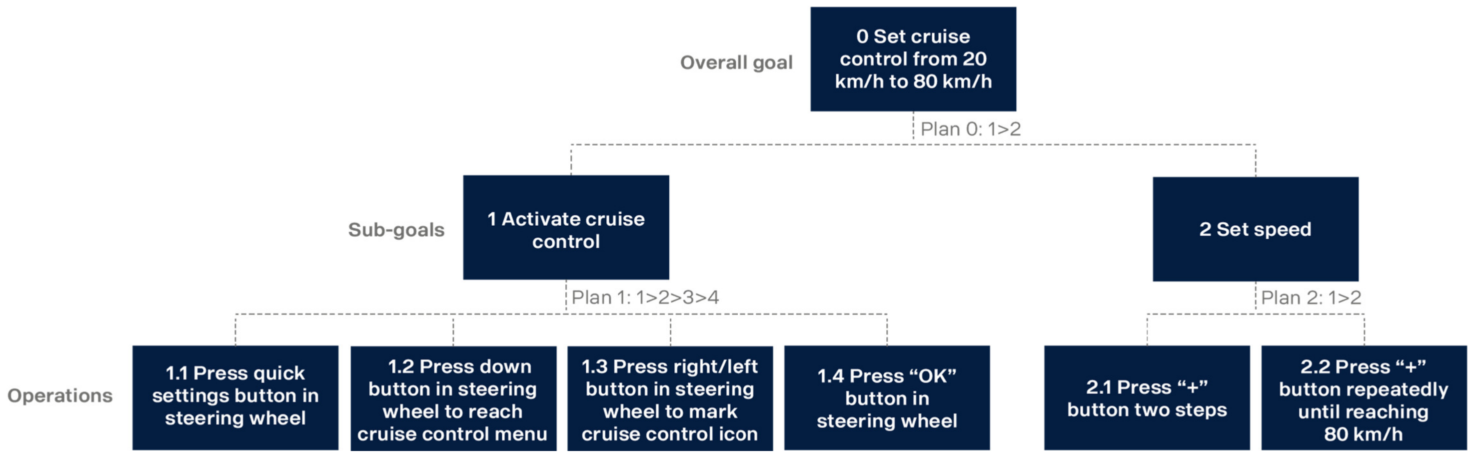
7. Rate the criticality of the errors types, in terms of completing the task, based on the following scale.

   a. Low criticality (L)
   b. Medium criticality (M)
   c. High criticality (H)

8. Propose improvement strategies that reduce errors and enable recovery possibilities.

EXAMPLE

Task step	1. Task type	2. Error type	3. Error description	4. Consequence	5. Recovery possibility	6. Probability (L/M/H)	7. Criticality (L/M/H)	8. Improvement strategy
1.1	A	A6	Press wrong button	Cannot proceed	Press another button	Medium	Medium	Better icons and placement of buttons
	A	A10	Fail to press setting button	Cannot proceed	None	Medium	High	Better icons and placement of buttons
	R	R1	Do not obtain information about cruise control settings	Might not be able to proceed	Explore the menu by pressing down button	High	Medium	Provide more information of settings menu
	R	R2	Do not obtain information about cruise control settings	Might not be able to proceed	Explore the menu by pressing down button	High	Medium	Provide more information of settings menu
1.2	A	A3	Wait too long to press down button	Dialogue window disappears	Redo operation	Medium	Low	Show information for a longer time
	A	A6	Press wrong button	Might not be able to proceed	Press another button	Low	Medium	Better placement of buttons
	A	A10	Fail to press down button	Cannot proceed	None	Low	High	Better placement of buttons
1.3	A	A3	Wait too long to press left/right button	Dialogue window disappears	Redo operation	Medium	Low	Show information for a longer time
	S	S1	Select the wrong cruise control setting	The wrong setting is activated	Change setting	Low	Low	Add descriptive text to icon
1.4	A	A3	Wait too long to press OK button	Dialogue window disappears	Redo operation	Medium	Low	Show information for a longer time
	A	A9	Do not press OK button	Cannot proceed	None	Low	Medium	Ask user to press OK
2.1	A	A7	Cannot handle cruise control switch	Cannot proceed	None	Low	High	Redesign switch button
	A	A10	Fail to press cruise control switch	Cannot proceed	None	Low	High	Better placement of buttons
	R	R1	Do not get information about the cruise control settings	Might set wrong speed	Increase/decrease speed	Low	High	Higher visability of information
2.2	A	A10	Fail to press cruise control switch repeatedly	Cannot proceed	None	Medium	High	Redesign switch button
	R	R1	Do not get information about the cruise control settings	Might set wrong speed	Increase/decrease speed	Low	High	Higher visability of information


Overall goal
0 Set cruise control from 20 km/h to 80 km/h
Plan 0: 1>2
Sub-goals
1 Activate cruise control
2 Set speed
Plan 1: 1>2>3>4
Plan 2: 1>2
Operations
1.1 Press quick settings button in steering wheel
1.2 Press down button in steering wheel to reach cruise control menu
1.3 Press right/left button in steering wheel to mark cruise control icon
1.4 Press “OK” button in steering wheel
2.1 Press “+” button two steps
2.2 Press “+” button repeatedly until reaching 80 km/h

TEMPLATE

Task step	1. Task type	2. Error type	3. Error description	4. Consequence	5. Recovery possibility	6. Probability (L/M/H)	7. Criticality (L/M/H)	8. Improvement strategy
