Assessing Vessel Traffic Service Operator Situation Awareness

ISBN 978-90-8891-148-4

Published by Uitgeverij Boxpress, Oisterwijk

Printed by: Proefschriftmaken.nl || Printyourthesis.com

Assessing Vessel Traffic Service Operator Situation Awareness

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties
in het openbaar te verdedigen
op dinsdag 16 feburari 2010
om 15:00 uur
door

JOHAN WILLEM FREDERIK WIERSMA

doctorandus in de psychologie
geboren te Zeist

Dit proefschrift is goedgekeurd door de promotor:
Prof. dr. A.R. Hale

Samenstelling promotiecommissie:

Rector Magnificus, voorzitter
Prof. dr. A.R. Hale, Technische Universiteit Delft, promotor
Prof. dr. K.A. Brookhuis, Technische Universiteit Delft
Prof. dr. P.T.W. Hudson, Technische Universiteit Delft
Prof. dr. J.A. Mulder, Technische Universiteit Delft
Prof. dr. ir. P.A. Wieringa, Technische Universiteit Delft
Prof. dr. B. Kirwan, University of Nottingham
Ir. T. Heijer, Technische Universiteit Delft

Acknowledgements

Many people have contributed to this work, and hereby I take the opportunity to thank you all. For the larger part of a decade Andrew Hale was the sole reader of different versions of this thesis, who must have read more than a thousand pages of my writing. His meticulous comments and the stimulating discussions we had after each new chapter or version have made this thesis much better.

A large part of the research described in this thesis was carried out in cooperation with MSR Rotterdam. Throughout these projects Wim van ‘t Padje always has contributed in the most positive and constructive manner. We have worked together for many long hours and his experience in VTS, in training operators and in working with the simulator were indispensable. The enthusiasm of Henk Regelink got the research of the ground and kept it going. When he noted that we were doing something valuable, he created the opportunities for us to do more studies. Other people at MSR have contributed at different stages of the project: The late Jeroen Hooijer, who passed away during the early stages of the work; Nathasja Mastenbroek worked with us on the SATEST experiments; René Butter participated in the development and application of PMI-P. I hold great memories of the time spend there with you all.

The Rotterdam Port Authority provided VTS operators both for the development of the methods and as subjects for the experiments. In Piet de Korte, who my first contact point, I thank all of you. In Finland I was supported by the Helsinki Port Authorities and worked together with Goran Granholm, of VTT Manufacturing Technology, Maritime and Mechanical Engineering in Espoo, Finland.

Many others in the maritime domain have contributed to this work, from Cees Glansdorp, who introduced me to the field and included us in the COMFORTABLE project; to Arne Wolters (NNVO) and Jos van Doorn (MSCN), who showed me at the end of the thesis that there is still a world to win in training situation awareness in the VTS world and many persons in the time between those two moments.

Tom Heijer and I worked together on the development of the SATEST methodology. Tom developed the software and we spent many hours discussing the results and implications for other domains. Ideas come from somewhere. When Tom and I were starting our work in the COMFORTABLE project, Barry Kirwan was doing a sabbatical in our Safety Science Group in Delft. He suggested us to read the work of Mica Endsley on situation awareness as an interesting approach to the project, which started the work described in this thesis. With my other colleagues at the Safety Science Group – present and past – I have worked on many different projects, mostly not in the field of this thesis. I thank you for all the discussions and perspectives on safety, security, and research methodology that you provided over the years.

I have presented my thoughts on situation awareness in relation to safety and security to many generations of students and discussed with them the merits and pitfalls of the different approaches. These discussions have helped me as much as I hope they helped you understanding the role of awareness in different domains.

The past ten years have been an interesting journey, much more than just work. For all support and motivation on the road I want to thank Thea van de Kraats, Gabriela Levin, and Sjors Jacobs.

The most important people I want to thank are of course Ljiljana and Emma. Time working on this thesis was time not spend with you. The piles of books and and boxes and scattered papers have created a mess and obstructed more than bookshelves. All these years you have given me unconditional love and support. I dedicate this book to you.

Leiden, January 2010.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

1.1 Background

The modern world cannot exist without the transportation of large quantities of goods and people around the globe. For centuries shipping has been the major means of transportation to support world trade. Through the Port of Rotterdam alone, over 421 million tonnes of goods are transhipped annually. Around 34,000 sea going vessels and over 130,000 inland vessels arrive in Rotterdam every year, together accounting for over 420,000 ship movements on arrival, departure and berth changes (Rotterdam, 2009, p. 5). Over the years this huge amount of transportation has only increased. Not only have the quantity of goods that are being transported increased over the years, the complexity of traffic has increased as well, for instance with the introduction of high-speed vessels.

The increased quantity and complexity of traffic, combined with the necessity to optimise traffic streams and logistics in harbours and port areas require a form of traffic control. In maritime transportation, traffic control is carried out by Vessel Traffic Services (VTS). According to IMO resolution A.857(20) a VTS is defined as:

> *A service implemented by a competent authority, designed to improve the safety and efficiency of vessel traffic and to protect the environment. The service should have the capability to interact with the traffic and to respond to traffic situations developing in the VTS area.*

VTS has been around since 1948, when the first harbour surveillance system was introduced in the port of Liverpool (O'Neil, 2000). The value of VTS in navigation safety was first recognised by IMO in resolution A158 (ES.IV), *Recommendation on Port Advisory Systems*, adopted in 1968, followed by Resolution A578(14) in 1985, *Guidelines for Vessel Traffic Services*. As a result of the improvements in efficiency, safety and the reduction of potential environmental pollution experienced by authorities using VTS, together with the rapid developments in computer technology, the number of Vessel Traffic Services has increased considerably and in 1998 there were about 500 of these services operational worldwide (IALA, 2002a).

VTS operators are responsible for a safe and efficient handling of vessel traffic. They perform this function in harbours, rivers, and approach areas all around the world. They monitor traffic, provide information on request and coordinate movement of ships in (emerging) conflict situations.

For gathering information on traffic situations, vessels' intentions, weather conditions, and other relevant parameters, VTS operators often have sophisticated tools at their disposal, such as double radar displays of traffic combined with synthetic images of harbour lay-out and vessels, and information systems of ships, harbour etc. A large number of communication tools of all kinds enable VTS operators to communicate with ships, other VTS centres, and allied services.

Technology development and improvement of instruments used at VTS centres are an on-going process. This new technology changes the performance of the system and of the operator. Instruments used at VTS centres are subject to constant development and improvement, with the idea to improve the performance of the operator in his work. To ensure that recently developed tools enhance VTS operator performance and that safety levels are improved or maintained, new equipment and technological devices have to be tested prior to implementation.

1.2 Problem definition and research questions

The Safety Science Group of Delft University of Technology got involved in the VTS field in the COMFORTABLE project, carried out within the 4th Framework of the Waterborne Transport Programme of the European Commission's Directorate for Transport (DG VII) between January 1996 and December 1998. The COMFORTABLE project (COMFORT – Advanced Benefits for Logical VTS Equipment) had the objective to develop new tools for VTS use to help operators recognise and assess traffic situations, including the evaluation of risk. Special attention was to be paid to human factors issues when integrating these enhancements into the VTS operator's Traffic Situation Display. The goal was to provide operators with better tools without increasing workload (Regelink, Coles, & Jarvis, 1999).

Work package (WP2) in the COMFORTABLE project studied if and how the tools that were being developed in the project contributed to safety. To determine this, a better understanding of the work of the VTS operators was needed. VTS is the place where all information of the traffic in a region is collected and combined into one comprehensive traffic image. The VTS operator plays a central role in the decision-making process in the traffic control system. Therefore the performance of operators is essential in the performance of the total VTS system (Butter, et al., 1998). In the COMFORTABLE study performance of the VTS operator was considered the central issue. A better understanding of this performance was considered a necessary requirement for answering questions regarding new technology, selection and training of personnel and safety.

It is necessary to understand how this technology affects VTS operator performance to understand if the new technology for VTS really improves the safety and efficiency of vessel traffic control. One central aspect of this performance is how well an operator is able to maintain an overview. An operator has to maintain an overview of many aspects of his task environment: The VTS operator has to know where the ships are, what they are doing and where they are going. An operator must also retain an overview of the infrastructure and work in progress. Moreover the operator must know the weather conditions, the tide and many other aspects that determine the boundary constraints of visibility and traffic manoeuvrability.

Prior to the work described in this thesis, research in the field of VTS mainly consisted of descriptions of tasks and functions based upon observation or functional analysis (*COST 301*, 1987) or task analysis combined with incident analysis (CEC, 1996). In a VTS station maritime traffic is displayed on radar screens. The representation of this traffic on the displays was referred to as the "*traffic image*". In a paper that compared the safety of different modes of traffic control (Wiersma, Glansdorp, & Stoop, 1995) I assumed that this traffic image referred to the mental image that the operator had of the traffic. Only later, when we got involved in the VTS research, did I learn that this *traffic image* referred to how traffic was represented in the radar display, not the mental representation of the traffic. The task of the operator was to observe this *traffic image* and to detect and respond to *symptomatic events*, events that required attention or action. The operator had to maintain an overview of what happened on this display, but the mental image that the operator required to do so was not mentioned explicitly. Indeed there was no indication that these two might be different.

At the start of the COMFORTABLE project the common approach to the evaluation of new equipment was asking operators to participate in the design and evaluation of these tools. Actual performance of operators had never been used in research for such an evaluation. The methods available that could be used to assess VTS operator performance, were those meant for training VTS operators. These methods focused upon communication between VTS operators and ships and knowledge of procedures. They did not take any account of nor separately develop the overview of the traffic. Only in recent years have there been developments in a methodology for training operators that explicitly accounts for how and how well VTS operators maintain their overview of the traffic.

The research described here uses the concept of situation awareness to fulfil the need for a method that studies the mental picture or overview of VTS operators. Situation awareness is a term designating a person's internalized mental model of the current state of the environment (Endsley, 1990, p. 1). Situation awareness is a relatively new concept, which has gained a lot of momentum in the past decades. The term was made popular through the work of Mica Endsley. The concept situation awareness has been applied in many domains to describe the mental model of operators in different environments.

The concept of situation awareness also presents a new perspective for the study of VTS operator performance by investigating the mental picture of the operator. This new perspective may lead to better understanding of the performance of VTS operators. The main objective of this thesis has been to study how situation awareness of VTS operators can be assessed and if the use and the assessment of this situation awareness lead to a better understanding of the performance of VTS operators. The main research question of this thesis is:

> *What does the assessment of situation awareness contribute to the understanding of VTS operator performance?*

This question cannot be answered in one step. Three questions need to be answered before the main research question can be answered:

2.1 What is situation awareness?
2.2 How can situation awareness be assessed?
2.3 How can situation awareness of VTS operators be assessed?

1.3 Research approach and thesis outline

The research described in this thesis has been carried out over a period of thirteen years, from 1996 to 2009. During that period the field of VTS has changed. The importance of situation awareness for the work of VTS operators has become well recognised and the term has become commonly accepted. Situation awareness is referred to in official documents regarding training of operators (IALA, 2003) and the effect of technological developments on situation awareness of VTS operators is taken into account (IALA, 2002b). Developments have also taken place in the field of Human Factors research concerning the concept of situation awareness. At the beginning of this study in the mid-nineties, situation awareness was in the focus of the scientific debate and conferences and journal articles were dedicated to the study of situation awareness in many domains (Vidulich, et al., 1994). Towards the turn of the millennium, the debate switched from the concept of situation awareness towards application of situation awareness and aspects of awareness (Scerbo, et al., 1999). At present (Summer 2009) research in the field of situation awareness is still ongoing. However, the basic methodologies for testing situation awareness are well established.

The focus of the work in this thesis is on the performance of VTS operators to improve understanding of how VTS operators carry out their task. At the start of this study, there was not much understanding of this performance and there were no tools or methods for predicting performance – it was considered a black box. The work described here has been dedicated to opening that box.
Situation awareness is a concept that has been applied with success in many domains where controllers need to maintain an overview of their work. The introduction of the concept of situation awareness in the VTS domain made people working there realise that this concept reflected something important for the work of VTS operators. Therefore two methods to assess situation awareness were developed. The first method was based upon the work of Mica Endsley, an acknowledged leading expert in the situation awareness field. Her SAGAT methodology (Endsley, 1995a) was the basis for our method called SATEST. The SATEST method is useful to answer certain (types of) research questions, but it does not answer all relevant questions.

One particular type of issue that cannot be well addressed by this method is that of the time dependency of information. Development of the maritime traffic process is slow compared to the development of processes in many other domains. Operators keep the same items in their situation awareness for a long time, enriching this information with observations and communication as time passes. The development of situation awareness over time is an important aspect of situation awareness that was underestimated in the beginning of the study and in developing SATEST.

When this issue became clear, a need for other methods for assessing VTS operator performance and VTS operator situation awareness arose. Therefore another method for assessing VTS operator performance was developed. This method, called PMI-P, uses communication between VTS and traffic as an indirect measure of situation awareness and as a direct measure of VTS operator performance. PMI-P can be used to answer different questions than can be answered with SATEST.

Together the two methods offer a wide range of opportunities to study situation awareness of VTS operators. This study improves the understanding of operator performance and can assist in the design of better tools to support VTS operators in their daily work.

The thesis is organised as follows:
Chapter 2 presents a framework for the examination of the concept situation awareness and for methods for assessing it in the context of VTS operator performance. This evaluation is based upon a literature study on situation awareness and VTS research and upon observations on VTS posts in the Netherlands.

Chapter 3 describes the development and use of a method for assessing VTS operator performance based on situation awareness, called SATEST. The method is derived from the SAGAT method developed by Mica Endsley and can be considered a traditional method for measuring situation awareness. This method was developed, tested and used in the COMFORTABLE project. Chapter 3 also presents the conclusions drawn from the use of SATEST.

Chapter 4 presents a second method for assessing VTS operator performance based on situation awareness, called PMI-P. This method uses the assessment of communication and relevant dimensions of behaviour to assess performance and situation awareness. The method was used in a project for the Rotterdam Port Authority.

Chapter 5 discusses the concept situation awareness in the context of assessing VTS operator performance. It also provides an answer to the question of what situation awareness can contribute to our understanding of human information processing and presents the main conclusions.

CHAPTER 2

Situation awareness

2.1 Introducing situation awareness

In the previous chapter the term situation awareness was presented as a key element in the assessment of performance of VTS operators. Situation awareness has been put forward as an objective measure of the understanding of supervisory situations, such as VTS. It is based upon the notion that the human operator uses available data sources to create and maintain a mental representation of the situation he must supervise. This representation, together with (his personal interpretation of) goals and objectives and his expectations or preconceptions determine his actions.

The construct of situation awareness as it is used today was developed in military aviation some twenty years ago by Mica Endsley (Endsley, 1988). Since then, researchers with a background in many scientific disciplines have studied the construct and tested its applicability in various contexts, within different fields of application, ranging from flying helicopters (Smyth, Malkin, & DeBellis, 1990) to anaesthesiology (Gaba, Howard, & Small, 1996) to emergency management in fire fighting (Companion, 1994; Schenk, 1994). Several conferences devoted to situation awareness (Garland, et al., 1996; Gilson, Garland, & Koonce, 1994), numerous symposia at meetings of the Human Factors and Ergonomics Society and other conferences (Vincenzi, Mouloua, & Hancock, 2004), and a special issue of Human Factors (37:1) are evidence of the fact that situation awareness is a construct which has attracted a lot of attention.

Situation awareness has been a topic of scientific research for more than twenty years. Numerous attempts have been made to define it and give it a place among other cognitive constructs such as mental models, workload, attention, working memory, etc.

An annotated bibliography (Vidulich, et al., 1994) demonstrates the attempts of the research community (in that bibliography merely from aviation) to get a grip on the construct of situation awareness: out of 233 papers, 61 mention a definition of situation awareness. In the same report, (Dominguez, 1994) provides a table with fifteen different one-line definitions from this bibliography (p. 6-7), and provides another one herself.

(Dominguez, 1994, p.5) states: "*When a new construct emerges and gains momentum in the academic and applied communities with such force as has situation awareness, it is only natural that those of us who pursue its meaning and its measurement should seek to define it as well. Unfortunately, the lack of an agreed upon definition of situation awareness has itself been a defining characteristic of situation awareness from the start. As was the case with mental workload, there are many definitions, and although the concept is accepted as important without qualification, nobody is willing to accept anybody else's definition (Wickens, 1992)*". The first version of this chapter was written some ten years ago, in the second half of the 1990's. Out of 147 papers, articles and reports reviewed at that time for the thesis (see Annex 1), 76 provided a definition of situation awareness, or discussed the concept. Overviews of literature had been presented by several authors; see for instance (Dennehy, 1996) and (Jones, 1996). At that time there was still a fierce debate going on about situation awareness. The March 1995 issue of the *Human Factors* journal marked the top of that discussion. With eight papers, this issue brought together the different perspectives on situation awareness of that period. The papers showed different theoretical approaches to situation awareness, each with their own methodological approach for assessing situation awareness.

The articles in Human Factors issue (37:1) are the most influential articles on situation awareness ever written. In order to update this chapter a bibliographic analysis on publications about situation(al) awareness was carried out using Scopus[1] in July 2009. The analysis showed that six out of the ten most cited articles on situation awareness are indeed from this issue (see Table 2-1). The most influential article is that in which Mica Endsley elaborates on her theory of situation awareness. That article has been cited over 500 times since publication. The other two articles in the issue are still in the top thirty of most cited publications, with 29 and 57 citations respectively (Flach, 1995; Smith & Hancock, 1995).

Since then the discussion has somewhat faded. Several papers have appeared on situation awareness theory, but they seldom have been more than a synthesis of ideas expressed earlier, or an explanation of situation awareness for a different audience, such as the paper by Stanton and colleagues in Safety Science in 2001 (Stanton, Chambers, & Piggott, 2001).

Around the turn of the millennium the first books on situation awareness appeared. The most awaited book was that of Mica Endlsey, published in 2000 (Endsley & Garland, 2000), in which she outlines her approach to situation awareness. This book is a

[1] Scopus is a large abstract and citation database of research literature and quality web sources. It offers 16,500 peer-reviewed journals from more than 5,000 publishers; including coverage of over 1,200 Open Access journals; 350 book series and coverage of 3,0 million conference papers publications (www.scopus.com, visited 20-07-2009).

comprehensive overview of the main approaches to situation awareness, both the theory and application. One of the latest books on situation awareness is "*A cognitive approach to situation awareness: theory and application*", edited by (Banbury & Tremblay, 2004). This book does not strictly follow the Endsley line of thought, although most chapters do quote Endsley's main publications. The book does address several issues that are not clearly worked out in Endsley's approach to situation awareness and therefore is a welcome addition to the existing literature on situation awareness. this book contains a chapter on the concept of situation (Flach, Mulder, & Paassen, 2004) and one on the role of awareness (Croft, et al., 2004). Methodological aspects of situation awareness measurement are also discussed, for instance the issue of obtrusiveness of the SAGAT methodology (McGowan & Banbury, 2004), that will be addressed later in this thesis in Chapter 3.

Publication	# Citations
Endsley, M.R. Toward a theory of situation awareness in dynamic systems (1995) Human Factors, 37 (1), Pages 32-64.	527
Collins, R.T., Lipton, A.J., Fujiyoshi, H., Kanade, T. Algorithms for cooperative multisensor surveillance (2001) Proceedings of the IEEE, 89 (10), pp. 1456-1477.	208
Endsley, M.R. Measurement of situation awareness in dynamic systems (1995) Human Factors, 37 (1), Pages 65-84.	164
Sarter, N.B., Woods, D.D. How in the world did we ever get into that mode? Mode error and awareness in supervisory control (1995) Human Factors, 37 (1), Pages 5-19.	121
Fletcher, G., Flin, R., McGeorge, M., Glavin, R., Maran, N., Patey, R. Anaesthetists' non-technical skills (ANTS): Evaluation of a behavioural marker system (2003) British Journal of Anaesthesia, 90 (5), pp. 580-588.	106
Gutwin, C., Greenberg, S. A descriptive framework of workspace awareness for real-time groupware (2002) Computer Supported Cooperative Work: CSCW: An International Journal, 11 (3-4), pp. 411-446.	91
Endsley, M.R., Kiris, E.O. The out-of-the-loop performance problem and level of control in automation (1995) Human Factors, 37 (2), Pages 381-394.	87
Salas, E., Prince, C., Baker, D.P., Shrestha, L. Situation awareness in team performance: Implications for measurement and training (1995) Human Factors, 37 (1), Pages 123-136.	86
Adams, M.J., Tenney, Y.J., Pew, R.W. Situation awareness and the cognitive management of complex systems (1995) Human Factors, 37 (1), Pages 85-104.	86
Gaba, D.M., Howard, S.K., Small, S.D. Situation awareness in anaesthesiology (1995) Human Factors, 37 (1), Pages 20-31.	76

Table 2-1 Most cited publications on situation awareness in period (1996-2005) in Scopus database

Based on this extensive literature, what then is situation awareness, and how can it be assessed? Most theories on situation awareness have in common that they regard situation awareness as an active part of knowledge that is combined with current information about the task or environment. This forms a representation of the state of the world that can be used to guide behaviour. The exact form of this representation is under discussion. An important issue that researchers do not agree upon is the question whether situation awareness is a process or the product of this process.

This chapter presents a model of situation awareness that can be used in the VTS domain and places it in a framework of decision making and information processing. The model follows from a critical review of the work of Mica Endsley, incorporating the ideas and experience that have been gained in the past ten years.

The interested reader who wants to know more on situation awareness and methods for measuring it, is referred to some of the landmarks of the situation awareness literature: The 1995 March Issue of the *Human Factors* Journal (37:1), which is completely dedicated to situation awareness, the book of Endsley and Garland (2000) *Situation awareness analysis and measurement,* the book of Banbury and Tremblay (2004) *A cognitive approach to situation awareness: theory an application* or can search the literature referred to in Annex 1

2.2 A model of situation awareness by Mica Endsley

The term situation(al) awareness has been used in the US military aviation since World War I (Press, 1986, as cited in (Endsley, 1988). Some of the definitions used by the military are very generic: "*An assessment of the situation based on the best possible information*" (Waddell, 1979, as cited in (Jones, 1996). Others are very operational: "*It's simply knowing what's going on so you can figure out what to do!*" (Adam, 1994). A definition that gives a better grip is the one proposed by the United States Air Force's operational community. They defined situation awareness as: "*A pilot's (or aircrew's) continuous perception of self and aircraft in relation to the dynamic environment of flight, threats, and mission, and the ability to forecast, then execute tasks based on that perception.*" (Judge & Bushman, 1992, p. 40).

Probably the best known framework, and certainly the most quoted one line definition of situation awareness (Bolstad, 1991; Bolstad & Hess, 1996; Carmody, 1993; Charness, 1996; East, 1996; Gaba, Howard, & Small, 1995; Hardy & Parasurama, 1997; Kaber, 1996; Vidulich & McCoy, 1995) and many others has been proposed by Mica Endsley (Endsley, 1988; Endsley, 1990, 1995b; Jones & Endsley, 1996; Robertson & Endsley, 1996). Endsley has developed her framework from research in (military) aviation. The definition of situation awareness was developed with a pilot in mind, but can be easily transferred to other domains. According to Endsley, "situation awareness can be conceived of as the pilot's internal model of the world around him at any point in time" (Endsley, 1990, p. 5).

Endsley proposes a model of situation awareness in dynamic decision making as depicted in Figure 2-1.

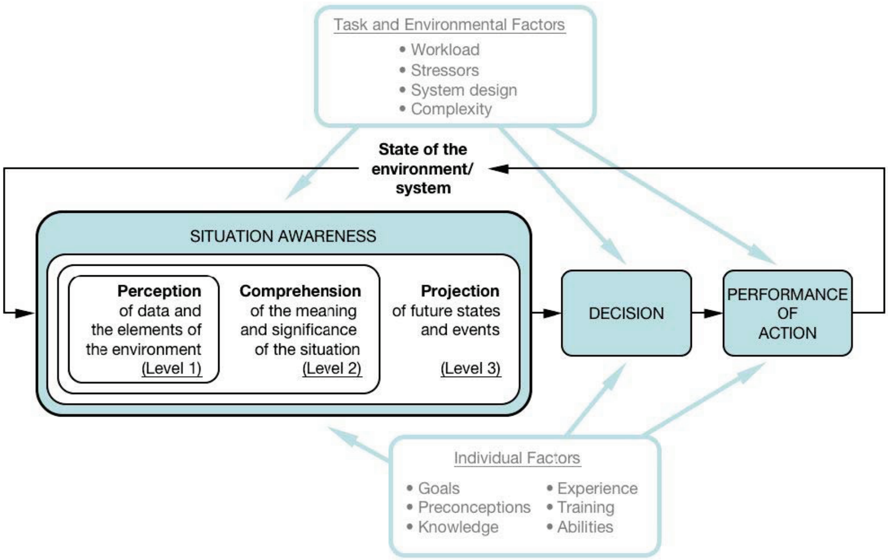


Figure 2-1 Model of situation awareness in dynamic decision making[2]

According to this model, a person's perception of the relevant elements in the environment, determined by use of system displays or directly from senses, forms the basis for his or her situation awareness. Selection and performance of actions are separate stages that will proceed directly from situation awareness.

The process is influenced by several factors, both on the individual and on the task/system level. On the individual level, factors such as abilities, experience and training influence the quality of the information gathering mechanisms. Goals, objectives, and preconceptions (expectations), on the other hand, provide schemata, that filter the data collection and steer the data interpretation that forms situation awareness.

Situation awareness will also be a function of the system that provides the information. The system design should provide not only the right information, but also this information must be provided in such a way that the operator can use it to maintain or adapt his situation awareness.

To establish a consistent terminology, (Endsley, 1994) makes a distinction between the process of collecting information to build a model of the situation and the model that results from this data collection. *Situation awareness* refers to the *state of knowledge* at any

[2] from: http://upload.wikimedia.org/wikipedia/en/6/61/Endsley-SA-model.jpg, downloaded 01-09-2007

moment in time. For the process to achieve, acquire or maintain this state of knowledge Endsley proposes the term *situation assessment*. This distinction differs from other efforts to define situation awareness (Dominguez, 1994; Sarter & Woods, 1995), but provides a better model, because it separates two distinct entities that can studied separately. In this model situation awareness forms the critical input to decision making, but is separate from decision making, which is the basis for all subsequent actions. *"Even the best trained and most experienced pilots can make the wrong decisions if they have incomplete or inaccurate situation awareness. Conversely, a pilot may accurately understand what is occurring in the environment, yet not know the correct action to take or be unable to carry out that action"* (Endsley, 1990, p. 9). Endsley defines situation awareness as:

> *The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future (Endsley, 1990, p. 12).*

According to this definition, this mental picture can be seen to consist of three levels:

Level I SA: Perception of elements in the current situation: the state of the environment, which an operator has to supervise, can be defined as a large collection of different parameters with specific, momentary, values. The completeness of the mental picture of the operator depends in part on the possibility to obtain information on these parameters: personal abilities and supporting equipment determine the subset of obtainable parameters and hence the limits of completeness of the picture.

Level II SA: Comprehension of current situation: numerous state parameters in the environment will be related by causal mechanisms. In any given situation some of these mechanisms are latent and others are active in affecting the way state parameters change over time. Comprehension of the situation can be described in terms of the available repertoire of mental models of these causal mechanisms, the choice of model(s) from this repertoire that the operator is currently applying and the quality (relevancy, accuracy, completeness) of the chosen models with respect to reality.

Level III SA: Projection of future status: apart from an assessment of the current state of the environment, a prediction of how this state will change in the near future is also important for the decisions and actions an operator may take. These predictions are based upon the consequences of external events or actions taken. The quality of these predictions in part determines the quality of the decisions.

Endsley stresses the temporal nature of situation awareness (Endsley, 1995b, p. 38). Situation awareness is something that is built up over time and changes constantly with new information becoming available, taking into account the dynamic nature of the situation.

In many contexts, situation awareness is also very spatial in nature. Often spatial relations also have a temporal dimension. This is the case in for instance ATC and VTS, with spatial relations between multiple aircraft (or ships) that change over time.

Situation awareness does not encompass all of a person's knowledge. It refers only to that portion pertaining to the state of a dynamic environment. Similarly, situation awareness is presented as a construct separate from others that may influence it, such as attention, working memory, workload and stress. Figure 2-1 presents mechanisms affecting and underlying the situation awareness process. In Endsley's terms this is called the *situation assessment*.

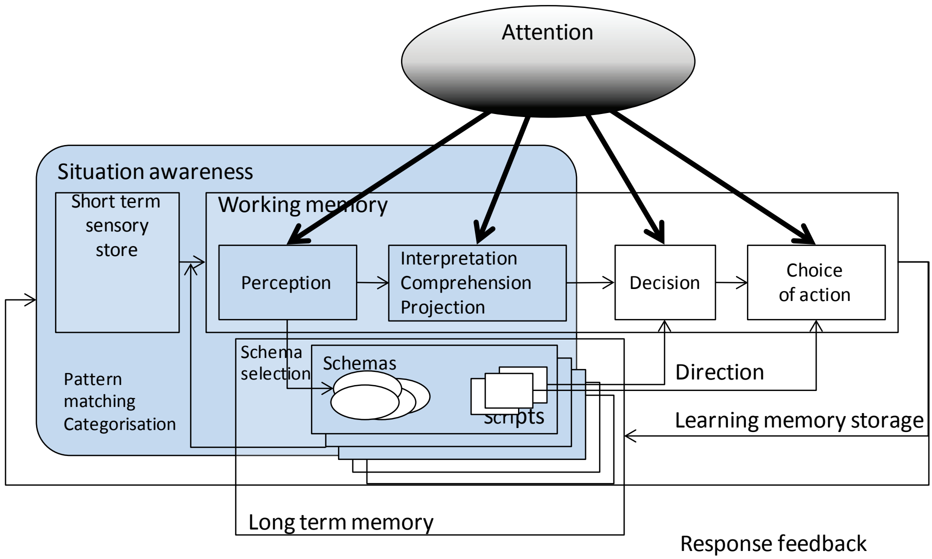


Figure 2-2 Mechanisms of situation awareness[3]

Figure 2-2 contains the following elements:

- Attention directs whether perceived data are interpreted at all (incorporated into situation awareness) and which data are used in the process of decision making and action guidance.
- A short term sensory store contains all the data that enter the person through sensory perception, that have not yet being processed (perceived)
- Only a part of the working memory is defined as being part of the situation awareness process. As stated, decision making and action guidance are not considered to be part of situation awareness or the situation awareness process.

[3] (Endsley, 1990)

- In Endsley's model, **schemata**[4] are parts of the long-term memory, used to filter data from the short term sensory store into perception, and to give direction to the interpretation of information. **Scripts**[5] are also part of the long term memory, but they direct more the processes of decision making and action guidance.

In Figure 2-2 situation awareness can be interpreted as a specific situation model. This model contains more or less fixed elements, such as schemata and scripts and dynamic information in the short-term sensory store and working memory.

Schemata and scripts are the basic elements of the situation model. They are filled with (dynamic) information from the short term sensory store and working memory and in turn provide filters and decision rules to direct the working memory.

If necessary, information from the working memory can lead to a change in the chosen schema or script. If this happens, then situation awareness must be built again: the new schema has to be filled with adequate data.

Attention influences the working memory and learning/memory storage influences the schemata and scripts (on a longer time scale). The action taken in a situation provides a response feedback to the short-term sensory store.

In summary, Endsley presents a model that describes situation awareness as a state of knowledge of the environment, which uses input both from memory and the environment to understand "*what is going on around you*" (Endsley, 2000a, p. 5). Situation awareness is distinct from situation assessment, the process of achieving situation awareness, but the two have a strong interaction.

2.3 Critical review of Endsley's work

2.3.1 Definition of situation and awareness

The term situation awareness consists of two words. A sharp and clear definition of both words seems necessary to establish consensus in using the term situation awareness.

Situation is the first word that needs to be pinned down. Most of Endsley's work is dedicated to the awareness part of the definition, leaving the situation wide open. Her definition of situation awareness as "*The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future*" quoted above leaves far too much room for the inclusion of all kinds of superfluous elements of the situation to become part of the situation awareness model. Endsley has

[4] Schemata are abstract knowledge structures or diagrammatic representations.

[5] Scripts are a memory structure or a form of knowledge representation that organise knowledge about stereotypical situations, such as dining at a restaurant or going to the movies. For further reference see (Schank & Abelson, 1977).

recognised this herself as she states "*Inherent in this definition is a notion of what is important*" (Endsley, 2000a, p. 5). Since the concept of situation awareness can be applied to the awareness of any (type of) situation, task or activity it seems necessary to pin down exactly what is meant by it. Therefore I prefer to narrow down the definition of a situation and propose one presented by (Pew, 1994, p. 18):

> *A situation is a set of environmental conditions and system states with which the participant is interacting that can be characterized uniquely by a set of information, knowledge and response options.*

Three aspects of the situation are particularly important for situation awareness:
- The complexity: The amount of data and interactions present in the situation;
- The coherence. Elements are either independent and need to be treated separately or they are related and can be chunked into larger bits of information. In the latter situation, expertise and training can help;
- The relevance. Not all information available in a situation is relevant, only that information that is necessary for carrying out the task.

To better understand the complexity, situations can be regard as "*a nested set of constraints that have the potential to shape performance*" (Flach, Mulder, & Paassen, 2004, p. 44). Rasmussen's abstraction hierarchy developed in 1985 provides a framework for representation of knowledge that can help structure the complexity and coherence. The abstraction hierarchy, generally referred to as AH, is one of the best known representation frameworks that has been proposed for describing complex work environments (Bisantz & Vicente, 1994), although most papers refer to it, rather than adopting it to solve a particular problem (Bisantz & Vicente, 1994, p. 84). The five levels of the AH are:
- Functional purpose level describes objectives, functions and limitations of the system
- Abstract function level describes the causal structure principles underlying the goals
- Generalised function level describes the standard functions and processes
- Physical function and level describes the physical components or equipment
- Physical form level describes physical appearance, location and condition of the components.

Complexity in situations represented in situation awareness will usually be based upon the understanding of the operator of the first three levels, fed with information about the state and functioning of the other two levels.

The second term of the concept situation awareness is *Awareness*. According to the Oxford English Dictionary (2nd ed., 1989):

> *Awareness refers to "The quality or state of being aware; consciousness". Aware is defined as: "(1) Watchful, vigilant, cautious, on one's guard (2) Informed, cognizant, conscious, sensible" (ibid.).*

Awareness is different from knowledge (Laughery & Wogalter, 1997, p. 1176) consider the difference analogous to a distinction made in cognitive psychology between short-term memory (sometimes thought of as what is currently in consciousness) and long-term memory (one's permanent knowledge of the world).

(Croft, et al., 2004) devote a chapter to the role of awareness in situation awareness. They state that the focus of Endsley on the explicit conscious knowledge will only provide an incomplete picture of situation awareness, given the failure to take into account the implicit knowledge (Croft, et al., 2004, p. 82). They advocate an approach that integrates implicit as well as explicit measures of situation awareness.

2.3.2 Product or process

Endsley's model of situation awareness makes a distinction between situation awareness and the assessment of the situation. The two strongly influence each other, but should be considered separately. Not all authors agree with Endsley in this respect. In the literature there is a discussion as to whether situation awareness should only be the *product* or *state*, as Endsley proposes, or that the *process* or *activity* itself should be part of situation awareness. The latter perspective is taken by (Sarter & Woods, 1995, p. 15). They view situation awareness as "*a variety of processing activities*".

Other authors, such as (Adams, Tenney, & Pew, 1995) argue that although it is possible to make a division of situation awareness into product and process, it is not always useful. The two aspects of situation awareness are interconnected. These authors give examples of aviation accidents to illustrate their point.

Endsley replies that, although there is great benefit in examining the interdependence between the process and the state of knowledge that is a result of this process, it is important to keep the terminology straight (Endsley, 1995b). This thesis supports Endsley's position. From a theoretical perspective there is no real need to mix the two, while for the purpose of communication the distinction between process and product is useful, reserving the term situation awareness for the latter only.

2.3.3 Situation awareness and information processing

Endsley places her model of situation awareness in a context of information processing, referring to (Wickens, 1992) theory of human information processing (Endsley, 1988; 1990; 1995b; 2000, p. 12). Her figure on mechanisms of situation awareness (Figure 2-2, this chapter) is one adaptation of Wickens' model. This figure seems to me unclear and imprecise. It does not clarify where situation awareness is to be located; it seems to be part of memory, but Endsley also includes perception, interpretation, decision making, and action guidance in memory. In an attempt to clarify this subject, (Endsley, 2000a, p. 12-18) introduces new figures that are not much clearer. It seems that Endsley has problems locating situation awareness as a functional block in models. Thus it seems that Endsley too has problems separating the product from the process.

The adaptation from (Wickens, 2000) model presented in Figure 2-3 is much clearer on this point. It locates situation awareness in working memory. Wickens stresses (2000, p. 261) that situation awareness is not the same as memory (the structure); it is what is maintained in it (the information contained). In his figure Wickens (2000, p. 295) includes part of the perception block into the situation awareness block. This seems to indicate that he wants to include part of the process into the situation awareness concept. The section above argues why I do not agree with this.

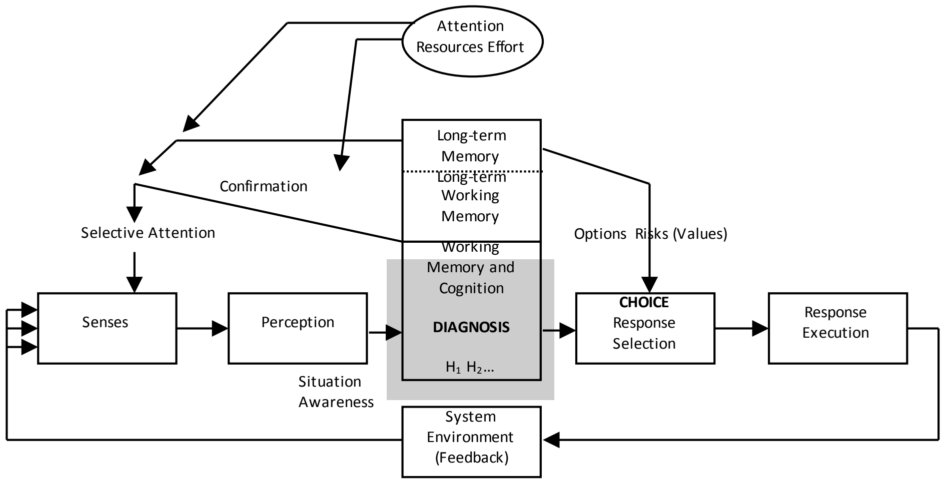


Figure 2-3 Key processes and components in information processing model of decision making[6]

Therefore Figure 2-3 adapts Wickens figure but separates situation awareness and perception. This leaves situation awareness as part of working memory (the grey rectangle), as it is depicted in Figure 2-3. In fact situation awareness can be regarded as the content of working memory. It takes its reference input from long-term memory, in the form of models, scripts etcetera and feeds these with information from sensory perception.

2.3.4 Levels of situation awareness and decision making

Endsley regards situation awareness as a necessary step in the decision making process. Others, including Wickens, can imagine decision making without proper situation awareness. An everyday example of this provided by Wickens is the experience of driving a car absent-mindedly: "C*onsider, for example, how well we can drive on an uncrowded highway even*

[6] Schank & Abelson, 1977. Adapted Wickens, 2000, p. 295

as our mind drifts away from awareness of the road conditions. Our driving responses are good, but our situation awareness is low." (Wickens, 2000, p. 261).

This example by Wickens illustrates the difference between situation awareness and consciousness: Situation awareness may be enough to perform the task, even though the driver is not consciously aware of that.

Although Endsley's model of situation awareness covers the relationship between the levels of situation awareness (*Level I SA*: Perception of elements in the current situation, *Level II SA*: Comprehension of current situation, *Level III SA*: Projection of future status, according to Endsley's definition) and decision making, I do not think the model is very clear in that respect. (Pritchett & Hansman, 2000) present a model that clarifies this relationship better. Their model is depicted in Figure 2-4.
This model describes the decision making process of an operator in terms of information processing and locates the levels of situation awareness presented by Endsley in this model. Level I, perception of elements correspond to the perception of information; Level II reflects the operator's understanding of the situation and Level III refers to the understanding of the need for action.

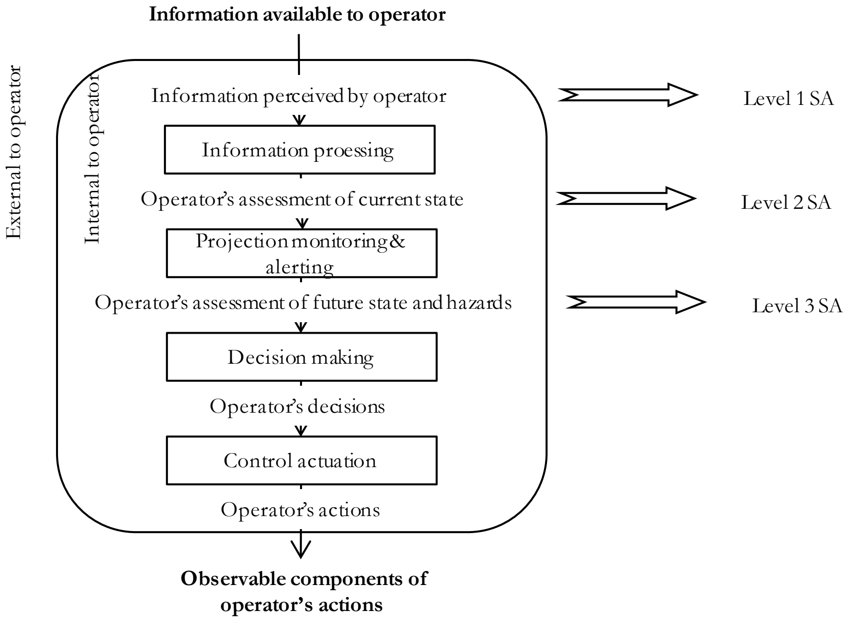


Figure 2-4 Levels of situation awareness and the decision process[7]

[7] Pritchett and Hansman, 2000, p. 193

In terms of the decision making model of Pritchett and Hansman it can be argued that the example of the car driving presented above describes a situation where situation awareness on the second and third level is low, while situation awareness on the first level is present and is actually guiding behaviour.

2.3.5 In conclusion

Surveying the literature on situation awareness, it can be concluded that there is no complete consensus on what situation awareness is. Most researchers follow the ideas presented by Mica Endsley in the late 1980's, but several issues are still under discussion. The main controversy seems to focus around the question if situation awareness has to be considered a state of knowledge and the product of a process (sometimes called situation assessment) or if this process itself should be part of the concept situation awareness.

Different authors have taken a different perspective on this issue. However, the discussion on the issue has died down. Most recent literature on situation awareness has stopped discussing the situation awareness concept and its theory. Now most research is trying to develop (experimental) experience with the concept and is leaving the theoretical discussion for what it is.

For the work in this thesis Endsley's model provides a useful framework. The distinction between the process of acquiring situation awareness and the product of this process (situation awareness itself) seems logical and has proven useful throughout this thesis.

Situation awareness is a psychological construct, not something that can yet be localised in the human mind. As such the definition of situation awareness is somewhat arbitrary. A good definition seems to be one that satisfies two objectives:

1. It is in agreement with other currently accepted theories and models, or at least not in disagreement with them;
2. It is useful in describing and explaining phenomena that are being observed; it increases the understanding of the world around us.

This thesis applies the concept of situation awareness in its most limited form. The definition for situation awareness used in this thesis is:

> *Situation awareness is the mental model that represents the understanding that an operator has of the current state of his task environment. This mental model is the result of the process of acquiring information called situation assessment. Situation awareness is stored in short-term memory, where it receives information both from long-term memory (the objectives that need to be achieved or functions that need to be carried out, and the models and scripts needed to interpret and process data presented in the environment) and sensory input that provides this information.*

The next section describes how this concept of situation awareness can be used to describe the work of VTS operators.

2.4 Situation awareness and Vessel Traffic Services

The previous section of this chapter has shown that there are several definitions of situation awareness. The question remains what the meaning of this is in the context of VTS and the performance of VTS operators. Is situation awareness a good concept to use in this domain? To investigate this issue, this section will first introduce the task of the VTS operator in the process of traffic management. Next the role of situation awareness in the work of the VTS operator will be described.

2.4.1 Vessel Traffic Service

According to IMO resolution A.857(20) Guidelines for vessel traffic services, adopted November 1997:

> *Vessel traffic service (VTS) is a service implemented by a competent authority*[8]*, designed to improve the safety and efficiency of vessel traffic and to protect the environment. The service should have the capability to interact with the traffic and to respond to traffic situations developing in the traffic area.*

Two main types of VTS are distinguished: A Port or Harbour VTS and a Coastal VTS. "A Port VTS is mainly concerned with vessel traffic to and from a port or harbour or harbours, while a Coastal VTS is mainly concerned with vessel traffic passing through the area. A VTS could also be a combination of both types. The type and level of service or services rendered could differ between types of VTS; in a Port or Harbour VTS a navigational assistance service and/or a traffic organization service is usually provided for, while in a coastal VTS usually only an information service is rendered." (Resolution 857(20), Annex 1, section 2.1.2).

The precise objectives of any VTS may depend upon particular circumstances in the VTS area and the volume and character of maritime traffic. A VTS is particularly appropriate in an area with high traffic density, complex navigation patterns or an infrastructure where the progress of traffic may be restricted. To perform their task VTS personnel should have access to a VTS traffic image. Therefore data need to be collected and evaluated to develop the traffic image. Figure 2-5 presents a general model for a VTS service, indicating the different modes of collecting information.

---

[8] A Competent Authority is the authority made responsible, in whole or in part, by the Government for the safety, including environmental safety, and the protection of the environment in the area

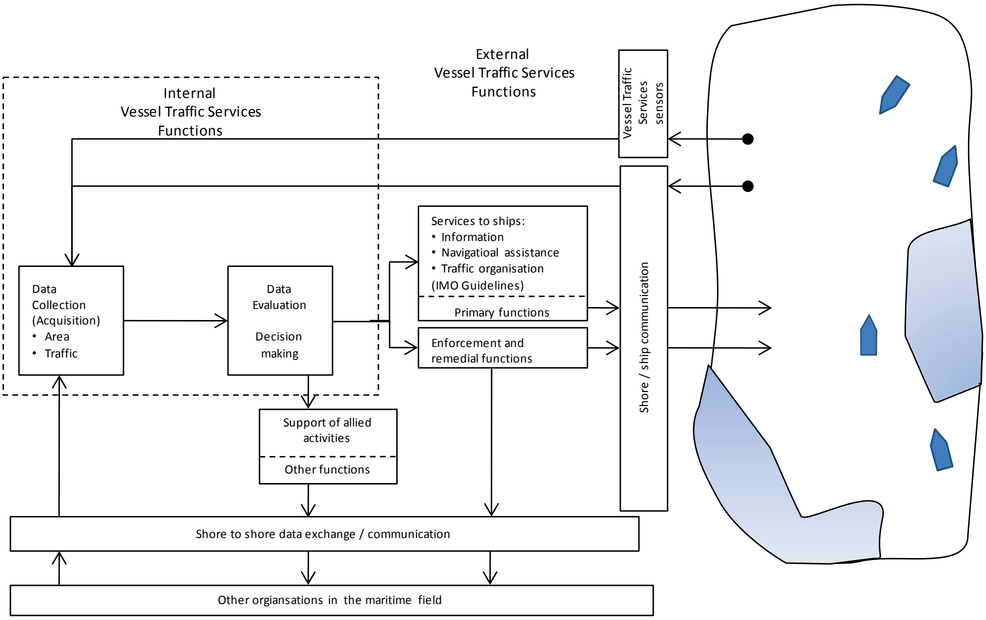


Figure 2-5 Generalised model for Vessel Traffic Services[9]

To obtain tactical information on the movement and identification of ships, VTS centres use radar, combined with closed circuit television and visual observation. VHF Communication is used to identify target ships. With the progressive introduction of the Automatic Identification System (AIS)[10] on board ships, this information too is used in VTS centres. The International Association of Lighthouse Authorities (IALA) VTS manual (IALA, 2002a, p. 25-29) presents an overview of technical options for a VTS centre.

The Competent Authority must assume responsibility for the proper and professional operation of the VTS and set up a legal framework for its functioning. Three major legal issues that need to be addressed are: First, the authority of central government, port or other maritime entity to impose control over vessels by means of VTS regulations; second, the relationship between VTS personnel and ship masters or pilots operating in VTS waters (determination of responsibility); and third, the determination of liability in the event of an accident involving a vessel participating in a VTS (IALA, 2002a, p. 29). Annex 5 of the IALA VTS manual (IALA, 2002a, p.125-134) presents an overview of some International and National legislative measures relevant to the provision of VTS. On the relation between VTS and Vessels, the IALA VTS Manual states (IALA, 2002a, p. 31):

---

[9] COST 301, 1987, p. 57.

[10] AIS is an autonomous and continuous broadcast system, operating in the VHF maritime mobile band. It is capable of exchanging information such as vessel identification, position, course, speed, etc. between ships, between and shore and through information broadcasts. (http://www.iala-aism.org/web/pages/AIS/cadreais.htm, downloaded 10-10-2008)

> *"Safe movement of marine traffic in a VTS area requires the VTS to have an understanding of the responsibility of the masters of vessels and vice-versa.*
> *"A VTS, with its specialised knowledge of the waterway has responsibility for managing the traffic in the area. While the master of a vessel, with his knowledge of its behaviour and his professional skills, has responsibility for the safety of the vessel. Taking into consideration these different, but related, responsibilities, any instructions from a VTS to a vessel should be "result orientated" only, leaving the details of execution to the master, officer of the watch or pilot on board the vessel.*
> *"Generally masters of vessels rely with confidence on the expertise and professionalism of VTS personnel and carry out instructions which are given. However, it should be recognised that there may be occasions when an instruction by a VTS is disregarded because the master considers that its execution might jeopardise the vessel."*

In earlier research projects on VTS no references are made to the concept of situation awareness. The term situation(al) awareness is for example not mentioned in the COST-301 reports (*COST 301*, 1987) or in the main report of TAIE (CEC, 1996)). Even though the concept of *situation awareness* was not widely known when the COST-301 project was finished in 1987, the underlying ideas are present in the COST-301 reports. The main idea that a VTS should maintain a traffic image forms the basis of managing the traffic. However, in the COST-301 reports this traffic image refers to an image created by technology. The role of the VTS operator, interpreting the traffic and maintaining a mental traffic image in his head is not made explicit.

2.4.2 VTS operator knowledge, skills and situation awareness

In the past decade the importance of situation awareness for the work of the VTS operator has become well recognised by the maritime world, to a significant extent through the work described in this thesis.

The IALA Vessel Traffic Services Manual is the guide that authorities use to consider all the aspects that can arise when planning, structuring and operating a VTS (IALA, 1993, p. 7). At the start of the project the first (1993) and second edition (1998) of the Manual were used. These manuals do not contain any references to situation awareness. In 2002 the third edition of the IALA VTS Manual was published (IALA, 2002a). This edition of the Manual states that one of the considerations to determine a period of duty for a VTS operator/supervisor should be *"the ability to develop and maintain situational awareness"* (p. 67). So not only has the concept been adopted, there also is an understanding that it takes time to build situation awareness, and effort to maintain it.

The IALA Recommendation on Standards for training and Certification of VTS Personnel V-103, published in 1998, presents an extensive list of the knowledge and skills a VTS operator needs to have. V-103 also includes methods for demonstrating this competence and criteria for evaluation of the competence (IALA, 1998, p. 24-31). The document states that tests need to be designed to determine the ability of candidates for VTS to demonstrate spatial and situational awareness (IALA, 1998, p. 15).

Situation awareness is also referred to in the IALA "Guidelines on AIS as a VTS tool". In these guidelines AIS is proposed to be an important tool to enhance situation awareness (IALA, 2002b).

The general VTS model from COST-301 presented in Figure 2-5 can be interpreted as a task model of a VTS operator. Situation awareness of the VTS operator is then covering the internal functions of the VTS services, the mental model that is necessary to carry out the external functions. The main difference with the original model lies in the interpretation of the figure: putting a (human) VTS operator in the loop. This interpretation is presented in figure 2-6.

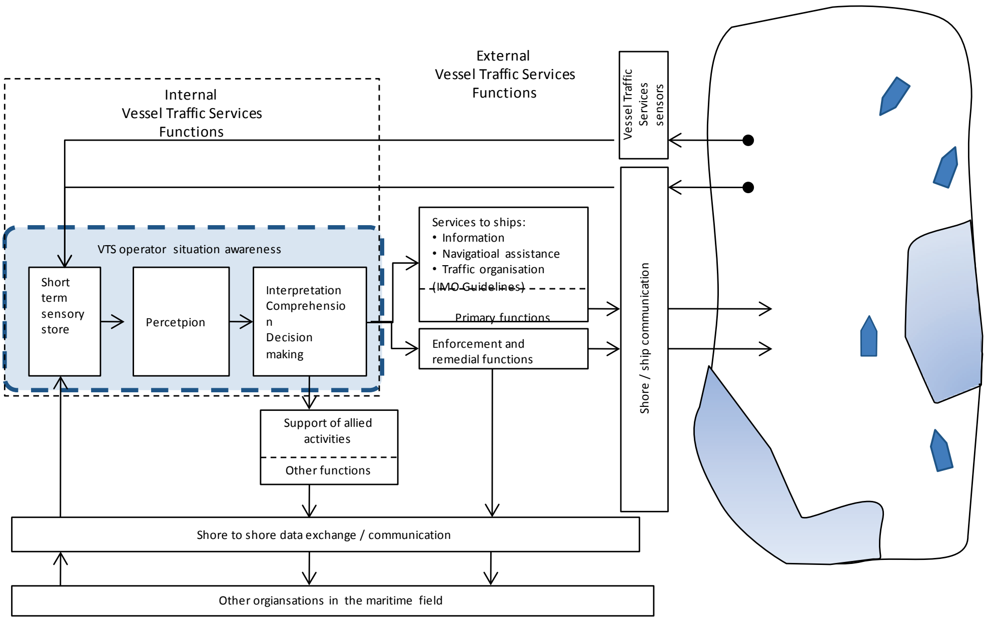


Figure 2-6 General model for Vessel Traffic Services including a VTS operator situation awareness[11]

The knowledge and skills of the VTS operator provide the scripts and schemata that, together with the dynamic information provided by the sensors make up the situation awareness. Together they form *Level I situation awareness: perception of elements in the current situation*. This refers to the extent to which the VTS operator is aware of all the information presented to him that is relevant in relation to the VTS task.
Endsley stresses the *temporal nature* of situation awareness (see section 2-2). Situation awareness is something that is built up over time and changes with new information becoming available, taking into account the dynamic nature of the situation. It is important

[11] Adapted from COST-301, 1987, fig 4.2 p.75

to notice that the dynamic changes in the situation (and therefore in situation awareness) develop in a different time frame for different aspects: Destination and cargo are fixed for one trip, course and speed are set for a certain track that can take from minutes up to hours, conflicts arise and disappear in minutes. Distances of ships from other ships or infrastructure need be monitored constantly. A method for assessing situation awareness must take account of the different time frames for different aspects of the traffic.

Another element mentioned in Endsley's model that is relevant for VTS is the *spatial nature* of situation awareness. The spatial relations between ships and between ships and shore change over time. This is another aspect that a method for assessing VTS operator situation awareness should consider.

At *Level II situation awareness: Comprehension of current situation* the understanding that the VTS operator has of the traffic situation is defined. Understanding the situation in relation to safety means that the operator knows which ships may have a conflict, either with other ships (risk of collision) or with the infrastructure (risk of stranding). These conflicts may be in the current situation but more likely are further ahead in time (which makes them part of Level III situation awareness). Relevant to this understanding of the situation is not only the position and movements of ships, but also other information, such as weather conditions, for instance fog and tidal conditions.

The understanding of the situation that a VTS operator has goes beyond the simple extrapolation of current course and speed of ships. This information provides merely a very rough measure of what is going to happen in the next few minutes. More important to the understanding of the situation are the tracks that ships are sailing and deviations from planned tracks. Even more important than the ship information the operator acquires from the VTS displays is the information he gathers through communication with the ships. Ships involved in a potential conflict communicate with each other and the VTS operator about how they plan to solve a situation. If such an agreement is made between the ships, and the ships keep to that agreement, the conflict ceases to exist for the VTS operator.

Track information and understanding of the communication that has taken place, either between ships or between VTS operator and ships, are very important elements in the understanding of the situation the VTS operator has.

Level III situation awareness: Projection of future status builds on the understanding of the situation which has been described in Level II situation awareness. *Projection of future status* contains the prediction of future position, course and heading of ships, and the consequences of those, such as conflicts that may develop. Only then can an operator decide if action needs to be taken. Level III situation awareness involves planning for all situations that have the potential of developing into a conflict, taking into account the circumstances.

Conflicts arise when it is not clear to all ships involved (and to the VTS operator) how a particular situation is going to be solved. As long as such uncertainty exists, the situation is to be regarded as a (potential) conflict. As long as there has been no or insufficient communication with all parties involved, the situation deserves extra attention. This is the core of the operational work of a VTS operator: predicting where conflicts may develop, and resolving such situations.

2.4.3 In conclusion

In the traditional perspective, the *VTS traffic image* was something that was created by technical means. The role of the VTS operator was merely to react to situations that were detected by the technology. Modern perspectives on VTS recognize the importance of the VTS operator in the traffic monitoring process. The traffic image is not just a representation on a display. In order to respond to situations, the VTS operator has to maintain a *mental* traffic image to recognise where problems in the traffic are or where they may develop. This mental representation can be well expressed with the concept of situation awareness.

There are two specific aspects of the VTS work that need to be taken into account in developing a method to VTS operator situation awareness. First, the assessment method needs to take the temporal nature of maritime traffic into account. Ships move relatively slow, but conflicts may develop quickly. Second, the assessment method needs to take the spatial nature of maritime traffic control into account. Ships are divided over a two dimensional area and operators have to make a mental picture of where ships are going and what infrastructure and other ships they will meet underway.

2.5 Criteria for a method for VTS situation awareness assessment

2.5.1 Criteria for selecting a method from the VTS field

It can be concluded from the previous section that situation awareness plays an important role in the work of a VTS operator. An operator needs to maintain a mental picture of the situation he is monitoring, updating it constantly with new information becoming available. Therefore the development of an assessment method for situation awareness for VTS may provide useful understanding of VTS operator performance.

There had been until the work described in this thesis no previous attempts to assess VTS operator situation awareness. The projects described in this thesis offer the first opportunity to investigate this issue. Therefore it was imperative to develop a method, or methods, that provide us with as much information about VTS operator situation awareness as possible.

The COMFORTABLE project in which the first method for assessing VTS operator situation awareness was developed focused on the development of new equipment for VTS stations. These new tools were not yet applied in real-life, but were tested in an experimental situation. Therefore application of the method in real-life was not necessary. For this project it was more useful if the method was suited for application in an experimental setting.

From the constraints of the COMFORTABLE project two criteria for an assessment methodology for VTS operator situation awareness emerge:

1. *Amount of (relevant) information* that can be obtained at one time. It is considered preferable to obtain as much as detailed, information as possible at one time;
2. Some methods can be applied in *real life* only, some in *experimental setting* only and some can be applied in both contexts. In general methods with more fields of application are preferred, but for this project methods were preferred that are suited for application in an experimental setting.

2.5.2 Criteria from the different perspectives on situation awareness

As section 2.1 has shown, there are different ways to look at the construct situation awareness. It will be no surprise that there are as many methods for assessing situation awareness, as there are different theoretical points of view. Several authors have provided classifications of these methods, along different dimensions. Section 2.6 will provide an overview of the discussion and types of methods resulting from it. This discussion will form the basis for the development of a method to assess the situation awareness of VTS operators described in Chapter 3.

As with the definition of situation awareness, researchers studying pilots in aviation have dominated the discussion on methods and often the methods developed are applicable only there. However, several methods have been designed and applied in other domains, including Air Traffic Control, car driving, or anaesthesiology, but there had been none up to the time of this thesis devised specifically for the vessel traffic system.

Many authors have presented a classification of situation awareness methods (Dennehy, 1996; Endsley, 1995a; Pew, 1996; Sarter & Woods, 1995; Vidulich & Hughes, 1991). In my opinion all of these have one shortcoming in common: they mix up measures and assessment techniques. *Measures* refer to *what* is being assessed by a method. While all assessment techniques assess situation awareness in one form or another, there is a large difference with respect to exactly what is being assessed. Examples of measures are system performance measures, operator performance measures, and subjective measures. *Assessment techniques* refer to the question *how* the method assesses situation awareness. Examples of assessment techniques are verbal protocols, queries, and observer ratings.

(Pew, 1996) for instance categorises the five following measurement categories:

- Direct System performance measures
- Direct experimental measures
- Verbal protocols
- Subjective measures
- Model-referenced performance measurement

The verbal protocol does not fit in this list; it is an assessment method, while the other categories refer to measures. Other categorizations have similar problems. Therefore this chapter presents a new classification. This classification is based on types of measures of situation awareness that can be distinguished, and assessment techniques that can be used to assess these measures. Five different measures of situation awareness are distinguished. They are:

- Recollections from operator memory
- Ratings on a situation awareness scale
- System performance
- Operator performance
- Physiological measures

The first two of these are direct measures of situation awareness that refer to internal processes, operator memory recollections and ratings on situation awareness scale. The first one refers to the mental picture of the operator directly. The second measure refers to an operator's evaluation of the quality of this picture. The other three measures are indirect measures of situation awareness that are assessed externally: situation awareness has to be inferred from the performance of the operator on a specific task, or from interpretation of situation awareness from performance or from physiological measures taken from the operator. For these methods the quality of the data depends on the correlation between the measurement and situation awareness. Figure 2-7 presents the different measures.

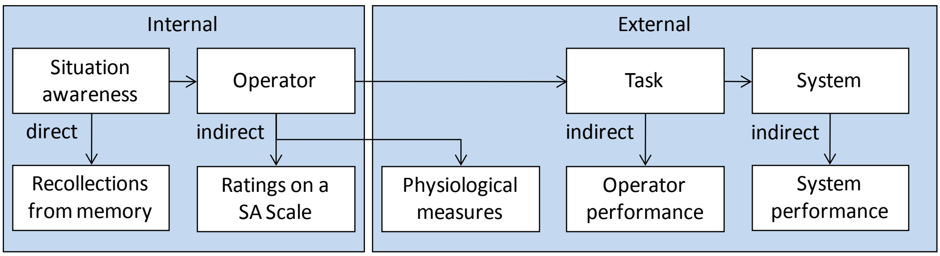


Figure 2-7 Schematic overview of situation awareness measures

As can be seen from Figure 2-7, only the recollections from memory are a direct measure of situation awareness. Even though ratings on a situation awareness scale are collected directly from an operator, they are not a direct measure of situation awareness. The other measures differ in how direct an indication of situation awareness they are. Each link in

Figure 2- takes them away one step further away from situation awareness. The quality of the measures does not only depend on the amount of steps, but also on the quality of the links (the arrows in Figure 2-7).

There is a wide variety in methods for assessing these five measures. All methods have their specific characteristics that make them applicable for certain applications and for certain types of research questions and contexts. These methods can be evaluated on a number of criteria. In addition to the two criteria mentioned in section 2.5.1 (amount of relevant information that can be obtained, and the use of methods in either real life or experimental setting), methods can be divided into methods that collect direct situation awareness and methods that collect indirect measures of situation awareness.

3. *Direct versus indirect assessment of situation awareness.* In general direct assessment of situation awareness is preferred over indirect assessment of situation awareness, since it provides information that does not need to be interpreted;

Other relevant criteria for evaluating methods for assessing situation awareness are the following.

4. In general, *simple* methods are preferred over *complex* methods. However, simple data collection often goes hand in hand with collecting only a small amount of information. As stated in the criteria in section 2.5.1, the amount of information collected should be larger and preferably as detailed as possible. Complex methods are more likely to provide that information.
5. Methods that measure situation awareness *on-line* are preferred over methods that only *provide post hoc* information on situation awareness, due to the problems with memory reconstruction of situation awareness. However, in general, on-line methods are more obtrusive than post-hoc methods;
6. *Unobtrusive* methods are preferred over *obtrusive* methods, since the latter may change operator behaviour. If an obtrusive method is chosen, attention will have to be paid to the extent to which the method disturbs the work of the operator.

2.6 Review of assessment techniques

There are several assessment methods available for the different types of measures. This section presents an evaluation of those methods, under the following headings:

1. Factual queries and memory probes
2. Rating scales
3. Incident reports
4. System performance measures
5. Experimental measures
6. Physiological measures

2.6.1 Factual queries and memory probes

Detailed information on all three Levels of situation awareness can be obtained by the use of queries to recollect information from memory. Probing techniques are especially designed to measure situation awareness in a direct way. The central idea of these techniques is that a subject's memory is probed for certain details of a situation at randomly chosen or preselected moments. Usually the methods are applied *on-line*, although *post-hoc* probing is possible. There are basically three major types of probing techniques (see Figure 2-8):

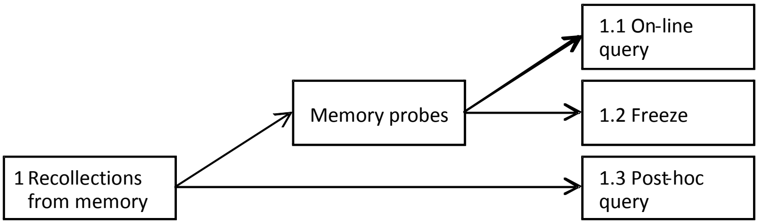


Figure 2-8 Classification of techniques for recollections of memories

2.6.1.1 On-line probing techniques

On-line probing techniques ask for information relevant to the (situation awareness of) the situation during the simulation run. The probing question is asked by the experimenter and may be presented as if a colleague or someone else from within the task environment asked it. An example of this is an assessor assuming the role of a ship and requesting specific information from a VTS operator.

STRONG POINTS:

- If the research focuses on a small part of the situation, questions can be asked pertaining to that detail, which enables the researchers to follow that particular aspect over time.
- If the query can be presented in such a way that answering the question is part of normal operation, then this method is an unobtrusive experiment.
- The methods can be used in online experimental settings
- The method can also be applied in real-life situations, providing that the probing does not disturb the main task of the pilot/operator.
- The methods provide a direct assessment of situation awareness

WEAK POINTS:

- Only a limited number of questions can be asked at a time, making it difficult to establish a good representation of the subject's total situation awareness.
- This method runs the risk that the question is recognised for what it in fact is: THE thing that the experiment is all about. If this happens, the question becomes very intrusive and can completely destroy the experiment.

2.6.1.2 Freeze technique

The second type of probing uses a *freeze* technique within the scenario. This is a typical experimental setting, in which subjects are aware that they are being assessed, and are informed about the procedure. At unannounced, but previously specified times during these scenarios the system displays are blanked and the simulation is stopped. The operators (e.g.) are requested to answer a number of questions about their perception of the situation at that time, designed to reveal their situation awareness. After completion of the query, the simulation continues. The best known technique of this type SAGAT, the Situation Awareness Global Assessment Technique, has been developed in aviation research by Mica Endsley (Endsley, 1990, 1991, 1995a). SAGAT is a global tool for measurement of situation awareness in military aviation. In SAGAT, pilots run a range of simulator flight scenarios. The questions asked when the screens are blanked pertain to all three levels of situation awareness, Level I (perception of elements); Level II (comprehension of current situation) and Level III (projection of future status). This includes a consideration of the system status as well as relevant aspects of the external environment. The accuracy of the responses is then compared with logged situation data. Thereby a measurement of situation awareness on all three levels is derived.

Endsley's method has been widely used and adapted, both in the aviation domain (Bolstad, 1991; Carmody, 1993), the air combat domain (Chubb, 1996; Endsley, 1996) and in other domains, such as anaesthesiology (Small, 1996); process control (Hogg, et al., 1995); team situation awareness for aviation crews (Prince, Salas, & Stout, 1996) aviation maintenance (Endsley & Robertson, 1996); chess (Durso, et al., 1996) and a laboratory experiment using an experimental task of perceiving moving blocks on a display (Kaber, 1996).

The methods that are adaptations of SAGAT usually consist of the use of the freezing technique in a laboratory experimental setting or simulation. SAGAT-like queries have to be developed in a domain-specific or even task-specific way. Several of the authors mentioned have tried to incorporate questions pertaining to all three of Endsley's levels of SA (perception, comprehension and prediction).

STRONG POINTS:

- During the break a number of questions can be asked about the situation. Depending on the type of task, breaks up to five minutes have been applied without causing problems for pilots in getting "back into the scenario". This time is usually considered enough to extract all relevant information.
- This (type of) method has been widely accepted as the standard methodology for the assessment of situation awareness. There is considerable experience with this method and it has been successfully applied in many domains.
- The method provides direct assessment of situation awareness, with potentially a lot of information, depending on the information retrieved at the breaks

WEAK POINTS:

- The method can only be used in (simulator) experiments or training sessions that can be stopped. If stopping the simulation is impossible or too disruptive, the method cannot be applied.
- The paradigm with breaks cannot be applied easily in real life situations, since operators need to be brought out of their task to answer questions.

2.6.1.3 Post-hoc probing techniques

A post hoc probing method consists of asking questions pertaining to situation awareness after the completion of the task. These questions can be asked in the form of questionnaires or interviews and contain references to specific events in the scenario. This type of method is not used much because of its weak points.

STRONG POINT:

- Post-hoc probing is unobtrusive, debriefing is quite naturalistic and an unlimited amount of questions can be asked.

WEAK POINTS:

- Post hoc probes assess the memories of the trial run, not the situation awareness.
- They are subject to reinterpretation of the situation, depending on the outcome of a scenario.

2.6.2 Rating scales

Methods using subjective measures of situation awareness can be divided into two very distinct groups:

- Methods that apply a one-dimensional scale of situation awareness, call upon an intuitive understanding of situation awareness. These methods are often used in a preliminary phase of a project.
- The other type of method using subjective measures, is the opposite. These methods apply a very detailed multi-dimensional concept of situation awareness. Section 2.6.2.1.2 presents an example of such a method. Scores on the different scales are combined into a situation awareness score. This type of method seems most suitable for researchers whose main interest is a better understanding of situation awareness and the underlying principles and mechanisms.

The simple one-dimensional scale can be applied both on-line and post-hoc. The multi-dimensional rating scales are applied post-hoc only. Both types of methods can be used with self-rating scales and with observer ratings. Figure 2-9 gives a schematic overview of rating scales.

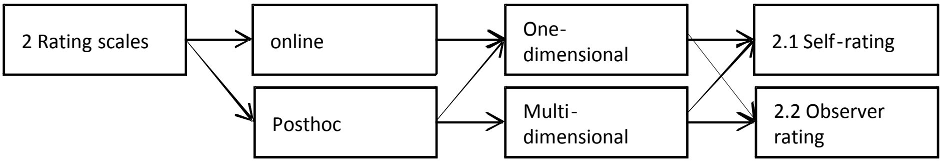


Figure 2-9 Classification of techniques for rating scales

2.6.2.1 Self-rating scales

2.6.2.1.1 On-line, one dimensional, self -rating scales

This is the simplest technique in which operators are asked to rate their own situation awareness. Such a method of assessing subjective situation awareness is proposed by (McGuinness, 1995). In his Simple Rating Scale SRC, pilots are asked mid-run to rate their understanding of the situation on a 10-point Likert-scale, thus addressing the pilot's own intuitive understanding of situation awareness.

STRONG POINTS:

- The method is very simple and can provide basic understanding of a task.
- Especially one-dimensional rating scales are simple and can be used on-line.

WEAK POINT:

- Self rating of situation awareness asks subjects for their awareness of their own awareness. This leads to asking people for their *unknown unknowns*. In the words of Wickens: "People are not aware of the things they are not aware of. "Therefore asking people to rate their level of awareness on a scale from 1 to 10 will not provide a terribly valid measure" (Wickens, 2000, p. 261).

2.6.2.1.2 Post-hoc, multi-dimensional, self-rating scales

Other subjective rating scales have been developed starting with a definition of situation awareness as a multi-dimensional concept. The best-known example of this approach is Selcon and Taylor's Situation Awareness Rating Technique SART (Selcon, Taylor, & Koritsas, 1991; Taylor, 1990; Taylor, 1996; Taylor & Selcon, 1990). SART has been designed as an evaluation tool for aircrew system design (Selcon and Taylor, 1990). The method is based upon a multi-dimensional description of situation awareness of (Taylor, 1990). It uses the ten dimensions of situation awareness described in Figure 2-10.

SART applies a seven-point Likert scale for each of ten dimensions. Scores on these ten constructs are grouped together to attain a score on a 3-dimensional scale of situation awareness. These dimensions are:

1. demand on attentional resources,
2. supply of attentional resources, and
3. understanding of the situation.

A rating of SART is carried out post-hoc in experiments. Subjects rate their own situation awareness, after completion of a specific task.

		LOW						HIGH
		1	2	3	4	5	6	7
DEMAND	Instability of situation							
	Variability of situation							
	Complexity of situation							
SUPPLY	Arousal							
	Spare mental capacity							
	Concentration							
	Division of attention							
UNDERS.	Information quantity							
	Information quality							
	Familiarity							

Figure 2-10 10-Dimensional SART scale[12]

STRONG POINTS:

- Multi-dimensional methods can provide a comprehensive set of dimensions to rate situation awareness, covering many relevant aspects.
- The rating scales that are used in the methods indicate the processes considered important for situation awareness.

WEAK POINTS:

- Self-rating on multi-dimensional scales suffers from the same disadvantages that it does on simple ones: Subjects tend to rate their situation awareness higher if they are successful in accomplishing their task.
- Post-hoc rating leads to reconstruction of situation awareness, related to the outcome of the experiment. Therefore the effects of self-rating are even amplified.
- Self-rating on multiple dimensions requires a lot of training for subjects. This is a weak point in the context of VTS research, where operators are not easily available.

[12] (Selcon & Taylor, 1990, p. 5-2)

2.6.2.2 Multi-dimensional, observer-rating scales

Observer rating is a method which requires an independent, knowledgeable observer to rate the quality of a subject's situation awareness. One example of a method that applies this type of scoring is the Cranfield-Situation awareness scale, developed by (Dennehy, 1997; Dennehy, 1996). The Cranfield scale was developed to monitor pilots' progress in developing skills necessary for situation awareness. It is based upon "*pilot actions and knowledge that the aviation community considered important in maintaining situation awareness*" (Dennehy, 1997, p. 1). Overall situation awareness scores are based on the addition of ratings on five subscales:

1. Pilot knowledge
2. Understanding and anticipation of future events
3. Management of stress, effort and commitment
4. Ability to attend, perceive, assimilate and assess information
5. Overall awareness

The Cranfield scale applies a 9-point Likert scale, ranging from unacceptable to excellent. The five dimensions are measured through scores on 22 descriptions. Several of these descriptions are specific for the aviation domain, especially the knowledge questions (e.g. Q.1-3: know the aircraft's systems, such as fuel and hydraulics?). Other questions can more readily be transferred to other domains (Q.2-4: accurately interpret incoming information?). The method can be applied as an *observer-rating* scale or as a *self-rating* scale.

STRONG POINTS:

- The methods provide a comprehensive set of dimensions to rate situation awareness, covering many relevant aspects of situation awareness.
- The rating scales that are used in the methods indicate the processes considered important for situation awareness.

WEAK POINT:

- An observer may have more information on the situation, but he also has less information of what is going on inside the subject performing the task. His observations are based on observable behaviour and are therefore comparable to other performance measurement techniques, and not really a direct situation awareness assessment.

2.6.3 Incident reports

Post hoc assessment of incidents and accidents can be a useful way to determine the importance of situation awareness in accident causation. A wide range of techniques are applied in this type of research, such as statistical analysis of experimental data and databases, video analysis, and in-depth investigation of specific accidents (Klein, 1996; Strauch, 1996).

The fundamental problem with this type of research lies in the first step of the process. Accident data collection related to physical objects and actions performed by operators is certainly possible. But trying to understand the situation awareness of people involved in an accident is almost impossible. In the words of Rodgers *cum suis*: "Occupations requiring interaction with complex, dynamic systems make the post hoc assessment of situation awareness problematic due to the lack of available information that directly relates to the cognitive processes of relevant personnel." (Rodgers, Mogford, & Strauch, 2000, p. 73). Still, in-depth investigation of accidents can reveal all sorts of underlying causes of accidents. It is possible that this type of research provides an understanding of situation awareness and situation awareness related issues in accidents.

STRONG POINT:

- Accident analysis can provide insight into the mechanisms of situation awareness *if* situation awareness is taken into account from the first data collection onwards.

WEAK POINTS:

- This type of analysis completely builds upon that data collected. If situation awareness has to be reconstructed from statistical analysis based on data where there was no attention to situation awareness during data collection, not much can be expected from it.
- Post hoc analysis of situation awareness suffers from memory reconstruction, especially in events with an adverse outcome.

2.6.4 System performance measures

A widely used group of methods for the assessment of situation awareness takes system performance as a measure of situation awareness. Many authors use performance measures in some way to measure situation awareness. (Pew, 1996) makes a useful basic distinction in performance between methods that use *direct system performance measures* (this section) and *direct experimental measures* (next section). The first method uses performance of the total system or performance of the operator on (part of) his normal task as a measure of situation awareness.

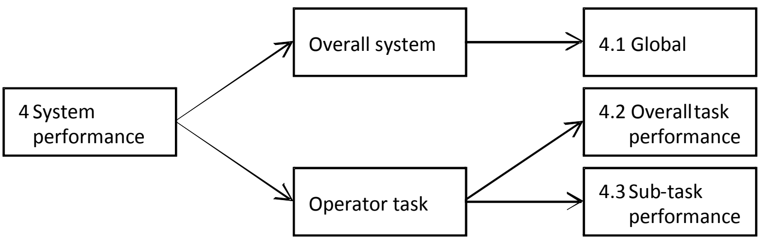


Figure 2-11 Classification of techniques using performance measures

Since situation awareness research is applied in many different fields of application, very many different performance measures are possible. Performance measures can be classified in three types (see Figure 2-11). The first group covers methods that assess the overall system performance, the other two types assess operator performance within the operation of the system as a measure.

2.6.4.1 Overall system performance

In general, direct system performance measures are objective and can be obtained non-intrusively. In general naturalistic test designs provide a good opportunity for assessing "normal" behaviour. However, as a measure of situation awareness they are indirect. Direct overall global system measures are appropriate only in situations where there is general agreement that the performance of the system in question is driven by situation awareness.

STRONG POINTS:

- System performance can be measured online non-intrusively.
- Performance of aspects or parts of the system that relate to situation awareness may be selected as measures.

WEAK POINT:

- System performance usually is not driven by situation awareness alone: other factors play a role in system performance. Therefore the performance has to be interpreted; situation awareness (or lack of situation awareness) has to be inferred.

2.6.4.2 Overall task performance

The performance of an operator on a task can be used as an indicator for the operator's situation awareness, if situation awareness determines how an operator performs his task. The same things that have been said for the overall system performance hold for the performance of the operator. Situation awareness may be a factor in operator performance, but other factors are important as well (such as training, motivation, arousal, and etcetera).

STRONG POINTS:

- Task performance can be measured non-intrusively.
- Situation awareness may be inferred from some situations

WEAK POINTS:

- Task performance usually is not driven by situation awareness alone: other factors play a role in the performance. Therefore overall performance has to be interpreted; situation awareness (or lack of situation awareness) has to be inferred.
- Performance on tasks in a natural setting cannot be controlled, as is desirable for experimental studies

2.6.4.3 Sub-task performance

If a (direct or indirect) link can be established between operator tasks and situation awareness that is needed to perform these sub-tasks, explicit assessment of the performance on these sub-tasks can be used as an indicator for situation awareness.

STRONG POINTS:

- Task performance can be measured non-intrusively.
- Performance of aspects or parts of the operator task that relate to situation awareness may be selected as measures.

WEAK POINT:

- Performance on sub-tasks in a natural setting cannot be controlled, as is desirable for experimental studies

2.6.5 Experimental measures

In the second type of performance measures, *direct experimental measures*, the researcher has changed something in the task environment (added a task). Performance on this additional task is taken as a measure of situation awareness. This extra task can be external, for instance an action to be taken by the operator which is not part of normal routine, or embedded in the task (looking up data in a database). Such a task may be added to mask that information is being collected to assess situation awareness. Experimental measures are presented in Figure 2-12.

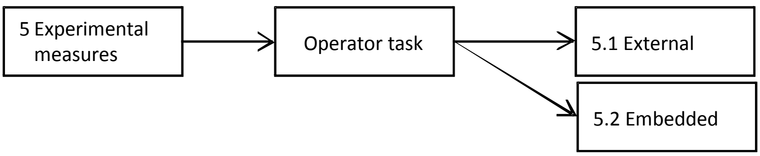


Figure 2-12 Classification of techniques using experimental measures

2.6.5.1 External tasks

As stated in the previous section, performance on tasks can be a good method for assessing situation awareness in situations where the relation between the task performance and situation awareness is clear. This type of measurement can be used in experimental settings if the experimental conditions can be created that evoke situation awareness based behaviour.

STRONG POINT:

- Performance methods assess final performance of systems and operator's actions. In many research situations, it is performance that actually counts, and situation awareness is a concept that is brought in to explain or predict certain behaviour.

However, it may very well be that if adequate performance measures are available, the concept of situation awareness is completely unnecessary.

WEAK POINTS:

- The strong point is at the same time the weak point of performance methods: they do not provide a direct measurement of operator's situation awareness.
- The relation between the task performance and situation awareness is not always clear.

2.6.5.2 Embedded tasks

If performance on the main task is not a good indicator for situation awareness, tasks that provide this information can be added to the task environment. An example of such a task may be answering to a co-pilot who is requesting information or filling in an on-line data display. The disadvantage of this method is that it can distract the operator from his original task, or make him aware that he is in a test situation. Therefore careful design of this task is necessary for minimising intrusiveness of the added task.

STRONG POINT:

- Performance methods assess final performance of systems and operator's actions. In many research situations, it is performance that actually counts, and situation awareness is a concept that is brought in to explain or predict certain behaviour. However, it may very well be that if adequate performance measures are available, the concept of situation awareness is completely unnecessary.

WEAK POINT:

- The strong point is at the same time the weak point of performance methods: they do not provide a direct measurement of operator's situation awareness.

2.6.6 Physiological measures

Physiological measures provide an assessment of the relationship between operator performance and correlated changes in operator physiology. Research in the field of mental workload measurement has shown that cognitive activity is associated with changes in various physiological systems. Well-known measures in this field include eye blink, heart rate, and respiration. All of these measures are indirect measures of workload; they measure physiological changes and these changes are used to make inferences about workload.

Several authors refer to the physiological aspect of situation awareness. The same measures that are used in the field of workload are used in the field of situation awareness, showing the overlap between the two concepts (Wilson, 1996, p. 142). Some authors even simply use workload measures to assess situation awareness, under the assumption that situation awareness varies inversely with workload (Metalis, 1993, p. 116). There is no evidence that this assumption is true, and also from a theoretical perspective there is no clear one-to-one relation between the two concepts (Endsley, 1995a). In a recent study (Vidulich, 2000) has demonstrated the independence of the two, showing how improvement of interface designs increases situation awareness, without generally reducing mental workload. Therefore, a physiological measurement of situation awareness cannot be that simple. (Westrenen, 1999) in his experimental research in understanding maritime pilots using physiological measures, for instance, rejects situation awareness because of the unclear link between situation awareness and workload.

(Byrne, 1996) argues that since maintaining situation awareness is an active process, psycho physiological measures may be useful. Especially Electroencephalography (EEG) and Heart Rate Variability (HRV) are proposed as useful measures in determining whether an operator is engaged in cognitive activity. Event Related Potentials (ERP) are used to determine whether an operator receives critical cues from the environment. The P300 amplitude[13] is widely used as a measure of expectancy. However, as a measure of situation awareness as defined in this thesis, it fails, since it does not reflect information content but only the process of retrieving information. Therefore it only refers to part of situation awareness.

(Stern, Wang, & Schroeder, 1996) use several physiological measures to assess operator predictions about the future (Level III situation awareness according to Endsley's definition). Specifically they apply assessment of heart rate and several oculometric variables, including the occurrence of saccadic eye movements, eye blinks and changes in the pupil diameter. The authors state that "expectancy leads to a decrease in heart rate, inhibition of blinking, and if one cannot inhibit a blink at a point in time close to an imperative event the blink is of shorter duration than normal" (p. 158).

STRONG POINT:

- The foremost advantage of physiological measures mentioned is the fact that they can be measured without interference with the primary task.

WEAK POINTS:

- Physiological measures applied to assess situation awareness have not been developed specifically to measure any aspect of situation awareness; existing methods have been used and their applicability in the situation awareness domain has been investigated. All in all, it seems to me that the authors mentioned have chosen to use physiological measurements of situation awareness because they are familiar with these measures. The link of those physiological parameters with situation awareness is still weak and sometimes even far-fetched.

[13] The P300 (P3) wave is an event related potential (ERP) which can be recorded via electroencephalography (EEG) as a positive deflection in voltage at a latency of roughly 300 ms in the EEG. The signal is typically measured most strongly by the electrodes covering the parietal lobe. The presence, magnitude, topography and time of this signal are often used as metrics of cognitive function in decision making processes. While the neural substrates of this ERP still remain hazy, the reproducibility of this signal makes it a common choice for psychological tests in both the clinic and laboratory. (http://en.wikipedia.org/wiki/P300_(Neuroscience), downloaded 10-10-2008).

- Physiological measures say nothing about information content. When they refer to the process of achieving or maintaining situation awareness (what is called situation assessment), they at best assess processes that support (achieving) situation awareness.
- Physiological measures therefore do not provide direct insight into situation awareness, and the best way to use them seems to be as a supporting measure, in combination with other measures of situation awareness.

2.6.7 Conclusion

In section 2.5 six issues have been defined which determine whether a particular type of method is applicable in a specific context:

1. Amount of (relevant) information;
2. Real life or experimental setting;
3. Direct versus indirect assessment of situation awareness;
4. Complexity of methods;
5. On-line or post hoc;
6. Unobtrusiveness.

		Amount of info	Real life	Experiment	Direct SA	Simple	On-line	Unobtrusive
2.6.1.1	On-line probing techniques	±	-	+	+	+	+	-
2.6.1.2	Freeze technique	+	-	+	+	-	+	±
2.6.1.3	Post-hoc probing techniques	+	+	+	-	-	-	+
2.6.2.1	One-dimensional self-rating scales	-	+	+	+	+	+	-
2.6.1.2	Multi-dimensional, post-hoc, self-rating scales	+	+	+	+	-	-	+
2.6.2.2	Multi-dimensional, post-hoc, observer-rating scales	+	+	+	+	-	-	+
2.6.3	Incident reports	±	-	-	-	-	-	+
2.6.4.1	Overall system performance	±	+	+	-	-	+	+
2.6.4.2	Overall task performance	±	+	+	-	-	+	+
2.6.4.3	Sub-task performance	+	+	+	-	-	+	+
2.6.5.1	External tasks	+	-	+	±	-	+	-
2.6.5.2	Embedded tasks	±	-	+	±	-	±	±
2.6.6	Physiological measures	±	+	+	-	-	+	±

Table 2-2 Situation awareness assessment techniques scored on criteria

Table 2-2 shows how the methods discussed in the previous sections score on the issues mentioned above. The methods are rated against the ends of the dimension mentioned in the table.

Table 2-2 provides understanding of the strong and weak points of the different methodological approaches. It does not yet provide a clear-cut answer, which of the methods mentioned is best suited for application in the field of assessing situation awareness of VTS operators. The constraints of the VTS work in general and the objectives of the research project under study will be essential in providing an answer to the question which approach is optimal. The requirements set by the different research projects will determine the approach. Table 2-2 will be used for selecting an appropriate measure for the different research projects which follow in the next chapters.

2.7 Selecting a method for VTS situation awareness assessment

It can be concluded from sections 2.4 and 2.5 that situation awareness plays an important role in the work of a VTS operator. An operator needs to maintain a mental picture of the situation he is monitoring, updating it constantly with new information becoming available. Therefore the development of an assessment method for situation awareness for VTS may provide useful understanding of VTS operator performance.

The previous sections have shown many different approaches towards assessing situation awareness. Table 2-2 shows the criteria that may assist in selecting an appropriate method for assessing situation awareness and scores situation awareness assessment approaches on these criteria. None of the methods scores positively on all criteria. Each technique has its strong and weak points. A suitable method has to be selected depending on the objectives of the research, the project definition and constraints.

There had been until the work described in this thesis no previous attempts to assess VTS operator situation awareness. The projects described in this thesis offer the first opportunity to investigate this issue. Therefore it was imperative to develop a method, or methods, that provided us with as much information about VTS operator situation awareness as possible. Therefore a direct measure of situation awareness was preferred. In general on-line methods provide more understanding than off-line methods.

The COMFORTABLE project in which the first method was developed focused on the development of new equipment for VTS stations. These new tools were not yet applied in real-life, but were being tested in an experimental situation. Therefore application of the method in real-life was not necessary, but it was important that the method was suited for application in an experimental setting.

VTS operators are in their work regularly confronted with disturbances. Therefore it was expected that the obtrusiveness of the method would not be a big issue, although the method had to be tested for it.

Concluding: Important criteria were:
- Direct assessment of situation awareness
- Method provides as much information on situation awareness as possible
- On-line methods are preferred
- Obtrusiveness of method has to be tested
- Method should be suited for experiments

From Table 2-2 there are two categories of methods that score well on these criteria:
- Methods that apply the freeze technique
- Methods that apply multi-dimensional rating scales of situation awareness rated either by the operators or by observers.

These methods provide the most understanding of situation awareness, while at the same time they fulfil the other requirements. The two methodological approaches are based upon different conceptions of situation awareness. There is a fundamental difference between the two approaches.

Methods that apply the freeze technique try to find situation awareness in the situation presented itself: perceived elements in the situation are built into models that lead to understanding and projection of future developments.

Methods that apply multi-dimensional rating scales of situation awareness try to find situation awareness by breaking down the concept into underlying mechanisms, such as attentional demand. This approach leads to better understanding of cognitive processes, but not necessarily to better understanding of the situation.

The COMFORTABLE project was interested in the work of VTS operators, not in cognitive processes per se. Therefore the freeze paradigm and the SAGAT method were chosen as a template for the development of the first method for the assessment of VTS operator situation awareness.

After the completion of the work with the first method, the method was evaluated against the criteria mentioned here. This evaluation is described at the end of Chapter 3 in section 3.5.1. This evaluation led to the approach for the second method for assessing situation awareness, which is a task performance based method. This method scored lower on the scale of assessing direct situation awareness, but it will be shown that the advantages of the method outweighed this disadvantage.

The next chapter describes the first method that was developed to assess VTS operator situation awareness. This method is a method for direct assessment of situation awareness.

CHAPTER 3

SATEST

3.1 Introduction

This chapter describes the development and application of a method for direct assessment of VTS operator situation awareness. The objective of this study is to increase the understanding of VTS operator situation awareness in relation to VTS operator performance. The basis for this method is Endsley's SAGAT methodology, described in section 2.6.1.2. SAGAT was first developed for use in military aviation, to test situation awareness of pilots in air-to-air combat. SAGAT queries on situation awareness refer to aircraft system status, such as own heading, own location, G-level, fuel level, etcetera and external environment, such as presence and position of friendly or enemy aircrafts.

Supervisory control tasks, such as VTS (and ATC) are very different tasks than flying an aircraft, and require a completely different type of overview. Section 2.4 has shown that the maritime traffic that has to be monitored in VTS consists of complex interaction between ships. The relations between ships and their potential interactions and conflicts are more important than the actual position, speed and course of individual ships. With increasing numbers of ships in a traffic area the potential number of relations and conflicts grows exponentially. VTS operators have to anticipate potential conflicts up to ten or twenty minutes ahead in time. The method that was developed to assess VTS operator situation awareness, described in this chapter, is not limited to asking details on individual parameters, but captures the complete image of the VTS operator at the freeze moment. The level of detail in representation of traffic is important for the assessment. When the work on this method started, the significance of this difference in tasks was not well understood. Only during the process of testing SATEST and analysing the SATEST data, did the significance of these differences become clear.

During the larger of the two experiments described in this chapter, in Rotterdam, several problems in the collection of the data occurred. As a result the data set that could be analysed did not contain enough data for statistical analysis. Therefore the results of the study have to be regarded more as a basis for discussion, than as a test of hypotheses.

3.2 Background to the development

A first method for assessing situation awareness in VTS operators was developed in the COMFORTABLE project. For further information see (Butter, et al., 1998).

At the start of the project, in 1996, observations were made of the VTS work and interviews were held with VTS operators and professionals involved in training VTS operators. Together with a literature study on available models for assessing VTS operator performance, this formed the basis for developing a method for assessing VTS operator situation awareness (Wiersma & Heijer, 1996).

The VTS simulator at MSR, a maritime training and research institute in Rotterdam, was used to develop and test the method in 1997. Several VTS operators, ranging from trainees to operators with more than 25 years of experience, freely participated as subjects during the development and try-out of the method.

The first experiment using SATEST was also carried out at MSR in Rotterdam in 1997-1998. VTS operators of the Rotterdam Port Authorities participated as subjects. The method and results of this experiment were presented at several conferences (Wiersma & Mastenbroek, 1997; Wiersma, Mastenbroek, & Wulder, 1999; Wiersma & Hooijer, 1997; Wiersma & Mastenbroek, 1999) and were also presented to and discussed at a user-forum of VTS operators and COMFORTABLE researchers in a workshop (Jarvis, 1997; Regelink & Jarvis, 1999). The method was received with interest and the value of the approach was recognised. Further development of the approach was recommended. The COMFORTABLE project leader was able to find additional EC funding, to conduct an extra experiment at VTT in Finland, applying the method in an archipelago area, with Finnish VTS operator as subjects (Wiersma, Jarvis, & Granholm, 2000).

3.3 Description of the method

SATEST is an experimental method for measuring situation awareness in VTS operators, using the "*freeze*" methodology and queries. It is based upon the Situation Awareness Global Assessment Technique (SAGAT) developed in aviation research (Endsley, 1995a). In SAGAT, pilots have to fly a range of simulator flights. At specified times (unknown in advance to the pilots) during these scenarios the simulation is stopped and pilots have to answer a number of questions about the situation, to reveal their situation awareness. The accuracy of the responses is then compared with logged situation data. Thus a measurement of situation awareness is derived.

In SATEST, scenarios of maritime traffic are run on a simulator. At specified moments, also unknown to the operators, the simulator is stopped and questions pertaining to the traffic situation are asked. The objective of SATEST is to get a (complete) recollection of important elements in the traffic situation. Individual scores are evaluated on the basis of the actual traffic situation, taking into account the specific characteristics of VTS operator work, such as accuracy and relevance of information.

Questions in SATEST pertain to all three levels of situation awareness; perception of elements; comprehension of current situation and projection of future status. This includes a consideration of the system status as well as relevant aspects of the external environment.

This section describes the choices made in developing and implementing SATEST and the requirements for a successful implementation of the method. It considers the following aspects: VTS traffic simulator; scenario; type of information collected; scoring system; and data analysis.

3.3.1 VTS traffic simulator

SATEST uses simulated maritime traffic, presented to the operator in a simulator. Depending on how extensive the simulator is, the task can consist of monitoring traffic, looking up information in databases, interacting with traffic or a combination of these tasks. Since the method follows the assessment procedure of SAGAT, the simulator must fulfil the following requirements:

r1. The simulator can be stopped at specified moments, at which the operator is tested. This is necessary to be able to use the "freeze" method,

r2. The traffic information of the "freeze" moment is logged in such a way that the answers of the operator can be compared to the actual situation.

The simulator itself is not an integral part of the method. In principle any simulator that can simulate maritime traffic can be used with SATEST. Having said that, it must be stated that if an operator performs his experimental task in circumstances comparable to those in his daily routine it may be expected that he will build up his situation awareness likewise. Therefore, using a realistic simulator and test environment increases the ecological validity of the research. This puts forward another requirement for the simulator:

r3. The simulator presents maritime traffic in a realistic manner.

SATEST was developed together with Maritime Simulation Rotterdam (MSR)[14]. This company had its own full scale interactive VTS simulator, which was used during the development of the method and in the experiment in Rotterdam.

---

[14] MSR is currently Part of the STG-Group

For the experiment in Finland another simulator was used. This was not a professional VTS simulator. Vessel traffic in the Helsinki VTS area was presented on a two screen PC-based system. Ships were moving along tracks according to a predefined scenario. Ships could not deviate from these tracks. Communication with the traffic was not possible. Operators were not aware how the simulator functioned. Post-test interviews with operators showed that they accepted the experimental set-up as a valid representation of their work environment and the differences in simulator set-up did not measurably affect their behaviour (see section 3.4.3.3.2 for more details).

Both simulators used in the studies in Rotterdam and Helsinki, could be stopped at will (r1) and record all information necessary for analysis (r2). Thereby the first two requirements were fulfilled in the experiments carried out in Rotterdam and Helsinki. There are no hard criteria for the third requirement (r3). The validity of the simulators was discussed extensively with operators and trainers during the experiments. Whether this requirement was fulfilled will be discussed in the sections that describe the experiments carried out with these simulators.

3.3.2 Scenario

The SATEST method does not need a fixed scenario to test VTS situation awareness, in principle any traffic scenario can be used, provided that it fulfils the following requirements. Comparable to the requirements for the simulator, the requirements for the scenario have both methodological and validity aspects.

- r4. The scenario is long enough to enable the VTS operator to build situation awareness of the situation between the breaks.
- r5. The scenario is of a sufficient quality to enable the VTS operator to forget that it is an experiment.
- r6. The scenario presents traffic and infrastructure that are tuned to the traffic conditions and work environment of the operators that are tested.

A scenario that presents realistic traffic in a realistic environment has more potential to produce results that have value for the VTS work in practice. It is not unusual for VTS operators to be interrupted during their work. Therefore having breaks in the scenario was not considered a problem in the validity of the task environment, but operators need time between the breaks to rebuild their situation awareness of the situation. The first requirement (r4) was tested during the try outs of the method.

To fulfil the other requirement (r5 and r6) the scenarios were developed by experts. They were adapted from existing training scenarios (in Rotterdam) or based on actual tracks of ships (in Helsinki). These aspects too were discussed and tested with professionals during try-out sessions, preceding the actual experiments, for validation.

At specified moments during the experiments the simulation is stopped and questions pertaining to the situation are asked. There are two approaches towards selection of these "*freeze*" moments: they can be selected at random or at moments of specific interest for the research question. Both were used in this research. The requirement for the break moments is as follows.

r7. After the completion of his task the operator is able to continue his task without problems

r8. In the description of the experiments there should be a rationale for the selection of break moments.

At specified times the simulation was stopped and the above questions relevant for VTS were asked about the situation. Responses to the questions asked were recorded by means of a personal computer using a computer program in Microsoft Visual Basic 4.0© for retrieving and storing information (Heijer,1996, unpublished).

3.3.3 Information collected

In the SATEST experiments the VTS operator is questioned about his knowledge pertaining to the exact moment the simulation stopped. The information collected at the freeze moments must give a clear picture of the operator's situation awareness at that moment. A description of VTS operator situation awareness on all three levels is presented in section 2.4. Based on this description, two important characteristics of the traffic were identified that need to be taken into account in developing a method: the temporal nature of the traffic and the spatial nature of the traffic. The spatial nature of the data is taken into account in the scoring system. The temporal nature of the data is considered in the experimental design.

A hierarchically structured score form was developed on which operators could fill in the following aspects of each situation.

The *first* level of information collected was an overview of the traffic situation. The score sheet consisted of a map of the VTS area, showing the infrastructure. This map was used to indicate:

- The exact position of all ships
- The speed and course vector of each ship.

On this form the operator made a complete reproduction of the traffic situation in the same format as it was presented on the VTS monitor. Small circles denoted ships; a vector indicated their speed and course. An example of form is given in Figure 3-1.

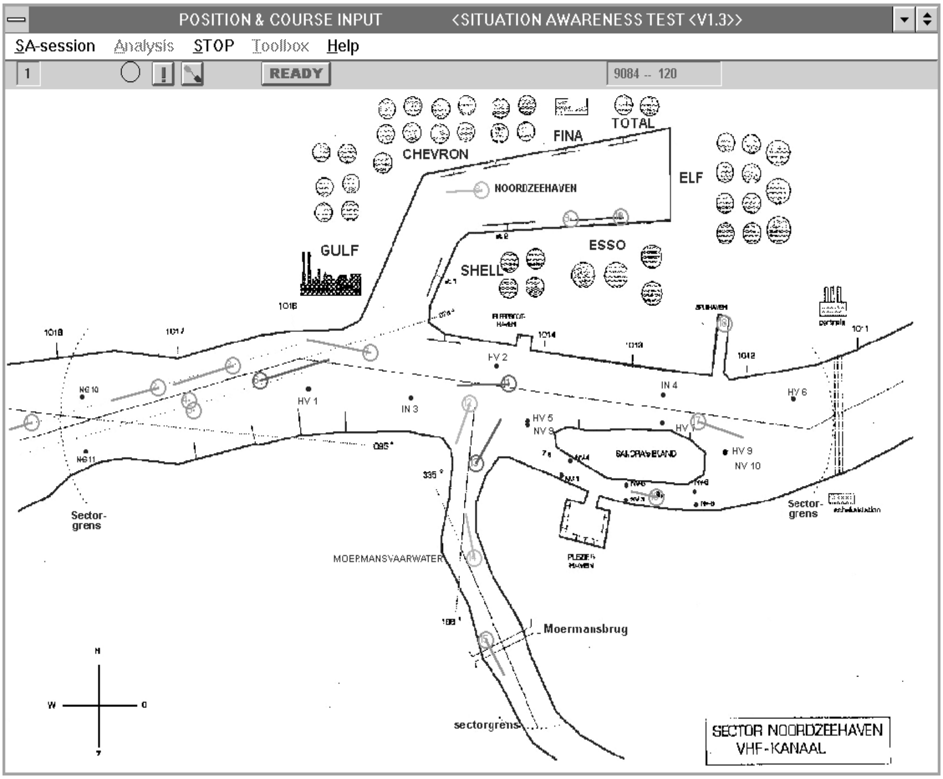


Figure 3-1 Filled in form (1) used in SATEST to describe traffic situation used in Rotterdam

The *second* level of the score form was used to collect information for each ship individually. This form opened every time a single ship was selected by clicking the mouse. On this form the operator was asked to fill in the following aspects:

- Label/name of the ship;
- Has there been communication with ship (Yes/No);
- Origin and destination (Categories provided depended on the area of the simulation);
- Type of ship (Seagoing, inland, tug, chemical);
- Speed (slow, normal, fast).

An example of form is given in Figure 3-2.

Level (1) and level (2) score forms together provided the required information on the *traffic image*. These are closely related to Level I situation awareness, perception of elements. Level (3) score forms provided information on the "macroscopic" aspects of the traffic situation: complexity, conflicts etc.

Figure3-2 Form (2) used in SATEST for information of individual ships

The *third* level score used the already completed form (1). The operator was asked to indicate in colour to which of the ships he had to pay normal attention (green), increased attention (orange) or high attention (red). And ships in conflict situations.

An example of a filled in third level form can be seen in Figure 3-3.

Figure 3-3 Form (3) used in SATEST with indication of increased attention

In this manner, information was collected on the operator's situation awareness of the traffic situation, including elements such as perceived conflicts and potential conflicts between ships, and predictions of where conflicts were likely to develop.
Table 3-1 shows how the traffic information collected relates to the different levels of situation awareness described in Chapter 2.

Ship data	SA Level
Name; origin; destination; type of ship	Level I
Position; course; speed; communication; agreements made	Level I
Needs extra attention	Level II, III
Is in conflict situation, type of conflict	Level II
Can/will be in conflict situation, communication is needed	Level III

Table 3-1 Relation of SATEST data with levels of situation awareness

The basic principle of the SAGAT paradigm (that was chosen as template for the development of SATEST) is the manner in which situation awareness data are collected. At specified moments the simulation is stopped and questions, which are relevant for the task, are asked about the situation. After answering the questions, the simulation continues and the operator has to continue his task, until the next break. Research in aviation has shown that a break of several minutes can be used without problems (Endsley, 1995a). This does not automatically mean that the same applies to the VTS domain, so this aspect was investigated. During the try-out experiment the issue of the length of freeze breaks was tested. All operators were able to continue their task without problems after the breaks, which were of approximately five minutes. The time between the end of one freezing moment and the start of another was typically a minimum of eight minutes. In the interviews at the try-outs (section 3.4.1) this was discussed and operators indicated that they considered the time long enough to enable them to regain their control and situation awareness of the situation.

3.3.4 SATEST Data analysis

The information gathered on the scoring forms was compared to the actual traffic situation. Thus a score could be achieved of the VTS operator situation awareness. To obtain an objective score, it was necessary that the scoring system be clear and unambiguous and that scores did not depend on the person doing the analysis.

SATEST data were analysed using two methods:

- Qualitative analysis by ranking the traffic displays completed by the operators;
- Quantitative analysis of completeness and accuracy of ship positioning in traffic displays.

3.3.4.1 Ranking of traffic displays

The analysis was performed by an expert operator, familiar with the scenario and the traffic development therein, someone who was able to determine what was important for the operator's situation awareness in the situation. Ranking of the traffic displays was performed by comparing the filled in score forms with each other and with the correct answers (taken directly from the scenarios). During the ranking the person performing the ranking was asked to think aloud about the considerations which led to the results obtained. Thus an insight could be obtained into the criteria that played a role in the ranking process.

3.3.4.2 Completeness and accuracy of traffic displays

The quantitative analysis compared the number of vessels and their location and course vector, indicated in the completed traffic displays, with the actual number of vessels and their actual location and course vector. Ships within one ship's length distance were considered to be located correctly.

The remaining data filled in by the subjects on the separate scoring forms, regarding the information about the ship, used for identification of the ships. This information was not separately scored for completeness and accuracy, since many operators had not filled in the complete forms, but had only used them to label the ships as they would do in their daily practise. Moreover, many operators did not use the official names to label the ships, but used abbreviations or invented their own labels. Since this corresponded with their normal routine, it was not considered right or wrong, and therefore these data were not scored separately. From the predictive data only the indication of increased attention was used regularly and analysed.

Although the analysis was fairly straightforward when performed by hand, automation of this analysis would not be simple. Automatic identification of the vessels is complicated by inaccurate locating of vessels, and by the fact that operators were asked to give their personal label to the ship, which does not necessarily correspond with the full name.

3.4 Application of the method

SATEST was developed, tested and applied in Work package 02 of the COMFORTABLE project. The aim of this Work package was to develop support systems for assisting VTS operator situation awareness in their daily work. Attention was paid to difference in traffic conditions in ports around Europe and the potential influence of this difference for support tools.

As a consequence, the method for assessing VTS operator situation awareness developed in this Work package was also tested in different situations. Experiments with SATEST were carried out in a high-density traffic port and approach area (the Port of Rotterdam), and in a non-dense traffic area (the Finnish Archipelago). The objectives of these experiments were to further develop and validate the method and determine its applicability in a wide range of situations.

Another opportunity for the development and evaluation that came out of COMFORTABLE was during the presentation of SATEST at the second of three workshops that were held during the project (Jarvis, 1997). The objectives of these workshops were to bring together researchers, suppliers, port authorities, managers, and VTS operators to discuss, evaluate, and contribute to future developments with the focus on the Human-Machine Interface. This workshop offered an opportunity to discuss the usefulness and validity of SATEST.

The remainder of this chapter discusses the results of the try-outs, the experiments, and the workshop, and conclusions that can be drawn about the SATEST approach for assessing VTS operator situation awareness.

3.4.1 Try-out

3.4.1.1 Aims

Situation awareness is a relatively new concept. This project was the first application of the situation awareness concept in Vessel Traffic Services. The methods used to study it in this project were also new and especially developed for this type of research. Therefore, it was necessary to conduct a number of test runs during the development of SATEST, to gain insight into the usability of the concept, and the applicability of the method and to investigate if the requirements set out in the previous section were met. All requirements except those for the data analysis were considered. The data analysis modules were developed and tested in the later experiments in Rotterdam and Helsinki.

3.4.1.2 Methodology

During the development of the SATEST method, several test runs and try-outs took place over a period of one year in 1996-1997.

SATEST requirements for the simulator:		
	r1.	The simulator can be stopped at specified moments, where the operator is tested.
	r2.	The traffic information of the "*freeze*" moment is logged in such a way that the answers of the operator can be compared to the actual situation.
*	r3.	The simulator presents maritime traffic in a realistic manner.
SATEST requirements for the scenario:		
*	r4.	The scenario is long enough to enable the VTS operator to build situation awareness of the situation between the breaks.
*	r5.	The scenario is of a sufficient quality to enable the VTS operator to forget that it is an experiment;
*	r6.	The scenario is long enough to enable the VTS operator to build his awareness of the situation.
*	r7.	After the completion of his task the operator is able to continue his task without problems
	r8.	In the description of the experiments there should be a rationale for the selection of break moments.

Table 3-2 Requirements tested during try-out

Each of the runs was preceded and followed by discussions with all researchers involved and with interviews with the VTS operators who participated as subjects. In these discussions the requirements presented in section 3.3 were evaluated. The requirements discussed are marked (*) in Table 3-2.

3.4.1.2.1 Traffic simulator and scenario

The development of SATEST was carried out in the COMFORTABLE project as a co-production of MSR and Delft University of Technology (Jarvis, 1997). Therefore it was a logical step to use the available VTS simulator of MSR in the process.

The MSR VTS simulator consists of a fully operational VTS terminal linked to an extensive maritime traffic simulator. Years of research effort have been put into designing and developing realistic traffic in this simulator, where ships behave as they would in reality, taking into account both ship characteristics and weather conditions, river flows and tidal effects. Numerous traffic scenarios can be chosen both in real maritime areas and imaginary areas, with pre-programmed traffic. A simulator instructor, who can also apply manual control of any of the ships within the scenario, supplies oral communication from ships, other VTS centres and services.

Instead of developing a new scenario for SATEST, a simulation session was chosen from the available training sessions at MSR (NNVO HC-03a). The advantage of this procedure was twofold: First the trainers were very familiar with the scenario and the developments therein and second it saved a lot of time.

The operators who participated in the experiment were not familiar with the scenario beforehand. The scenario was forty-five minutes long and took place in an imaginary harbour, to avoid variance in familiarity with the situation among subjects. In this period ships entered, sailed through and left the VTS area, some ships set sail and others docked. The complexity of the traffic situation, the number of active ships and ships on potentially conflicting courses, varied through the scenario. The total number of ships in the scenario was 18. All ships sailed along a pre-programmed track. The traffic in the scenario followed regular paths in the waterways, without unexpected or undesired movements of ships. All necessary communication between operator and ships was present in the scenario. At no point did the operator have to instruct ships to change course or speed. This was different from scenarios used in other experiments, see for reference section 4.3.2.1.

The VTS operator had to communicate with the ships to inform them about activities on the waterway and to solve emerging conflicts. The simulator instructor played the role of all captains and other VTS posts in the communication. The VTS operator therefore communicated extensively with this trainer.

3.4.1.2.2 Data registration

The first test runs with SATEST were conducted with paper and pencil, whilst the computational scoring system was developed. The operators first received a paper with the outlines of the infrastructure on which they could fill in the ships with a speed and course vector. When they had finished this first part, they filled in another score form for each ship individually, indicating the information requested.

The main advantage of using a paper and pencil test in the first phase of the development was that changes in a paper query are much easier to implement than in a computer score form, especially if not only the questions change, but also the manner of presenting them. The disadvantage of the paper and pencil method is that it takes a large amount of paper for each operator.

Whenever possible, the questions were presented in the form of multiple choices, to facilitate scoring and to minimise the time necessary to complete scores. The questions resulting from these trials are presented in Table 3-3. The items marked (*) were presented in the form of multiple choice. The score forms (in Dutch) are presented in Annex 2.

	Item
1.	Ship identification number
2.	Subject name
3.	Trial number
4.	Has there been communication with ship? (*)
5.	Name or label of ship
6.	Origin of ship
7.	Destination of ship
8.	Load or type of ship
9.	Speed of ship (*)
10.	Special remarks
11.	Agreements made
12.	Does this ship deserve special attention? (*)
13.	Is this ship in conflict situation? (*)
14.	Can or will this ship be in conflict situation later? (*)
15.	What type of conflict? (*)

Table 3-3 SATEST paper score form questions

Early in the project the possibility was proposed to develop a computerised version of the method. Advantages of a computer version of the method are simplification in storage, retrieval and processing of data. It was also expected that VTS operators would prefer data entry by computer over using paper-and pencil. The computer version of the method is the one described in section 3.3.3 above.

3.4.1.2.3 Subjects

Six VTS operators of the Rotterdam Port participated free of charge in the experiments during the try-out, five males, and one female. The experience range of the operators was large: Two of them had just finished their training for VTS operator and were ready to start working, the others were experienced VTS operators and were also trainers in the education of VTS operators. These operators not only knew the VTS work, but also the simulator very well. The variance in experience was considered useful to explore further the issues regarding the build up of situation awareness and the acceptability of the simulator environment.

3.4.1.2.4 Procedure

The simulator instructors received a briefing protocol to ensure that they would act the same for all subjects, and that they would operate the simulation session identically. A written protocol stated at which moments during the scenario the instructors had to take the initiative in communication.

The VTS operators who participated were already familiar with the simulator, on which they had practised. They were first given an introduction to the experiment and materials used. This introduction explained in four pages the objective of the project, the difficulties in assessing VTS operator performance, what the concept of situation awareness entailed, and gave a step-by-step description of the experiment and score forms used. The protocol used for this introduction (in Dutch) is in Annex 2.

Next the subjects were familiarised with the testing procedure by way of a try-out standard simulator run of five minutes with only four ships, followed by a freeze with a scoring session. During this try-out questions could be asked and incorrectly filled in score forms were pointed out. Special attention was given to operators who did not fill in the form completely, although they were aware of more information. Sometimes operators thought they had filled in "enough already", without trying to be complete. It was emphasised to these operators that it was very important to fill in the queries completely, with everything they were aware of.

The actual experiment consisted of the previously mentioned simulation run, which was standard for all subjects. At four pre-specified times the simulation was stopped and subjects were asked to fill in the traffic image and the score forms.

3.4.1.2.5 Analysis

At this stage no formal method had been established to score the results of the operators. The focus of the project was still on the data collection and retrieval, not on the analysis. The results were examined by a skilled VTS operator trainer who was very familiar with the scenario (and who knew the conflicts and points of interest). In a debriefing session after the experiment the operators and this trainer discussed the results of the experiment and how well the operator had performed. The scores achieved in the experiments were not really important, except for the operators themselves; they were very interested in how well they had performed on the test. More interesting than the actual scores achieved was the question if the method fulfilled the requirements set out at the beginning of this section. The debriefings were used to get an answer to these questions.

3.4.1.3 Results

The attitude of the operators towards the experiments was very positive. In general the operators did think that good situation awareness was necessary for their work and that a method for assessing it could be useful. They considered SATEST a useful addition to existing methods of operator performance assessment and were pleased to be involved in this research, to develop the method.

3.4.1.3.1 Fulfilment of SATEST requirements for the simulator

The simulator could be stopped at will and did produce all the information necessary for the analysis, fulfilling the first two requirements. However, the simulator screen could not be blanked at the freeze moments, but showed the final situation. Operators had to be taken away from the simulator to prevent a long "final glance" to fix the positions of ships in their memory. As a result the procedure for scoring was refined: The experiment leader took the operator away from the screen as soon as the scenario stopped. Furthermore the computer with the scoring forms or the paper scoring forms were placed in such a way that the operator could not see the screen when filling in the forms.

Even though the simulator was designed and used to train VTS operators for their work, the issue of fidelity was addressed in the debriefing interviews. The operators stated that the simulator was adequate in this respect, although some operators would have preferred a simulator with an interface that exactly matched the one they used in their workplaces.

3.4.1.3.2 Fulfilment of SATEST requirements for the scenario

What has been stated with respect to the simulator also held for the scenario used; it was especially designed to reflect the natural working conditions of the operators. The scenario presented traffic and infrastructure that was well suited for Rotterdam VTS operators.

Although the simulator and scenario were of high quality, the VTS operators did not completely forget that it was an experiment. However, the simulated environment and scenarios were so good that some operators every now and then used the names of waterways in their actual work environment, instead of the official names given to the waterways in the artificial area.

The issue of resumption of the task was discussed in the debriefing. The breaks were four to eight minutes long, depending on the time needed to fill out the scoring forms. The operators stated that they could almost immediately resume their work after the breaks, and did not need recovery time. The breaks were set five to sixteen minutes apart, apart (the time between resumption after one break and the start of the next break), which was considered enough by all operators to resume situation awareness of the traffic situation.

3.4.1.3.3 SATEST information collected

The issues on requirements of information (selection of freeze moments; type of data collected; and collecting information that can be measured) were studied in this try-out phase. The information collected on VTS operator situation awareness was discussed in the debriefing. The remarks of the operators were taken into account and led to improvement of the score forms for later experiments.

3.4.1.3.4 SATEST scoring system

The paper score form used at the beginning of the trial sessions was constructed in such a way that the time needed to fill in the form was minimised, to make sure the breaks in the VTS run were as short as possible. However, the time needed to fill in the complete questionnaire still ranged from four to eight minutes. This time was therefore often longer than the five minutes Endsley allows (Endsley, 1995a), but the VTS operators did not seem to have a problem returning to their task after a break.

During the trials it became clear that some operators preferred to work with the computer forms, while others preferred to work with paper. Apparently some operators were not accustomed to working with a mouse, which was necessary for exact and fast input of the information in the computer. In later experiments both computer and paper versions of the score form were offered to the operators. Both methods yielded the same results and took the same amount of time.

3.4.1.4 Conclusions

The concept situation awareness was received with enthusiasm by the VTS trainers and operators. The method seemed to be adequate and most of the requirements set were fulfilled in the try-out in the Rotterdam situation: the simulator and scenario were sufficient and the scoring system was already well worked out towards the end of the try-outs. This left only the questions of data analysis and issues of reliability and validity for further study.

3.4.2 Rotterdam: Full scale test

3.4.2.1 Aims

The first full-scale test of SATEST was an experiment carried out in Rotterdam in the first half of 1997. This experiment was designed to provide a baseline measurement of situation awareness in a dense traffic port and approach area by using a ‘conventional’ VTS Traffic Situation Display. Future enhancements in the display (developed in other parts of the COMFORTABLE project or elsewhere) could then be evaluated against the results of this experiment. The experiment was also used to develop the SATEST method for data analysis and to validate the method.

Due to problems during the tests, the data of 6 of the 21 operators that participated in the experiment could not be analysed. As a result the analyses could not be performed as planned and hypotheses could not be tested. This changed the function of this experiment in the total study. The present function of this experiment is putting forward hypotheses and shaping thoughts about VTS tasks and situation awareness.

3.4.2.2 Methodology

3.4.2.2.1 Traffic simulator and scenario

The simulator and scenario described in the try-out experiment were used again in this experiment, but with a new set of subjects. The situation and traffic in this scenario are considered representative for the traffic in the Rotterdam area (see Try-out experiment). The complexity of the traffic was not constant throughout the scenario. The freeze moments were chosen to sample different levels of complexity in terms of number of ships present and potential conflicts. The number of ships in the scenario ranged from eleven to eighteen.

3.4.2.2.2 Hypotheses

This test was the first full experiment in VTS operator performance. Therefore hypotheses had to be formulated regarding this performance and its relation to situation awareness.

The first hypothesis was that experience leads to better performance. Over the years operators learn to interpret traffic situations better, knowing where to focus, while at the same time keeping overview. Novice operators are not yet able to perform as well as experienced operators. The assumption is that it will take a few years of practise until an operator is not a novice any more. Experienced VTS operators are therefore hypothesised to have better situation awareness than novice VTS operators and trainees. In the experiment experience of operators is used as an independent measure of how good an operator is.

Hypothesis 1

SATEST is regarded a valid instrument for assessing VTS operator situation awareness if it demonstrates differences in situation awareness between novice VTS operators and experienced VTS operators.

Better performance can be shown in several fields. With increasing complexity of the situation, there will come a moment for everyone, where it is impossible to keep every detail of the situation in mind. The second assumption is that a good operator will at that moment maintain situation awareness of the important elements, disregarding less important information. Being able to distinguish important information from information that can be disregarded and maintaining situation awareness of the first is hypothesised to differentiate good operators from not so good operators.

Hypothesis 2

In a demanding traffic situation the attention of a good VTS operator focuses upon potential conflicts, in the meantime disregarding less important parts of the traffic situation. SATEST will distinguish this from the less good VTS operator, who will have his attention divided over all elements contained in the situation, without regard to their importance.

Maintaining overview of a situation alone is not sufficient. An operator also has to act upon the situation as it develops. Early recognition of potential conflicts and communication with the ships involved is considered positive. The assumption is that experience helps an operator to learn when and how to react to a developing situation. Good VTS operator performance and good situation awareness include solving evolving potential conflicts at the earliest moment by communication to ships. If ships are aware of an oncoming potential conflict, they will take the appropriate action themselves most of the times. The role of the operator in those situations is to make ships aware of the potential danger.

Hypothesis 3

SATEST will distinguish the more experienced VTS operator with better situation awareness, who recognises a potential conflict earlier and acts upon it earlier. In this scenario the action will consist of elimination of the potential conflict through communication.

3.4.2.2.3 Subjects

In the setup of the experiment 24 VTS operators were to participate. They were to be VTS operators working at the Rotterdam Port VTS, divided into three categories: 8 trainees, 8 operators with some (three years) experience, and 8 experienced operators, with more than 10 years of experience as VTS operator.

However, due to technical failures the full experiment could not be completed. Several of the subjects could not be tested properly. It also turned out in the course of the experiment, that the VTS operators were more experienced than expected. There were fewer trainees and operators with less than three years of experience available for the experiment than required. Except for the trial runs, no trainees participated in the experiment. To solve this problem for the analysis, operators were divided into two groups: operators with less than 3 years of experience as VTS operator were categorised as *Novice*, operators with more than 3 years of experience were categorised as *Experienced*.

Twenty-one subjects (20 male, 1 female) participated in this experiment. 8 Operators were categorised as inexperienced (0-3 years of experience, average 1.6) and 13 were categorised as experienced (6-30 years of experience, average 15.1).

After this, the problems were not over. Out of 21 operators that participated, the data of only 15 operators could be used for the analysis. The data of the remaining operators were lost due to problems with bugs in the data registration software and other (computer) malfunctions. In the final data set of fifteen operators, four operators were categorised as *novices* and eleven operators were categorised as *experienced,* making statistical analysis impossible.

The operators were taken out of their daily operational routine for a day of experimenting. For a number of operators the experiment could be combined with training, for which they had to be at MSR. For these operators, the experiment was carried out after they had finished their day's work at the MSR training centre. The other operators had to travel especially to MSR for the experiment after their normal work. Subjects participated voluntarily, without extra payment.

3.4.2.2.4 Data registration

At the start of the experiment, only part of the computerised SATEST score form was available. During the experiment, the fully computerised version of SATEST was developed. This version of SATEST runs on a stand-alone personal computer. When both versions were available, operators were offered the choice between a paper-and-pencil test and the computerised version of the method. Some VTS operators, the ones who were not very skilled in using a computer mouse device, preferred the former method. Earlier try-outs had already shown that there were no differences in the results of the two methods, and that they took the same amount of time. Therefore, it was decided to allow the use of both the paper-and-pencil score forms and the computerised answers in the analysis.

This was the first experiment using situation awareness to assess VTS operator performance. Therefore the experimenter discussed the results of the scoring with the operators, to evaluate how they reached their answers. This discussion did not lead to changes in the scores themselves.

3.4.2.2.5 Procedure

First the operators were introduced to the objectives of the experiment. A standard protocol (in Dutch, see Annex 2) was used, explaining the goals of the research and the SATEST method. After that they were given some time to get familiar with the VTS simulator and the SATEST program and score forms were introduced. Next the experiment started with a 60 second trial run. At the end of this period, the operator had to fill in the SATEST score form. This trial was necessary to familiarize the operator with the SATEST scoring system.

After this test-stop the scenario continued. At predefined stopping moments (see next section) the simulation was stopped. At these moments the screen blanked and the operator was asked to fill in the SATEST program.

The operators were asked to indicate for all ships the following (see section 3.3.3):
- Position of all ships
- Course and speed vector
- Information on ships, name, origin, destination, type of ship
- Communication with the ship
- Conflicts between ships or between ships and infrastructure
- Whether there had been communication about these conflicts
- Which ships needed extra attention

The total amount of time for the experiment was approximately one hour and a half.

3.4.2.2.6 Scenario and stopping moments

The experiment was carried out using one fixed scenario of 45 minutes, in which actions of the operators did not have any influence on the actual movements of the ships. The scenario was an adapted version of a training scenario used in the training of VTS operators. Within this scenario several situations developed with different levels of complexity to which the operator had to respond. The complexity of a situation was expressed in terms of the number of ships present and the number of ships that were potentially involved in a conflict. The order in which the situation developed was fixed within the scenario. The scenario contained several of those potential conflicts.

Breaks in the scenario were chosen based upon expert judgement of the VTS trainers. Three situations were selected with different level of complexity.

1. The MODERATE situation contained eleven ships and two potential conflicts each involving three ships. This situation was considered moderately difficult.
2. The SIMPLE situation contained fifteen ships, but only one potential conflict between three ships. This situation was considered simple.
3. The COMPLEX situation contained eighteen ships and three potential conflicts involving two or three ships. This situation was considered complex.

Figure 3-4 presents the MODERATE situation used in the experiment. The conflicts used in the scenario are described in detail in Annex 2. Having just one scenario in the experiment was a consequence of using the full-scale VTS simulator at MSR. It was, in the time available for the study, not feasible to develop several scenarios with conflicts in different orders.

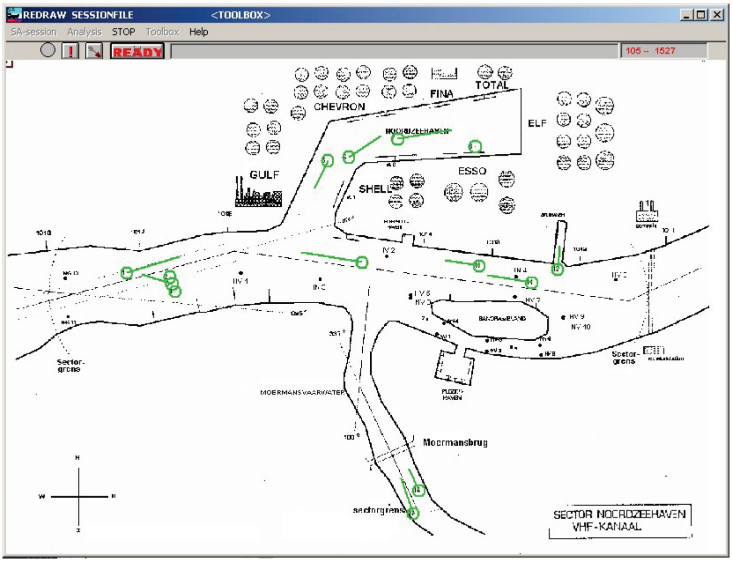


Figure 3-4 Example of a traffic situation with a moderate conflict

For each of the three situations, three breaks were selected around the selected MODERATE, SIMPLE, and COMPLEX conflict situations. These moments where the simulation was stopped were chosen based on expert judgement of the developers of the scenario and were discussed with the operators participating in the try-outs described in section 3.4.1. The three conditions were labelled EARLY, NORMAL and LATE.

1. The EARLY condition was defined as the first moment an experienced operator could be expected to recognise a conflict;
2. The MIDDLE condition was the moment at which all operators should have recognised the conflict;
3. The LATE moment was the last moment an operator could effectively act, to avoid the conflict. This last moment was not the Closest Point of Approach (CPA) of the ships; VTS operators are expected to react to a situation long before this moment.

Figure 3-5 shows the pre-conflict (EARLY) situations of the moderate conflict situation. Although to the casual observer the circled ships may be far apart, an experienced VTS operator will recognise that these ships are in a potentially conflicting situation, some ten minutes ahead. Experienced operators usually do not wait long when they see potential conflicts and try to solve them before they get critical.

Each VTS operator was presented a scenario with five breaks. The first break was after one minute in the scenario, prior to the conflicts. This was the on-line try-out, to practise the recording environment. Next three experimental breaks were presented. Each operator had one break early in a situation, one middle break, where all operators would normally act to solve conflicts, and one break late in a situation. A final break was presented after 45 minutes to all operators.

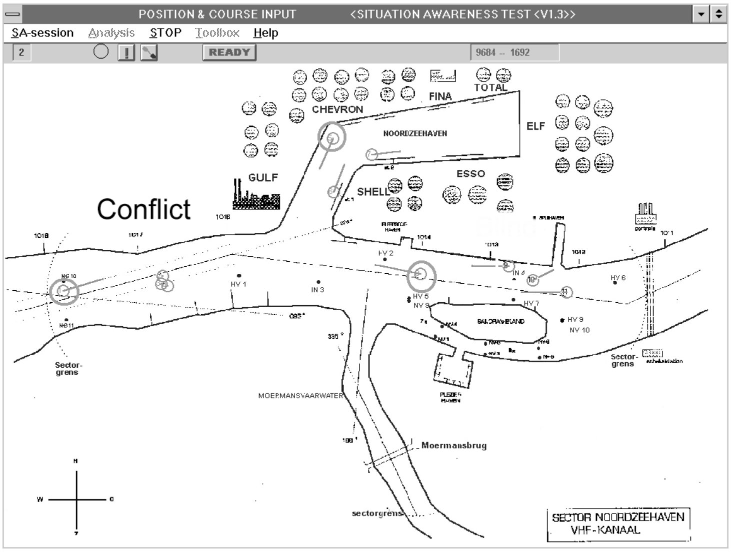


Figure 3-5 Potential conflict situation in early condition, some 8 minutes ahead of point of no return

At this time all conflicts in the scenario had been resolved and the situation was in a final stage, with no conflicts pending. This break also announced the end of the experiment. The presentation schedules of the breaks around conflicts are presented in Table 3-4. Operators were assigned at random to one of three possible schedules, presented in this table.

Time	Complexity	Condition	Schedule 1	Schedule 2	Schedule 3
0’	Start scenario		*	*	*
5’	Moderate	Early	*		
8’	Moderate	Middle		*	
10’	Moderate	Late			*
14’	Simple	Early		*	
17’	Simple	Middle			*
20’	Simple	Late	*		
35’	Complex	Early			*
37’	Complex	Middle	*		
40’	Complex	Late		*	
45’	End of scenario		*	*	*

Table 3-4 Time line of breaks in the scenario and schedules for operators

3.4.2.2.7 Analysis

The experiment in Rotterdam was the first test with the SATEST method. The methods for analysing the data were tested and further developed during the analysis of the results of this experiment. The data were analysed using two methods:

- The first method for analysing the data was qualitative analysis by ranking the completed SATEST score forms.
- Second, the qualitative analysis provided the input for a quantitative analysis of SATEST data.

Ranking of traffic displays

The objective of ranking the traffic displays was to get a grip on the criteria and dimensions that were, or could be, used to discriminate between "good" and "bad" performance on SATEST.

One qualified examiner of VTS trainees who was trainer in VTS refreshment courses (from which the scenario was taken) performed the ranking. This trainer was very familiar with the scenario and the traffic development performed therein.

Ranking of the traffic displays was performed by comparing the completed score forms with each other and with the correct answers (taken directly from the scenarios). Differences between the traffic situations (ranging from simple situations to very complex situations) made it hard to use absolute measures. Therefore an ordinal ranking of the score forms between operators on the same schedule was applied.

During the ranking the person performing the ranking was asked to think loud about the considerations which led to the results obtained, and the verbal comments were written down. Thus an insight could be obtained into the criteria that played a role in the ranking process. This was done to be able to objectify this scoring procedure in the future. In the Rotterdam experiment these comments were considered more important than the actual ranking score, because they would help to understand how a score was being derived.

The ranking of the final stop, which was the same for all operators, was repeated by the same scorer after two months to validate the consistency of the scoring.

Completeness and accuracy of traffic displays

The first quantitative analysis performed was comparing a simple count of the number of vessels indicated in the traffic displays with the vessels actually present in the situation. This preliminary analysis aimed at assessing the operators' performance. Both the number of ships reported and their position and course were taken into account.

The second analysis was more an analysis to assess the situations. A normative "correct" score of where operators needed to put their focus of attention was defined by the VTS trainers. This normative score was constructed through discussion between the developers of the scenario and the operators participating in the earlier try-out experiment. The normative "correct" score was compared to an analysis of the focus of attention that operators indicated in the completed traffic displays and communication with traffic. This analysis was carried out by hand. The normative scores are presented in Annex 2.

Results

Fifteen operators were used in the final data set. The novice and experienced operators were not divided equally over the different conditions (see Table 3-5 and Table 3-6).

Schedule	Novice operators	Experienced operators
Schedule 1	3	2
Schedule 2	1	4
Schedule 3	0	5

Table 3-5 Participating novice and experienced operators divided over schedules.

	Minute									
	5	8	10	14	17	20	35	37	40	45
Novice	3	1	0	1	0	3	0	3	1	4
Experienced	2	4	5	4	5	2	5	2	4	11
Total	5	5	5	5	5	5	5	5	5	15

Table 3-6 Participating novice and experienced operators divided over stopping moments.

3.4.2.2.8 Ranking of traffic displays

To test if SATEST can be used to demonstrate the differences between novice and experienced operators, the completed score forms of the fifteen operators that formed the dataset for this test were ranked for all of the breaks.

The break at 45 minutes had been tested with all fifteen operators. The other breaks were filled in by five operators each. Even though the break at 45 minutes was not the most interesting in terms of traffic situation, the scoring for this break was repeated after two months (2^{nd}) to check the consistency of scoring by ranking.

The results of the two rankings of the break at 45 minutes are presented in Table 3-7, where rank 1 is the best score. Experienced operators are labelled E1-E11. Novice operators are labelled N1-N4. The correlation between the two scores shows a reasonable consistency over time (0.7934, $p<.001$).

	Rank Minute 45		
Operator	1st	2nd	Ave
N1	1	2	1.5
N3	3	1	2
E10	2	3	2.5
E1	4	6	5
E11	5	5	5
E5	9	4	6.5
E3	7	8	7.5
N2	6	10	8
E7	10	7	8.5
E4	8	12	10
E2	12	9	10.5
E8	13	11	12
E9	11	13	12
N4	14	14	14
E6	15	15	15

Table 3-7 Ranking of operator SATEST scores on break at 45 minutes

Table 3-7 shows that in the 45th minute novices score better with an average rank of 6.38 than experienced operators, who have an average rank of 8.59. Since this break was at the end of the traffic scenario, where the interesting traffic situations have dissolved, not much consideration should be given to this observation.

Comparison of the other breaks in the scenario provides a better basis to discuss the differences in performance between novice and experienced operators. These breaks were selected around interesting situations in the scenario (see section 3.4.2.2.6). Table 3-8 shows the ranking of operators over these other three breaks in their scenarios. The columns show the ranking of the five operators tested for each of the breaks within their groups. The final column shows the average ranking score for each of the operators.

	1st	2nd	3rd	Ave
Group 1				
E1	1	1	1	1
N1	2	2	2	2
N2	4	3	3	3,3
N3	3	5	4	4
E2	5	4	5	4,7
Group 2				
E3	1	1	1	1
E4	2	2	4	2,7
E5	4	4	2	3,3
N4	5	3	3	3,7
E6	3	5	5	4,3

Table 3-8 Ranking of operators SATEST scores over scenarios

Group three only consists of experienced operators and is left out of this analysis. There are not enough scores to test the results statistically, so only the direction and size of the differences can be shown.

The consistency of scoring over operators over the different situations in the scenarios is large in general, especially for the high ranking operators (E1, N1, and E3). They have consistent ranks over all breaks. This indicates that these operators did perform consistently well throughout the complete test run. The ranks assigned to the lower ranking operators are less consistent, indicating that these operators performed better on some breaks than on others, even though their overall performance was less adequate than that of the top ranking operators.

The scores of novice and experienced operators also have been compared to investigate if experienced operators score better on more complex situations than novice operators. Table 3-9 shows the averaged score of the ranking of experienced and novice operators over the three breaks. Table 3-9 shows that over the three test situations experienced operators score somewhat better with an average rank of 2.9 than novice operators, who have an average rank of 3.3.

	1st	2nd	3rd	Ave
E (1-11)	2,8	2,9	3	2,9
N (1-4)	3,4	3,4	3,2	3,3

Table 3-9 Average ranking of experienced and novice operator scores over all breaks

In Table 3-10 the breaks are grouped according to the complexity of the situation. Group three only consists of experienced operators and is left out of the analysis. Ranking of the completed score forms shows that experienced operators score better on the simple and moderate situations, but that no differences can be shown between the groups on the complex situations. The effect of the consistency of operators is larger than the effect of the situation (see also Table 3-8).

	Simple		Moderate		Complex	
Rank	14’	20’	5’	8’	37’	40’
1	E3	E1	E1	E3	E1	E3
2	E4	N1	N1	E4	N1	E5
3	E6	N3	N2	N4	N2	N4
4	E5	N2	E2	E5	N3	E4
5	N4	E2	N3	E6	E2	E6
Ave (N)	3.5		3.3		3	
Ave (E)	2.7		2.8		3	

Table 3-10 Ranking of operator SATEST scores grouped according to complexity of situation

The scores of novice and experienced operators also have been compared to investigate if experienced operators see potential conflicts earlier than novice operators. Table 3-11 shows the same breaks grouped according to the position of the break in the scenario: early, normal or late in the situation. The table shows that experienced operators score better in the early breaks, but no differences between experienced and novice operators can be observed in the other breaks. Here too, the effect of the consistency of operators is larger than the effect of the situation (see also Table 3-8).

	Early		Normal		Late	
Rank	**5'**	**14'**	**8'**	**37'**	**20'**	**40'**
1	E1	E3	E3	E1	E1	E3
2	N1	E4	E4	N1	N1	E5
3	N2	E6	N4	N2	N3	N4
4	E2	E5	E5	N3	N2	E4
5	N3	N4	E6	E2	E2	E6
Ave (N)	3.75		3		3	
Ave (E)	2.5		3		3	

Table 3-11 Ranking of operator SATEST scores grouped according to timing of situation

Verbal comments made during the ranking were scored to determine underlying criteria for scoring. From these discussions, the most important aspects in the qualitative evaluation of the traffic scenes can be elicited. The following aspects were considered important for evaluation of a traffic scene (not in order of importance):

1. **Recognition of conflicts.** Indication of all ships involved, speed and course indicated;
2. **Accuracy** with which ships in the conflict are reproduced;
3. Indication of where operators put their **focus of attention** in the scenario;
4. **Overview of the complete traffic area** including ships that are not involved in a conflict situation, for these ships a global indication was sufficient;
5. **Accuracy** with which ships, not involved in conflicts, are placed was considered relevant, but not as relevant as the other criteria.

3.4.2.2.9 Completeness and accuracy of traffic displays

In a first attempt to quantify the results of the SATEST scores, a simple analysis was performed in which for all breaks, the actual number of ships reproduced in the different trials was counted (see Annex 3). This analysis showed that mistakes were made on two sides:

In 37 instances operators did not report all ships present in the situations. In one case an operator missed 6 ships that were in fact in the area. This operator (E6 in break 40) was ranked lowest of all on this break.

In seven instances operators reported ships that were no longer in the situation (These ships had left the traffic area). In one case an operator reported 3 ships that were in fact no longer in the area. This operator (E2 in break 5) was ranked lowest of all operators in that break. In the other instances operators mentioned only one ship that had already left the area. The average number of mistakes for novices (1.44) was slightly higher than that of experienced operators (1.32).

3.4.2.2.10 Attention of operators

Table 3-12 presents how operators divided attention over the traffic. The light grey indications present the ships for which the experts had indicated that they needed increased attention. Ships for which more than half of the operators indicated increased attention are underlined. The numbers in Table 3-12 refer to the number of operators (experienced and inexperienced combined) indicating for each ship that increased attention is needed.

	Minute									
	5	**8**	**10**	**14**	**17**	**20**	**35**	**37**	**40**	**45**
	(N=5)			(N=5)			(N=5)			**(N=15)**
Viroin	<u>4</u>	2	<u>3</u>		<u>3</u>	<u>5</u>	<u>4</u>	<u>3</u>		<u>4</u>
Zp Montelena										
Liguria		1							<u>3</u>	
Subro Valour	<u>5</u>	<u>4</u>	<u>4</u>						1	<u>3</u>
Vinga Polaris				1	2	<u>3</u>		1		1
Benwil				2	1					
Hoo Creek		<u>3</u>	<u>4</u>		1	2				
Caraibe				<u>3</u>	<u>3</u>	<u>3</u>		1		1
Van Uden 22										
Deutschland								1		
Artisgracht					1		<u>3</u>	2	1	
Jan sr.					1		<u>3</u>	<u>3</u>	1	<u>3</u>
Brise					1			2	2	<u>4</u>
Palva					1			2		<u>4</u>
Alice p.g.							1			1
Kraan reiger								1		1
Maria Elisabeth										
Mineral Hoboken										

Table 3-12 Attention of operators in Rotterdam experiment

Table 3-12 presents results from experienced and inexperienced operators in one figure. The table shows that the attention of the operators follows the line of attention indicated by a diagonal pattern. There were not enough data available to present the results from experienced and inexperienced operators separately.

Several things can be concluded from Table 3-12. First the assessors had a lower threshold for indicating increased attention than operators had. Assessors indicated 4.7 ships on average needing increased attention (marked yellow), while 2.4 ships were indicated by more than half of the operators on average (marked with diagonal grid). The attention that needs to be assigned to ships according to the assessors follows the scenario. Throughout the different (SIMPLE, MODERATE, COMPLEX) situations of the scenario, the focus is shifting towards different ships. Only one ship needs attention throughout the total scenario (the Viroin). Assessors and operators agree on the ships that do *not* need attention. In only 3 cases out of the total of 180 possibilities (98%) a majority of operators thought a ship needed attention while the assessors deemed it not necessary. In cases where assessors indicate that attention is required, the majority of operators agree only in 45% of the cases (21 ships out of a total of 47). There is no indication that certain operators are over represented in this.

These results were discussed with the operators after completion of the trials. The operators indicated that attention to ships in a potential conflict situation was mostly given (and scored) when the conflict had not been communicated to the ship. Once this communication had occurred, the operators did not regard the situation as a conflict any more, and turned their attention elsewhere. Unfortunately, there were no data available to analyse these statements in the data set.

3.4.2.3 Discussion and conclusion

The aims of the Rotterdam experiment with SATEST were threefold:

- Develop and validate the data analysis of the method;
- Increase understanding of VTS operator performance; and
- Study the usability of situation awareness in the VTS context

The experiment was set up to provide a baseline measurement of situation awareness in a dense traffic port and approach area. The idea was that differences in experience of VTS operators would lead to improved situation awareness, and that these differences could be assessed with SATEST. From discussions with operators and assessors it was expected that these differences would be visible in the situation awareness of complex situations and that experienced operators would recognise and respond to conflicts earlier than novice operators and trainees would. Three hypotheses were formulated to study the relationship between situation awareness and performance. All three included the comparison between experienced operators and novice operators.

As it turned out, the data set that could be analysed contained only data of four inexperienced operators and eleven experienced operators, instead of the expected 24 data sets. Moreover, the individual differences between operators within the novice and experienced groups were large. As a result no statistically significant conclusions could be drawn over differences between groups of experienced and inexperienced operators. Instead the results of the experiment were used to develop at methods for analysing the SATEST data and discuss these with operators and other experts.

3.4.2.3.1 SATEST data analysis

Both qualitative and quantitative data analyses were used. The results of these analyses are compared to study if the results are internally consistent.

Qualitative analysis

The concept of situation awareness was new in the VTS domain and VTS trainers were at first hesitant to give absolute judgements on operator performance, based on situation awareness scores. Therefore, absolute judgements on the completed SATEST score forms were not used as a method. Instead ranking of the completed score forms across operators was applied as the first method of analysis.

The trainers performing the analyses found it difficult to compare completed score forms from different situations. Apparently the differences in situations can be large, even within the scope of just a few minutes. They were only able to compare within traffic situations.

The results show large differences between operators. Operators rank high or low consistently over the different situations. The differences cannot be explained by experience: among the best scoring operators are both experienced and novice VTS operators (see Tables 3-8, 3-10, and 3-11).

Quantitative analysis

In the quantitative analysis a first attempt has been made to quantify the dimensions put forward as underlying the judgement in the qualitative analysis. The quantitative analysis in this experiment was very simple. The number of ships recollected is just a first indication of how an operator understands the situation. It became clear in the course of the experiment that operators understand the traffic situation in terms of what they consider relevant for their task as traffic controller. Potential conflicts are important, ships leaving the traffic area or the conflict area can be ignored. In the completed SATEST score forms this can be seen in the form of missing ships or other inaccurate recollection of information. With a simple quantitative scoring system, without regard to the relevance for a (potential) conflict of a specific ship, a situation may receive a lower score than justified. The mere counting of the number of ships in a completed score form as applied in this experiment is therefore too simple as a method of analysis. In the next experiment the quantitative analyses were therefore further developed.

3.4.2.3.2 VTS operator performance and situation awareness

Several remarks can be made about the performance of VTS operators in the experiment relating to their performance and situation awareness.

On a number of occasions there were score forms where operators had indicated that they did not need to give attention to ships that according to the assessors needed extra attention because they were on a conflicting course with another ship. In discussions with operators about their results it became clear that several of these operators had recognised the potential conflict and had communicated it with the ships. In communicating the conflict, they shifted the responsibility for the situation to the ships, and did not think that the situation required more attention than monitoring if the ships followed the action they had agreed upon. In their perception this was not giving extra emphasis to the ship, but a part of regular monitoring behaviour.

For the scoring in SATEST, it seemed as if these operators had missed a potential conflict situation, while in fact a closer analysis showed a difference in interpretation of the same perceived situation. This is a complicating factor for the SATEST scoring system: An operator, who anticipates a conflict early and responds to it by communicating to the ships involved, may receive a lower score, while his actual behaviour (early communication) is better.

Discussion with VTS operators about their results and how they achieved them revealed several important issues for the evaluation of SATEST scores. Operators do not regard reproduction of the complete traffic situation as important for evaluating a situation. Getting the most important elements right is considered more important than getting all the elements. A clear difference can also be noticed between individual VTS operators in their strategy of remembering data and positioning the ships. While some operators remember the position of the ships as such, others reconstruct the position of a ship based on information from their communications or the last position of the ship that was observed. This last result opens new fields of understanding of VTS operators' strategies in performing their work and raises questions for further research. Later experiments focused upon these issues.

3.4.2.3.3 SA in a VTS context

At the time of the experiment in Rotterdam, situation awareness was a new concept in the VTS field. It was therefore important to see that the concept and the method for assessing it were easily accepted by the VTS world. Both VTS trainers and operators easily understood the concept and its importance for VTS work. Applying the method in the experiment also led to greater understanding of the VTS work. It started a discussion about issues that were not made so explicit before: particularly the mental picture of the traffic situation. The SATEST method was well received by operators. Especially the simple representation of ships in a traffic situation on paper started discussion on the work and raised interest in operators about their performance. The researchers understood that more work needed to be done in the analysis of the results. In particular the quantitative analysis needed to be more elaborate.

One unforeseen drawback of the method was that, once tested, operators remembered the traffic scenario very well. Even after several years operators recalled to the author how they had missed particular ships in the traffic situation during the experiment. This clearly demonstrated that operators can only be tested once for a particular scenario. That is a factor that makes replication of experiments difficult. Apparently there need to be scenarios developed for each experiment separately.

3.4.3 Helsinki

3.4.3.1 Aims

Within the Comfortable project the Finnish Research Institute VTT developed a path prediction tool for low density traffic situations as part of a support system for VTS operators. One of the display options suggested in order to reach this goal was the short term path prediction tool.

The proposed path prediction tool is a graphical display option that allows the viewer to immediately see the vessel's speed, drift and turning motion. The tool has been successfully used on board ships for tactical navigation. On the bridge of a ship the same data is already visible in numeric form but the information must typically be gathered from different screens and indicators. The interpretation of this information has, up to now, therefore depended on the mariner's skill and experience.

SATEST was used to make measurements to test if the tool developed at VTT helped the VTS operator and improved operator performance in an archipelago area with low density traffic. This second test of the SATEST method was carried out in the summer of 1998 in Helsinki in cooperation with VTT in Espoo. The objectives of this experiment, apart from the test of the VTT path prediction tool, were twofold:

1. To give the VTS operators a chance to discuss the usefulness of the VTT tool with the developers, thus ensuring that in an early stage of development of new tools, the end-users were consulted
2. To test the SATEST method developed at Delft University of Technology and MSR, this time in a VTS area with low-density traffic.

3.4.3.2 Methodology

3.4.3.2.1 Traffic simulator and scenarios

The visual displays presented to the VTS operators resembled as far as possible the set-up in the Helsinki VTS centre. A large picture projected on a computer screen in front of the operators showed an area of approximately 1.5 x 1.5 nautical miles, which covered the major part of the VTS area. A chart of the Helsinki VTS area was used. Critical parts of the area were zoomed in on another computer screen. The operator could choose the area on which he wanted to focus by moving around the zoomed screen, using a mouse. No other equipment was provided to the operator. Both screens were placed in front of the operator.

These displays showed an electronic chart of the selected area together with graphical representations of tracked radar or transponder targets. This set-up is similar to that of the VTS. Because the software interfaces in the VTS and in the simulator environment were different, the degree of user interaction was kept to a minimum. When this was discussed with them, operators considered the interface a good enough representation of their normal equipment to carry out their work normally.

The radar (or transponder) targets shown during the test run were simulated and followed predefined tracks. Actions of the operators with the ships could not lead to changes in course and speed of the ships. In the Helsinki area this was not considered a problem, since operators very seldom instructed ships to change course or speed. Moreover the scenario did not require such actions of the operators.

At the beginning of the experiment the zoom screen was always set to a predefined area at the centre of the VTS area. The simulator was run by another operator, who also took care of the video recording and of problems that arose during the experiment.

For this experiment two different, but equally complex, traffic scenarios were designed and implemented in a simulator environment. Both of the scenarios, A and B, represented realistic traffic situations with regard to the fairway routes in use and to the number of ships, in this case 5 to 6 vessels.

The scenarios were constructed out of a VTT database of ship tracks. In this database ships and their tracks are stored. Ships followed predefined tracks, which could be along the standard paths that ships travel in the area. Alternatively they might deviate from those paths at any moment in time and space. From this database ships were selected and placed in such an order that during the scenario interesting situations developed. The use of this standard database made the traffic realistic.

Prior to the experiments the scenarios were shown to several experts, including representatives of the VTS, to ensure that the traffic was realistic. The traffic scenarios were recorded as AVI files, with one frame every second. This gave the option to replay the scenarios at any speed.

3.4.3.2.2 Hypotheses

The following hypothesis was formulated for the experiment. The predictive display improves the traffic image, by providing tactical traffic information at an earlier stage. This information provides a basis for corrective actions and is therefore important to the VTS operators' understanding of these situations. Situation awareness is a measure of the understanding of a traffic situation. Part of situation awareness of a situation is the distinction between important elements in a situation and elements that are less important now and in the future (Level II & III in Endsley's classification).

Hypothesis 4

SATEST is regarded as a valid instrument for assessing VTS operator situation awareness if it demonstrates differences in situation awareness of important elements in the traffic situation between work environments where the predictive display is used compared to work environments without this tool.

In the experiment SATEST data are supplemented with information from interviews and video analysis.

3.4.3.2.3 Subjects

At the time of the experiment VTS in Finland had been operational for only two years. There were two main VTS stations, one in Helsinki and one in Turku. The VTS organisation was also small. At the time, only six VTS operators worked in the port of Helsinki, and all six of them participated in the experiment as subjects. The operators ranged in age from 35-59 years and had VTS experience ranging from 4 months to 2 years, right from the beginning of the operation of VTS in Finland. Prior to working in Helsinki they had no experience as VTS operator. The operators did have considerable maritime experience as captain or pilot. Five operators were male, one female.

The operators volunteered to participate in the experiment and received no money for their participation. However, they all were glad to participate in an experiment which enabled them to express their opinion on tools being developed for them, the tools at their disposal and other factors involving their work.

Three operators had no experience with the predictive display that was tested in this experiment. The other three operators had the experience of one short (two days) course with the display.

3.4.3.2.3 Data registration

SATEST

The SATEST software was installed on a stand-alone computer linked to a printer, to the side of the VTS terminals.
After every break, the results were printed immediately. These prints were used in the analysis. The SATEST files, that were recorded were only used as back-up in case of failure. During the experiment no failures in the SATEST program occurred.

Video

The experiment was filmed on video. The video shows the operator sitting behind the VTS console and also records all remarks.

Interviews

After the completion of the experiment the operators were interviewed. The interview was recorded on video, for analysis purposes. In the interview the following information was collected:

Operator information
- Name
- Age
- Experience as VTS operator
- Previous experience
- Experience with predictive display

Traffic situations

The AVI files were used in a walk-through talk-through of the experiment to verify that the traffic scenarios used were realistic. The following topics were discussed:
- Incidents and conflicts
- Focus of attention
- Was the traffic realistic?

Simulation

After the completion of the experiment the operators were interviewed. The interview was recorded on video, for analysis purposes. In the interview the following information was collected:
- Was the simulation realistic?
- Recommendations on the simulator

Usefulness of prediction on different levels
- Strategic level
- Tactical level
- Recommendations

SATEST
- Comments
- Recommendations on this type of experiment

General Comments

3.4.3.2.4 Procedure

The operators were tested individually. First the operators were introduced to the objectives of the experiment. After that they were given some time to get familiar with the VTS simulator and with moving around with the mouse. Next the SATEST program was explained. The experiment then started.

The first scenario presented to the operator was stopped after 60 seconds for a test-trial. At this moment the operator was not expected to have full situation awareness of the traffic situation. The objective of this stop was merely to familiarize the operator with the SATEST score form, which the operator had to fill in. This test-trial had proven to be necessary in previous studies, to let the operator practise with the SATEST program, without interfering with the experiment. The results from these test-trials were not analysed later.

After this test-stop the scenario continued. At these predefined stopping moments the simulator was stopped. At the moments the screen blanked and the operator was asked to fill in the SATEST program. The operators were asked to indicate for all ships the following:

- Position
- Course and speed vector of ships
- Predicted future position of the ships in 3 minutes time.

Compared to the study in Rotterdam the information collected in this experiment was different with regard to how *projection of future status* (Level III situation awareness) was assessed. The information collected would demonstrate if operators profited from the predictive display, by being better able to indicate the predicted future position of ships.

The operators were requested to think out loud and to comment on the traffic situation as if they were advising the traffic.

After the first scenario had finished, operators had a second run with the other scenario. Both scenarios were stopped three times to fill in SATEST. After the completion of the second scenario, the operators were interviewed.

The total amount of time for the experiment was approximately one hour and a half, including the interview.

3.4.3.2.5 Stopping moments

During the preparation of the experiments, within each of the scenarios, three stopping moments, called breaks, were chosen. The breaks depicted moments where certain developments in the traffic situations could be noticed with the predictive display, whereas the standard vector without the predictive tool did not display this information yet. This could be information concerning a conflict with other traffic or with the infrastructure. Annex 2 shows a list of all points of interest in the scenarios. The point of interest in the first break of the first scenario was the following: "*The vector display shows Portboard-Portboard passing of HANNA and AILI, whereas the predictive display already shows Starboard-Starboard passing.*"

Figure 3-6 presents the part of the traffic situation in the break showing HANNA and AILI. The left side of the figure shows the situation without the predictive display. The right side of the figure shows the same situation with the predictive display. The predictive display shows with a curved line that AILI is already turning.

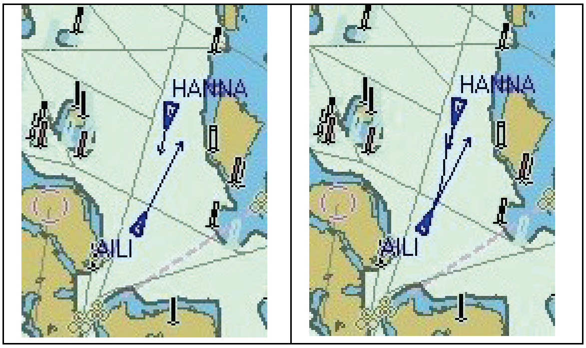


Figure 3-6 Traffic situation in Helsinki scenario, left without and right with predictive display

3.4.3.2.6 Experimental design

Each operator viewed both scenarios, one with radar targets and conventional speed vectors and the other with radar targets and path prediction. The test matrix is shown in Table 3-13.

Operator	Test 1	Test 2
1, 5	Scenario 1 Path Prediction	Scenario 2 Vectors
2, 6	Scenario 2 Path Prediction	Scenario 1 Vectors
3, 7	Scenario 1 Vectors	Scenario 2 Path Prediction
4, 8	Scenario 2 Vectors	Scenario 1 Path Prediction

Table 3-13 Helsinki test matrix

In addition to SATEST, operator actions such as timing and content of VHF communication were recorded. Additional input was collected through user response and feedback.

3.4.3.2.7 Analysis

SATEST

SATEST forms can be analysed in two ways, qualitative analysis by ranking, and quantitative analysis of accuracy. In previous experiments the focus was on qualitative analysis. In the Helsinki project a quantitative approach was applied.

Ranking

In earlier experiments a ranking of the completed score forms was used as preliminary analysis. The experiment leader based this ranking on the interpretation of the scored results. In the current experiment it appeared not to be feasible to give a ranking of the results. The experiment leader was not used to giving a judgement on the performance of VTS operators, and did not feel comfortable doing it. The qualitative analysis was therefore limited to pointing out a number of interesting elements in the scenarios and the score forms. These comments have been incorporated in the quantitative analysis. Therefore only quantitative results are included in this section.

Interpretation of traffic at breaks

In other experiments it was shown that a quantitative analysis of the SATEST score forms should take into account more than a mere description of the inaccuracy of scoring. It should also account for the characteristics of the situation. Therefore for each of the breaks a description was made of the situation and of the points of particular interest in these situations.

In this experiment the basic analysis methods used in Rotterdam were extended. No longer were individual ships the focus of the analysis, but more complex traffic situations were labelled as "points of interest". In these situations ships would meet in potentially critical circumstances, such as close encounters or passing on the "wrong" side. These situations cannot be simply defined by the position of ships, but need the position, speed and heading of both ships relative to each other. The complete set of descriptions can be found in Annex 2. These descriptions were used as reference in the analysis. The completed SATEST forms were analysed to see if the operators were aware of these particular situations, and to analyse how good their situation awareness of these situations was.

Scoring

The SATEST forms were analysed by hand by overlaying them with transparencies with the actual position of ships and scoring the differences using a measuring rod with millimetre scale and a protractor. Three types of transparencies were used for all breaks:

- Correct locations of ships with heading vector
- Correct future positions

- Tracks of ships

The scale of the SATEST form was 1:60,000. Deviations from correct location were therefore given in counts of 60 meters.

- When the ships were located at a deviation of less than 60 meters, the location was called *exactly correct*. In the form this meant that the position indicated overlapped with the correct position.
- When the deviation was between 60-120 m. the location was called *correct*. In the form this was indicated by a partial overlap between the indicated position and the correct one.
- In all other circumstances, the deviation from the correct location is given in a count of 60 meters.

All SATEST forms were analysed, using the transparencies, together with description of the actual situation and the important elements in the situation.

The points of interest in the scenarios were analysed by interpreting the SATEST data, combining position, speed and heading of ships in a situation. For these points of interest, it was not so important that operators located the ships exactly at the right location. The issue here was whether operators put the ships right relative to each other, with regard to position, speed and heading.

Video analysis

The video was used to study how operators watched the screens during the simulation session. It turned out to be very possible to determine to which of the simulator screens the operator was directing his attention. The video was also useful to determine the comments made during the simulation session.

Attention at breaks

Special attention in the analysis was given to the operator's attention during the last ten seconds before each break. Important issues were at which of the screens the operator was looking and if the operator in this period changed the settings of the screen or communicated with one of the ships. This analysis was used to explain certain aspects of the scoring on SATEST.

Interviews

The interviews were analysed from video. Of each of the interviews a report was made containing the personal details and the main comments on both the predictive display and on the SATEST methodology.

3.4.3.3 Results

3.4.3.3.1 SATEST

Distance from correct location

Figure 3-7 shows the distribution of the accuracy of location in the SATEST score forms. The total amount of ships in the scenario for all breaks of all operators together was 37 times 6 = 222 ships.

The scenarios were analysed to determine which ships were important for VTS and which were not. Ships were labelled important if they were in the proximity of the shore, or on a track that would lead them close to shore (the Helsinki traffic area has many small islands along the fairways), or when a ship was on a crossing course with another ship. Ships that were regarded unimportant in the situation (N=98) are marked red in Figure 3-7. In orange the ships are shown that were regarded important in the situation (N=114). The scores are put in percentages to enable comparison of the scores of both types of scores. The correlation between the two is high (0.9783; p<0.001). It is interesting to see that operators in this study do not have better situation awareness of important ships than they have of unimportant ships.

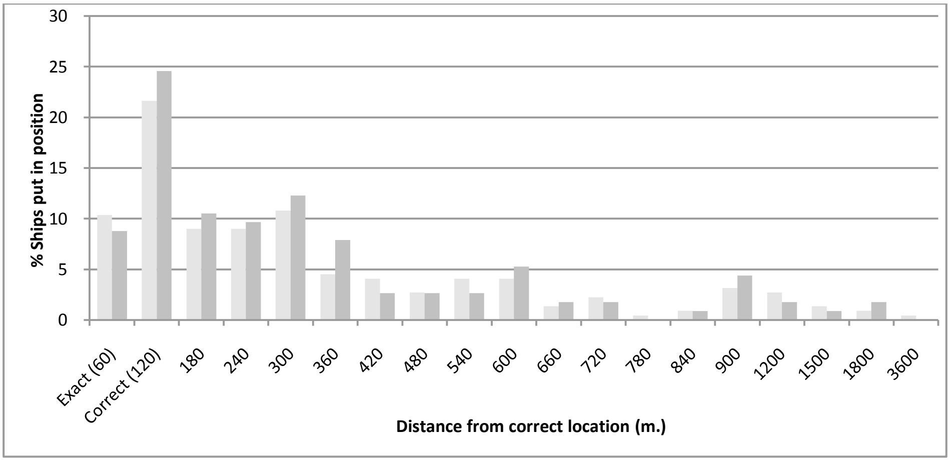


Figure 3-7 Frequency distribution of location accuracy in SATEST score forms for all breaks

Eleven situations in the scenarios were marked as points of interest. In seven of these situations the predictive display was expected to be able to help the operator (see section 3.4.3.1). Figure 3-8 shows the SATEST scores on these situations in the scenarios in the form of percentages scored *exactly correct* or *correct*. With this limited amount of data it cannot be tested statistically, but the results suggest that operators do benefit from the predictive display.

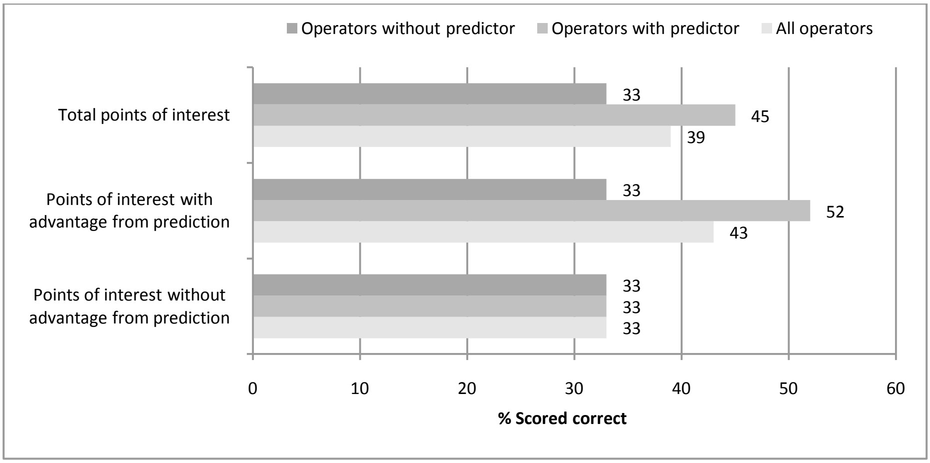


Figure 3-8 SATEST scores on points of interest scored correct with or without predictive display.

This observation was supported by the video analysis. In the ten second period before the breaks, those operators with predictive displays who were actually looking at their displays correctly located 64% of the points of interest in the scenarios that could be observed correctly. Operators who were looking elsewhere in that period, or had their displays set to an area from where they could not see the points mentioned, located only 31% correctly.

3.4.3.3.2 Interviews

All operators were interviewed after finishing the experiment. The interviews took approximately fifteen minutes and were conducted in Finnish. The following topics were discussed in the interviews.

Traffic situations

In general the traffic situations were regarded as quite realistic. One operator remarked that in the scenarios there were more near misses than in reality.

Simulation

In general the simulation was considered adequately realistic. Several operators remarked on the fact that missing radio communication is a handicap, since they did not receive feedback on their communication with the traffic. However, they indicated that it did not affect their behaviour.

Usefulness of path prediction

The operators were of the opinion that the path prediction tool was potentially useful. It enabled them to see how a situation was developing at an early stage, especially at turning points.

SATEST

Operators found it in general an interesting experience to participate in the SATEST experiment. Some operators were already familiar with the method, which had been presented at a COMFORTABLE workshop (see section 3.4.4).

3.4.3.4 Discussion and Conclusions

The goals of this experiment were twofold:

1. To give the VTS operators a chance to discuss the usefulness of this tool with the developers, thus enhancing the consultation with users at an early stage of development of tools.
2. To test the SATEST method developed at Delft University of Technology and MSR in a traffic area with low-density traffic.

SATEST was used to study the effectiveness of a path prediction tool that had been developed as part of a support system for operators. To test this tool, scenarios were developed that provided maximum usage of the tool and SATEST breaks were set at moments within the scenarios where operators could profit from the tool.

Despite the fact that only six operators participated in the experiment, the advantages of the predictive tool could be assessed with SATEST. The effects were even larger when video observations of operators were used to eliminate noise in the data, such as operators who could have used the tool, but were looking elsewhere at times where the predictive display could have offered real assistance to the operators in understanding the traffic situation.

In this experiment the basic analysis methods used in Rotterdam were extended to the analysis of particular "points of interest" in the scenarios. The focus of the analysis changed from individual ships to more complex traffic situations, labelled as "points of interest". In these situations ships would meet in potentially critical circumstances, such as close encounters or passing on the "wrong" side. It was very possible to use the completed SATEST score forms to analyse these situations. Analysis of these situations showed a clear advantage of the predictive display.

The addition of video observation to SATEST proved to be valuable to the method. With these recordings, better understanding could be gained of how the operators built their situation awareness. The SATEST method itself again seemed to be a useful instrument in measuring VTS operator behaviour.

The combination of the different methods used in this experiment, including situation awareness measurements, video observation and interviews gave useful information on the use and acceptance of the path prediction display used in the experiment.

This experiment was the first one using a VTS simulator at VTT. Taking that into account, it must be stated that the simulation was very successful. Operators were able to use the simulator well and accepted the picture shown. Their main complaint about the simulation concerned the fact that there was no communication possible between the operators and the ships. In daily operational work this communication is the main source of information on the traffic. It should therefore be implemented in a future version of the VTT VTS simulator.

3.4.5 Workshop

Between April 1997 to June 1998 the COMFORTABLE consortium organised three workshops. The aim of these Workshops was to bring together the researchers, suppliers, port authorities, managers and operators of VTS to discuss, evaluate and contribute to the future developments under review for VTS operators, with the focus being on the Human/Machine Interface (Regelink and Jarvis, 1999).

The SATEST methodology for assessing VTS operator situation awareness was presented at the second Workshop, in November 1997. At that time the Rotterdam experiment was finished, while the Helsinki experiment had not started. This workshop was the first time that the concept of situation awareness was discussed by a large group of people working in the VTS field, both on operational and managerial level. It was also the first presentation of the SATEST methodology to an audience in the VTS field.

In the workshop participants were shown a Traffic Image, via an LCD projector, and asked to study the image as if a VTS operator. The image was then turned off and the participants were asked to reconstruct the traffic image with as much information as possible on a blank chart. On seeing the traffic image again it was apparent that the non-trained operator had difficulty in recounting the important information with respect to a potentially dangerous situation that was developing within the traffic image. Even the experienced VTS operators had some difficulty and not all of them marked the data, previously shown to them, accurately. The Traffic Image was advanced and the exercise was repeated. This demonstration formed the introduction to the concept of situation awareness and the SATEST methodology to the participants.

In the discussion that followed, it was concluded that situation awareness could provide a tool with which a baseline of performance could be established (Regelink and Jarvis, 1999, p.19). The importance of the tool for the future in order to monitor the capabilities of VTS operators, technically and personally, was also recognised. The workshop concluded that this was an interesting topic giving an average baseline of performance and testing additional / new tools on the performance level.

3.5 Discussion

This chapter has described the development and testing of SATEST, a method for assessing VTS operator situation awareness. The method is based upon the work of Mica Endsley, whose concept of situation awareness and methodology are generally accepted as the standard in the field of situation awareness research. Many researchers have developed methods for the assessment of situation awareness in other domains based upon this work. Since there was no previous experience with the assessment of situation awareness in vessel traffic, it seemed logical to take the work of Endsley as a starting point. From that work the SATEST methodology was developed.

3.5.1 Evaluation of SATEST results

The results of the experiments are mixed. The literature study of Chapter 2 concluded with criteria for selecting a method for VTS operator situation awareness. Section 2.7 provided a number of criteria that formed the basis for the development of the method. In section 3.3 these criteria were further detailed in requirements for the method, which were tested during the try-out described in section 3.4.

Evaluation of the criteria and requirements show that some of the criteria can be discussed briefly, which is done in this section, while others require more discussion and will be dealt with later in this discussion.

The following criteria were mentioned in section 2.7:

1. Direct assessment of situation awareness
2. Method provides as much information on situation awareness as possible
3. On-line methods are preferred
4. Obtrusiveness of method has to be tested
5. Method should be suited for experiments

ad 1. SATEST is a method for direct assessment of situation awareness, in line with the theory and approach of Endsley and her SAGAT methodology.

ad 2. SATEST collects a lot of information on VTS operator situation awareness. The analysis and interpretation of this information is not straightforward and requires considerable understanding of the VTS work. Assessors are very capable of ranking scores of different operators within one break, but comparison of scores between breaks turned out to be impossible. This issue will be addressed in the next section 3.5.2.

ad 3. SATEST is an on-line method.

ad 4. The obtrusiveness of SATEST is subject to discussion. On the one hand Endsley states that the intrusiveness of methods like these is not a problem (Endsley, 2000b, p. 161). In VTS operators are used to interruptions of their works so for them it should not be a problem at all. Most operators did not seem to be disturbed by the testing. This aspect of obtrusiveness therefore caused no discussion. However, there were also operators who, right after returning to their "work" after a break, began to search the screen, to see if they had missed anything in the SATEST task. These operators were clearly doing an experiment, instead of carrying out their normal task. Moreover, asking operators to reproduce all ships in the traffic situation, may have led them to larger awareness of what was going on. In my opinion, the obtrusiveness of SATEST is not undisputed, and should be studied in future experiments.

ad 5. Certain experiments can be performed well using SATEST, while the method is not suited for others types of experiments. This issue will be addressed in section 3.5.3.

Two methods for analysis of SATEST data were tested in the experiments. Ranking of completed score forms provided a qualitative understanding of scoring within breaks. The evaluation of the results of the experiments show that assessors are quite capable of ranking scores of different operators on one break, but comparison of scores between breaks proved to be impossible for them.

For quantitative scoring, the experiments did not yield statistically testable results. This was due to the combination of two factors. There was a lack of complete data sets caused by problems in the experiment in Rotterdam and the unavailability of enough operators in Helsinki. Combined with the issue that the data from different breaks of one operator cannot be combined easily, this has resulted in an insufficient number of data points for statistical analysis.

More important still is the inherent complexity of the VTS operator tasks and the temporal aspect of situation awareness in the VTS work. These issues make quantitative analysis of the completed score forms difficult. To better understand the first issue, a comparison is made between the original method for analysis of situation awareness SAGAT and the SATEST method in the next section. Section 3.5.3 will address the temporal aspect of VTS operator situation awareness.

3.5.2 Comparison between SAGAT and SATEST

The method developed follows the methodology and terminology of Endsley's well-known method SAGAT described in section 2.6.2.1. SATEST uses a VTS simulator to create a task environment. Assessment of situation awareness is done during breaks in the scenario, called "freezes". At these breaks VTS operator must reproduce their picture of the traffic situation at that moment. There are a significant difference between the SAGAT method and the SATEST method for VTS. Following a strict SAGAT approach probing questions should have been asked such as: *Enter speed for ship X,Y,Z*, or *Indicate the ships that currently are on a conflicting course.* Instead in SATEST operators were requested to reproduce[15] a detailed description of the complete traffic. The reason for this approach was the specific characteristics of the VTS work. The dynamics of maritime traffic develop much more slowly than in aviation. A VTS operator builds up his situation awareness of a situation through slowly changing parameters in the traffic. Some of the ships are in a VTS area for several hours at the time. In the scenario used in the experiments some of the ships were present from the beginning to the end, 45 minutes later. Over this whole time situation awareness is built up and enriched, through changes in position, heading, course, and speed of the ship, and communication with the ship. An operator combines all of this information on ships that need to be monitored. Everything any of the ships do will have to be monitored by the VTS operator to decide if the situation changes, and if potential conflicts are resolved or new potential conflicts arise. Communication that takes place in the beginning of the scenario may play a role half an hour later. The situation awareness of the VTS operator does not contain every information item singly. An operator combines all information into one coherent picture. It was therefore considered best in SATEST to test this complete, integrated picture and not the individual information items separately.

From a testing perspective this offers a challenge for the scoring. The score forms cannot be interpreted in a straightforward way by counting the number of ships that have been reproduced. Both completeness and accuracy of the forms have to be taken into account. It is not arbitrary how these two are scored. The question for discussion is how complete the score forms need to be and how accurate the information needs to be. The data show that there needs to be some sort of weight for different parts of the information reproduced by the operator to derive a score. Endsley does it differently. Scoring in SAGAT does not use weighting factors to indicate importance of certain elements in the situation. The rationale for this is that forgetting or losing track of any element in a situation may cause disaster (Endsley, P.C. 1997).

This was arguably so for Endsley's subjects, fighter pilots. However, this argument does not hold in the context of VTS: Ships that are moored need to be watched occasionally, but are definitely less important than ships on a conflicting course. Therefore even among elements (in the case of maritime traffic ships) there is a hierarchy of importance. This holds even more for different types of elements. Speed and course of a ship are more important than a name, origin or destination of a ship.

However, the experiments have also shown that this information is very context dependent. Information that is vital at one moment may be completely obsolete at another

[15] *Reproduction* in this sense refers to an immediate recollection of what an operator remembers of the situation. *Reconstruction* refers to the process of using the reproduced information to work out a more complete, detailed, or more correct picture. The distinction between the two is not absolute.
During the breaks in SAGAT limited information is gathered. In SATEST operators reproduce everything in the situation they remember. On the one hand this is an advantage, since richer information is collected, that can be used to perform more complex analyses, but on the other hand this has led to reconstruction of traffic instead of reproducing situation awareness

moment during the same scenario. This makes it difficult to use scores of operators over different breaks to understand how situation awareness develops in a scenario. It requires careful analysis of different breaks, to define for each break what elements are important. The experiment in Helsinki has demonstrated that this is feasible, but it is a labour intensive process. This temporal dimension will be addressed specifically in the next section.

3.5.3 SATEST and temporal situation awareness

The concept of situation awareness in the work of Endsley is very much focussed on the present. The present situations in the "*freeze*" moments and cues in that present are used to determine situation awareness in SAGAT. Thus a prediction of the future (level III situation awareness) is determined by cues from the present.

In maritime traffic cues are not always so clear. A situation with ships approaching each other gradually develops into a potential conflict. VTS operators differ in when they regard a situation as a potential conflict. They also differ in when they want to deal with such a conflict. Some operators let a situation develop for a while, where others may respond early. Although there seems to be agreement among experienced operators and trainers about the latest acceptable moment at which communication should take place, there is no agreement about the best moment for communication about conflicts. In the Rotterdam experiment this aspect was dealt with by testing cross-sections several minutes apart, prior to a particular potential conflict. It was not possible to test the same operator at these short intervals, since he would not have had time enough to get back into the scenario if the time slices were too small. Therefore no assessment could be made of the development of situation awareness of a particular operator. The data collected reflected situation awareness in general at these breaks, averaged over the group of operators. However, in the analysis it was hard to find the conflict back in the results of the SATEST score forms. The reason for this was that operators stopped giving the potential conflict extra attention when they had communicated about it with the ships involved. After that the ships on potentially conflicting course were just monitored like any other ships. Therefore they no longer received extra attention. This issue was not accounted for by assessors when the scoring for this scenario was developed, but became clear only in the analysis after the experiments had ended.

For a better judgement of VTS operator situation awareness it seems necessary to monitor communication better. The next chapter will elaborate on this issue.

3.5.4 What can be tested with SATEST

Considering the issues discussed in the previous two sections the question arises whether the SATEST method provides a suitable, or indeed the most suitable approach in assessing VTS operator situation awareness.

The approach to assessing situation awareness developed in SATEST is well suited for use during training of VTS operators. Situation awareness is important in the work of the VTS operator and is a prerequisite for good performance. Testing of situation awareness in candidates and training of situation awareness therefore is useful. The completed SATEST score forms can be discussed with the operators after the training run, to explain to them what was good, and what not, thus increasing understanding of how well operators are maintaining an overview of the situation. Interesting in this respect is that after I presented the work of the Rotterdam experiment at a VTS conference (Singapore, 2000) several people approached me and told me that they wanted to use the method in their simulator-based training. They were not planning to use the whole SATEST method, but just wanted to use a paper outline of the traffic area they were training in. On this paper trainees could fill in the traffic. With this simple tool, they could integrate situation awareness assessment in their training. However, since then the method has not become a part of the official VTS training methods. Section 5.3.3 will discuss how this may be achieved in the future.

When SATEST is applied like this, the complex matter of comparing scores of different operators on the same trial, with an objective interpretation of the results can be disregarded, since the results of the trial are used in a direct one-on-one discussion with the candidate/operator.

As a method for testing of new VTS systems, the starting question for this chapter, the appropriateness of the SATEST testing method is not so clear. Several problems remain:

- The temporal nature of the VTS operator work is not well covered with a method which applies breaks to test situation awareness at particular moments during the scenario, but does not address the ongoing process of constantly changing situation awareness.
- To get statistically significant results for proof of differences between systems, larger numbers of subjects are necessary, but they are not generally available.
- Scenarios can be used only once, because they are remembered very well.
- Scoring procedure: Each break of each scenario has to be studied to determine the relevant elements; there is not yet a general procedure. Therefore the scoring process is tedious
- The absence of an external criterion for good VTS operator performance makes it difficult to interpret the test results.
- Comparison of systems is only possible if the researcher can interpret the different results attained between the systems and the relevance of these differences for performance. Right now there is not enough understanding of the process of VTS to generalize differences in experimental results to general valid statements.

3.6 Conclusion

The objective of this study was to increase the understanding of VTS operator situation awareness and its relation to VTS operator performance by developing a method for the direct assessment of situation awareness of VTS operators.

This chapter has shown that situation awareness is indeed a useful concept in describing the work of the VTS operator. The SATEST method presented in this chapter needs more work. In subsequent research there needs to be more attention for the role of communication in the work of the VTS operator. Communication may be used to analyse when operators become aware of conflicts and when they will solve them.

The next chapter describes the development of a method for assessing VTS operator situation awareness based upon analysis of communication.

CHAPTER 4

Performance measuring instrument PMI

4.1 New approach towards assessing VTS situation awareness

The work on SATEST was very useful in providing understanding of the VTS-work, the concept of situation awareness and its application in a VTS context, and also of the limitations of the method.

One of the things learned about maritime traffic management was that it strongly relies on the ability of the operator to evaluate situations as they develop. Early prediction of conflict situations as they emerge enables the operator to optimise his workload over time. Early detection of conflicts and early decisions on how to handle them are considered to be valuable strategies in preventing accidents or delay later on.

It became clear in working with SATEST that the method covers situation awareness at any particular moment very well, but it is not very suitable to observe how situation awareness develops in a situation (Wiersma, 2005).
Because the SATEST method applies freezing the situation at predefined points in time during the scenario, it is difficult to perceive at what moment the operator starts considering a particular traffic situation as a conflict. To answer questions of this type would require many tests, with multiple time slices. This is an inherent limitation of a method that uses time slices to assess an ongoing process, such as situation awareness.

Therefore a new approach to testing VTS operator situation awareness was developed, applying a performance-based method. The main objectives for developing a second method for the assessment of VTS operator situation awareness were to:

- Increase understanding of the relation between situation awareness and performance in a VTS context by using assessment of observable behaviour;
- Overcome some of the limitations of SATEST by addressing the time dimension of VTS operator situation awareness.

In section 2.6.4.3 the strong and weak points of an assessment method using sub-task performance were described. Such a method is considered useful if a direct link can be established between the sub-task and situation awareness. In Table 2.2 the sub-task performance method scores well on the amount of information that can be obtained, even though it does not assess situation awareness directly. The method also scores well on experimental conditions, such as use on-line and unobtrusiveness, but is not considered simple.

The method developed is based upon an ongoing assessment of the situation and upon observable behaviour. For the assessment of the situation a new concept will be introduced: Required Situation Awareness (RSA). RSA refers to the situation awareness needed to correctly handle a situation. The concept is described in section 4.1.2 The observable behaviour is assessed by assessing communication, which is used as a measure of situation awareness. Situation awareness is inferred from this observed communication.

The method that was developed as a result of this, the Performance Measuring Instrument (PMI) was again developed in close cooperation with MSR (MSR, 2001, *in Dutch*). It was used in a research project for the Rotterdam Port Authority that investigated the relation between age and VTS operator performance. This project offered a good opportunity for validation of the method, since all Rotterdam VTS operators participated in the project.

In the research project conducted in Rotterdam the focus changed from assessing situation awareness towards assessing VTS operator performance (Wiersma, Butter, & van 't Padje, 2000).. After the completion of the study the results were used to further investigate the relationship between communication performance and situation awareness. To perform well operators do not only need to communicate the right issues to the ships, but they also need to do that at the right moment. Two studies were conducted to investigate this topic using the data from the experiment in Rotterdam. The first study addresses the relationship between communication and situation awareness, and the second study investigates the influence of timing of communication. The results of these studies were used to re-evaluate the assessment method as a tool for understanding situation awareness.

This chapter first describes the new approach towards assessing situation awareness that was chosen in section 4.1, and the method that resulted from this approach in section 4.2. It next describes in section 4.3 how this method was used in the study carried out in Rotterdam, and the validation of the method that resulted from this study in section 4.4 to 4.6. Section 4.7 describes how a selection was made of effective and ineffective operators. This selection is used in section 4.8 that provides a re-evaluation of the method as a tool for understanding situation awareness.

4.1.1 Proactive monitoring versus reactive monitoring

To perform his task the operator continuously monitors traffic, scanning the water and his screens for events in the expected traffic flow. Two operational modes can be distinguished, a reactive mode and a proactive mode. The choice for a certain type of traffic control seems to be more dependent on regional differences than defined by characteristics of the traffic or geographical conditions.

- In the **reactive mode** the operator waits for traffic to contact him. He does not take the initiative to contact ships, and supplies only the information that is requested by the ships. In some traffic areas the reactive mode is the only mode accepted, and active traffic control is regarded as a nuisance and an infringement on the authority of the ship's master or pilot. This type of traffic control has been observed in Scandinavian VTS posts, such as Helsinki VTS;
- In the **proactive** mode the VTS operator actively searches for emerging conflicts and deviations from "normal" traffic and informs ships as soon as he sees anything of importance. This type of traffic control is characteristic for VTS in Rotterdam.

Although each single event can be handled in a reactive manner, being proactive is considered the better strategy by experienced operators and trainers, including those in VTS areas where the reactive mode is the dominant behaviour, since it allows the operator to optimise his attention and to prevent being overloaded when the task load increases. Besides, to an operator who acts proactively, signals coming from the ships can be regarded as a backup system: If the operator has missed a certain event, he is alerted to it in time to still take appropriate action. On the other hand, an operator who only reacts to signals may miss a certain event, only because the ship does not report it to him. There is then one less barrier protecting against mishap.

4.1.2 Required Situation Awareness

To be able to carry out his task correctly, the VTS operator needs a certain amount of situation awareness. At this moment it is time to introduce a concept that was first presented at a HPSAA conference (Wiersma, 2000), which is called Required Situation Awareness (RSA).

The concept of situation awareness is a *psychological construct* that describes something inside the head of a person, in this case the mental model of a situation. This mental model is supposed to have a relation to the outside world, but strictly speaking there does not have to be such a relation. One can have situation awareness of something that is not really there. It is the old Cartesian dilemma of the relationship between what is inside a human mind and the outside world. Situation awareness is a description of the mental part of the equation.

To make situation awareness useful as a concept for training and research it is necessary to connect it with its counterpart in the *outside world*. In the actual situation there has to be something the person needs to be aware of. It is an inherent property of the situation itself that demands knowledge, skills, and awareness from the person in order to deal successfully with that situation. This property is what this thesis calls the Required Situation Awareness (RSA). Figure 4-1 visualises the RSA concept.

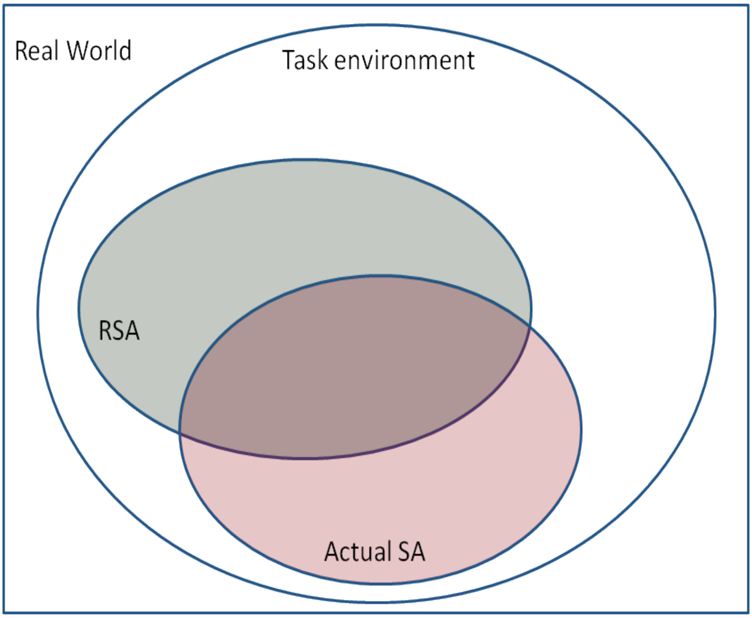


Figure 4-1. Venn diagram showing RSA and actual situation awareness in the task environment

As shown in Figure 4-1, RSA is part of the total actual situation awareness a person can have of the task environment: it is that part of situation awareness that is necessary to perform the task at any particular moment in time during the scenario. What is *Required* in a situation is a complex issue which has relations to the demands on attentional resources used in SART, mentioned in section 2.6.2.1.2, and will be later defined using expert judgement.

In optimal circumstances, actual situation awareness and RSA should be equal. In most situations the two will overlap but not be identical. Some parts of RSA may be missing in actual situation awareness to the detriment of performance, whilst at the same time situation awareness may contain elements that are not needed at that moment.

Any method measuring situation awareness will also not cover exactly either situation awareness or RSA. Figure 4-2 shows seven different areas of interest that can be distinguished: the first four of them being assessed by the situation awareness measuring instrument X, the other three not:

1. The part of RSA that is in actual situation awareness of an operator that is assessed by instrument X (RSA+; SA+; X+);
2. The part of actual situation awareness of an operator that is assessed by instrument X, but that is not in RSA (RSA-; SA+; X+);
3. The part of RSA that is assessed by instrument X, but that is not in actual situation awareness of an operator (RSA+; SA-; X+);
4. The part of the task environment that is tested by the instrument, but is neither RSA nor actual situation awareness (RSA-; SA-; X+);
5. The part of RSA that is in actual situation awareness of an operator, but that is not tested (RSA+; SA+; X-);
6. The part of RSA that is not in actual situation awareness of an operator, and that is not tested (RSA+; SA-; X-);
7. The part of actual situation awareness of an operator that is not in RSA, and that is not tested (RSA-; SA+; X-).

An assessment methodology should maximise the areas (1-3), and minimize the areas (4-7). The discrepancy between what needs to be tested (numbers 1-3 & 5-7) and what is actually tested (number 1-4) is usually regarded in validity studies of methodologies, although usually without regard for the difference within this three areas. The more complex relations between RSA, situation awareness and the assessment methodology are not addressed.

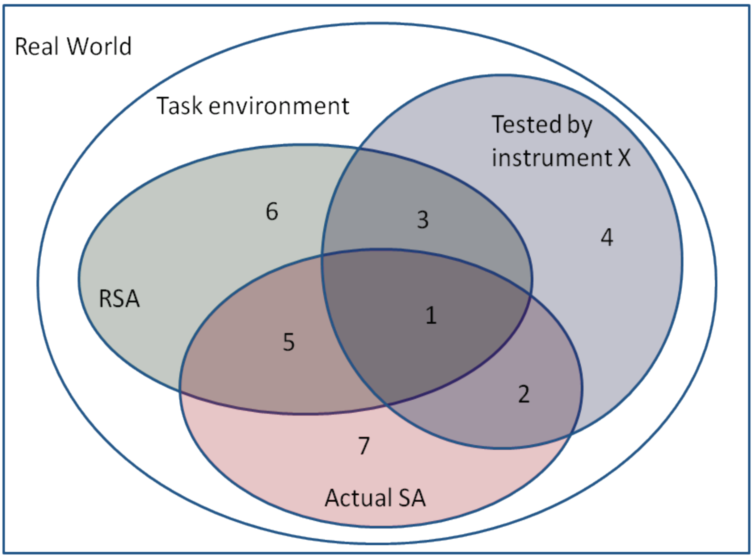


Figure 4-2 Venn diagram showing RSA, actual SA and tested SA in task environment

Methods for assessing situation awareness can take two approaches towards assessing situation awareness: A normative assessment approach considers situation awareness from the perspective of RSA, and puts the emphasis on the areas 1 and 3. A naturalistic approach takes actual situation awareness as starting point and focuses on areas 1 and 2. The approach taken in this chapter combines the two approaches.

Even though methods for assessing situation awareness use the idea explained here (information necessary to perform the task is used to establish a situation awareness score), the author has not yet found a method for assessing situation awareness that explicitly addresses the issue of RSA. This may be partly due to Endsley's opinion, described earlier (see section 3.5.2) that all information in the situation is equally important for the task of pilot and that a pilot may miss none of the cues presented (an assumption which may be true for fighter pilots in military aviation, but does not hold for maritime traffic control, given the slower dynamics and the lack of deliberate confusion attempts by the enemy). Therefore, according to Endsley, weighing of different elements in a situation is not necessary, because all information in the task environment is what is called here RSA. This simplification makes work easier for the researcher, but does decrease the validity of assessment methods.

The advantage of using the concept of RSA is that it allows analysis of situations, without actually assessing situation awareness. This situation-driven RSA depends on the amount of information and complexity of information that needs to be monitored, processed, and acted upon. It thus provides an external measurement. In the absence of a valid objective measure, the results of inter-subjective expert judgement are as close to an objective measurement as can currently be acquired. Exactly this approach was chosen in the development of the method PMI-P, described in this chapter.

4.2 Development of PMI-P

The idea of RSA has been used in the development of the second method for assessing situation awareness, PMI-P. PMI-P is one module from a larger instrument for monitoring VTS operators. This method, called *PMI* (Performance Measuring Instrument), was developed by Delft University of Technology and MSR in order to get more insight into the practical job performance of Vessel Traffic Service (VTS) Operators (Butter, & Wiersma, 2002; Wiersma & Butter, 2002). The method *PMI* consists of three modules. First, a performance module based on a standardised simulation scenario with a detailed assessment protocol breaking down VTS performance into the various types of communication between VTS operator and the vessel traffic (PMI-P). Second, a well-being module that consists of questionnaires focusing on factors that influence performance, such as sleeping problems and recovery need. Third, a medical module aimed at eyesight and hearing.

This thesis mainly addresses the first module, PMI-P. The second and third modules of the method were developed and applied separately from the performance module, and will not be discussed extensively here. The complete *PMI* method has been described in a report for the Rotterdam Port Authorities (MSR, 2001).

PMI-P used the same assessment environment as SATEST, the VTS simulator that MSR used for training purposes. By means of this simulator, the operator was confronted with a scenario that invited him to execute his job as a VTS operator just as in daily practice.

The main difference with SATEST was that in PMI-P an experienced observer rated performance in an ongoing process, instead of assessing a direct measure of situation awareness during breaks, as applied in SATEST.

4.2.1 Rating communication performance

The RSA concept introduced in the previous section is a theoretical concept. It is an inherent property of the situation that cannot be translated directly into a scoring system, but has to be translated to something that can be assessed. PMI-P assesses the communication performance of operators as the measure of RSA. The measure was made operational as "*The information that has to be given to the ships at the right moment in time during the scenario in order to achieve safe and efficient shipping*".

In a series of four sessions two VTS trainers decomposed events in the chosen test scenario and arranged them along a time scale. The sessions used both on-line test runs, where one of the trainers performed the task of the VTS operator while the other noted the communication, and off-line test runs where the scenario was stopped at critical points and adequate communication in those situations was discussed. In this way the PMI-P score form was developed, using expert judgement for elicitation. The completed score form was subsequently tested by having three experienced operators run the scenario and scoring their communication. Their runs were scored by two VTS trainers, to test whether they were internally consistent in their scoring of the communication. After each of the runs the results were discussed with the operators to determine whether their communication had been appropriate and whether it could be scored on the score form. The final result was a score form that could be scored by one person with good understanding of the traffic development in the scenario, but who did not need to be a VTS trainer. In the research project in Rotterdam two assessors were used, since it was the first time the method was used and at that time the internal consistency of the scoring system was studied as part of the project.

For each event in the scenario an observable action (some kind of communication) was defined. Observable actions included calling a ship; giving necessary information (for instance fog warning); giving additional information (for instance information that is not critical at this moment but that may become useful later on); and active monitoring of deviations (calling a ship's attention to it). The following categories of calls were distinguished:

Code:	*Type:*
A1	Reactive response to a call from a ship with important, necessary information. Information which has to be presented;
A2	Reactive response to a call from a ship with additional information used to clarify the traffic situation;
S	Reactive response to a call from a ship, acknowledgement without providing information (Standard procedure call);
W	Proactive call to a ship based upon own observation of traffic, without request from ship.

In terms of situation awareness levels, the categories **A1**, **A2**, need at least level II situation awareness (comprehension of current situation), while for the **S** type of communication level I situation awareness (perception of elements) is enough. Communication of the **W** type needs level III situation awareness (prediction of future). The relation between the PMI-P scores and situation awareness will be discussed in more detail in section 4.8. In terms of RSA, the most interesting items are type **A1** and **W** communication, since they contain information that an operator needs to be aware of to perform his task well. The other categories merely provide additional information. It can even be argued that the categories **A2 and S** are not part of RSA at all, but only present additional information, that may even be distracting. However, the distinction has not been made this strictly, since the relevance of information is not always easy to determine. Section 4.8 will address this issue in more detail.

The score form was set up in such a way that it allowed for individual differences in communications strategies with regard to the exact timing or content of a particular message to the traffic. Certain items in the score form were duplicated in several places, when the situation allowed that communication could take place earlier or later, or when communication could be more extensive (in the form of an **A1-** or **A2-**type communication), or minimal (in the form of an **S-**type communication). Observed actions were scored with an indication of whether an operator performed the action earlier or later than was expected in the scenario, according to the normative protocol that resulted from the expert sessions. The relation between the timing of PMI-P scores and situation awareness will be discussed in more detail in section 4.6.4.2.

Code:	*Type:*
<no code>	In time, according to the score form before the next time slot;
E	**Early**: Information presented more than one minute earlier than time set in score form, through anticipation of traffic

development;

L **Late**: Late messages. Information presented more than one minute later than time set in score form.

By focusing on observable events, PMI-P tested the RSA that was defined for this particular scenario. Inherent in understanding the situation is the awareness of a number of possible outcomes of a situation. The operators' actions do not just follow these outcomes; they set them in motion and can change them if necessary. In their communication with ships, operators talk about these outcomes (for instance who goes first and how do ships pass each other). They do not tell ships exactly what manoeuvres they want (change speed or course). If an agreement has been made between ships and VTS who goes first, the operator monitors that agreement. Only in rare circumstances, for instance a pending accident if no immediate action is taken, operators directly tell ships what manoeuvre to perform.

In the scenario used in the study, all operators were tested on exactly the same scenario. That implies that all communication was foreseen and that there was no need or possibility for the operators to orders ships on a course or speed that was not anticipated developing the scenario. To achieve this in a scenario where an accident occurred required special care. The ship that caused the accident was not responding to communication with the operator in the period leading to the accident and performed a strange manoeuvre that actually led to the accident. The operator would try several times to have communication with the ship and inform other traffic about the lack of communication he encountered. The attempts of the operator to prevent the accident would be in vain due to the impossibility to communicate with the ship. The credibility of the scenario was discussed in the try-outs and considered satisfactory.

4.2.2 Post-hoc rating scale of behavioural dimensions

The performance measures aimed to assess situation awareness were combined with a post-hoc “subjective assessment” of the VTS operator performance by the two assessors, VTS trainers. For the use of the method one assessor would have been enough, but to study internal consistency of the scoring, two assessors were used in this study. The subjective measurement was assessed using behaviourally anchored rating scales. The dimensions which were measured on a four-point Likert scale are presented in Table 4-1, with a score of 1 being good and 4 poor:

This rating was assessed after the scenario had been completed and the PMI-P score forms had been filled in. This type of scoring on behaviour scales is commonly performed in assessing VTS operator performance. It is the type of assessment that is used in normal training and examination of VTS operators (NNVO, 2009).

	Dimension	Description
1.	Finding solutions, improvising	Comes up with adequate solutions to developing problems
2.	Global overview	Guides individual ships explicitly taking into account other traffic
3.	Setting priorities	Addresses complex situations by solving most important problems first
4.	Use of voice	Uses voice clearly and convincingly
5.	Flexibility	Deals adequately with sudden changes in the environment
6.	Coping with stress	Acts calmly and effectively in complex situations
7.	Decisiveness	Acts quickly in complex situations
8.	Motivation	Shows willingness to perform well during the test
9.	Global effectiveness	Keeps throughout the simulation a good balance between safety and efficiency

Table 4-1 Behavioural dimensions in PMI

Assessors indicated that scoring the PMI-P form during the scenario made them more aware of the communication that actually took place, which enabled the m to provide better scores on the behavioural scales. Therefore the assessment was not completely independent. However, the PMI-P score form provides a minute-by-minute review of communication performance and total scores were not calculated immediate, while the scoring on the behavioural scales was a typical overall assessment of performance.
The scores derived from this assessment are used in section 4.7 to create two sets of operators: one set of effective operators and one set of ineffective operators. These sets will be used to study how the more implicit judgements from these global dimensions relate to detailed scored behaviour on the PMI-P score forms.

4.3 Application of PMI-P

PMI was developed and validated in a research project carried out for the Rotterdam Port Authority. The objectives of the project were to develop a monitoring instrument for VTS operator performance, wellbeing and health and to establish a standard assessment and a benchmark of VTS operator performance that could be used later as a reference. In this extensive research project all 110 VTS operators employed by the Rotterdam Port Authorities participated (MSR, 2001).

4.3.1 Aims

The aim of the project was the development of a validated instrument for monitoring VTS operator performance. The project did not study situation awareness in-depth. That part of the study was done afterwards, when the FLO project had already finished, and will be reported in section 4.8. This section describes the data collection in the FLO experiment and the assessment of the reliability of the data. The other parts of the FLO research are not discussed in this thesis.

4.3.2 Methodology

4.3.2.1 Traffic simulator and scenario

This project used the same VTS simulator of MSR in Rotterdam as was used in the previous experiments with SATEST in Rotterdam. A description of this simulator has been presented in Chapter 3, that describes the work with SATEST.

The traffic scenario used in this experiment was different from the previous experiment. It was set to take place in the same fictitious area used in previous experiments. The scenario was situated in a sector that had to be surveyed with a split screen using two monitors. A more detailed description of the scenario and the PMI-P score form with the responses can be found in Annex 4.

The traffic build-up and density were comparable to the Rotterdam situation. The scenario contained 25 vessels, took 30 minutes and developed in 6 phases, lasting 3 to 8.5 minutes. Each phase was characterised by a number of events which required a response from the operator.

In one of the phases of the scenario used in this experiment an accident took place. The accident could not be avoided by any action taken by the operator, because the ship that caused the accident did not respond to calls from the operator, but the operator could react to it. While the accident took place and drew the attention of the operator, other traffic in the area continued, forcing the operator to divide his attention between the extraordinary event and the normal traffic. Most of the scenario took place in fog (visibility dropping to 100 m).

4.3.2.2 Subjects

The entire population the Rotterdam Port Authority (RPA) participated in the experiment. In total PMI was administered to 110 VTS operators. This high extent of participation was not completely out of free will. The operators were sent by their employer to take part in the experiment, during working hours. The outcome of the research could be directly relevant to the VTS operators, because the results of the study were to be used in a discussion about changes in the RPA policy on early retirement schemes (in Dutch: Functioneel Leeftijds Ontslag, or FLO). For that reason there was a certain concern that some operators might want to influence the outcome of the research. They could do that for instance by underperforming in the test. The assessors were aware of the (possible) problem and indicated if they thought an operator was not performing his best. This happened only on one occasion. The results of this subject were eliminated from the dataset.

4.3.2.3 Data registration

4.3.2.3.1 Score forms

PMI-P data were collected using two types of score forms. During the trial detailed observation forms were used to record the communication performance of the operator (see annex 4). These registration forms were developed especially for this scenario. The score form lists the phases of the scenario and indicates all events in the scenario with their timing. For each event all possible relevant verbal responses of the operator are indicated with their appropriate timing in the development of the scenario. On these forms the assessors could indicate which specific responses the tested operator gave to the stimuli in the scenario. In other words, the assessors registered which information the operator gave to the traffic. The aim of the score form was to be as complete as possible, making the scoring procedure easy for the assessors. This meant that an operator did not have to give all responses mentioned in the form to obtain a good score. When an operator gave a response that was not anticipated, it was noted on the form. These responses were evaluated separately, to see if any important remark had been missed in the score form.
The possible responses were coded with the corresponding types of information, according to the categories set out in section 4.2.1.

Since this was the first use of such a PMI-P test, there was not yet a standard method for developing this score form. The score form used in this experiment was in fact the first one developed ever. The development of the scenario raised questions about the level of detail necessary for scoring and about the (normative) categories of types of responses distinguished.

A second score form was used by the assessors after the test to make a "subjective assessment" of the VTS operator performance, using behaviourally anchored rating scales (see Annex 4). The dimensions are described in section4.2.2. This second form was filled in by the assessors after the test run had been completed. The method of assessing VTS operator performance with such scales is closer to the competence based assessment of VTS operators that the assessors were used to in their normal work as VTS trainers described in the IALA Guidelines on the assessment of training requirements for VTS personnel (IALA, 2001) and the IALA Recommendation V-103 on standards for training and certification of VTS personnel (IALA, 1998).

4.3.2.3.2 Assessors

In total three assessors participated in the experiment. The assessors were experienced VTS trainers with 10-25 years of experience as VTS operator and more than five years of experience as VTS trainer. These trainers were the same ones that had also been involved in developing the PMI-P score form. Therefore they had a very clear understanding of the scenario and of the communication they expected to take place at any moment in the scenario. They also had some experience scoring communication on PMI-P forms, which they had gained during the development and testing of the score form, as described in section 4.2.1. Using two assessors was not necessary to obtain the scores on either of the tests, but was used to enable a study into the inter-rater reliability on the scores on PMI-P and the rating scales.

Out of these three assessors two were selected randomly in any trial. These two assessors scored the performance of the operators independently. One of them played an active role in the experiment. He functioned as test leader who operated the simulator and functioned as "ghost" (playing the role of other ships, communicating with the operator etc.). The other assessor was just monitoring the experiment and filling in the score form. At the end of each trial the two assessors compared their completed score forms and discussed any large discrepancies between the two. These discrepancies could be caused by different perception of what had happened, or by a different interpretation of communication. The differences were used in the inter-rater reliability study.

4.3.2.4 Procedure

The experiment was embedded in the total PMI research which cost an operator half a day. During this time all three different tests, for health, well-being and performance, were administered.

For the performance part (PMI-P), the operators were first briefed about the general background of the research. They were told to act in the simulation run in the same way they would in daily operation. The operators did not have to do anything special in the test, except perform as normally as possible. The operators knew that there performance was assessed, but could not actually see the scoring, since this was performed by assessors sitting out of sight of the operators.

4.4 Data clean up

Data from the experiment were cleaned-up by removing unclear data from the dataset. In the simulation run the assessors had two tasks. The first task was to run the scenario and to communicate to the operator the calls coming from the ships. The scenario was written out with a time line indicating when certain issues needed to be communicated. The second task was to score the communication from the operator on the PMI-P score form (see annex 4). The two forms were combined into one that contained both the remarks and communication needed to present the scenario and the coded categories of the PMI-P score form. This facilitated the scoring of items along the time line of the scenario. As a result several lines of the form contained both information on the scenario and coded response categories. Moreover the form contained empty lines that were put in to facilitate reading by the assessors. A detailed look at the completed PMI-P scores form revealed that assessors had not only put their scores at the designated areas, but also at other locations on the form. Not all scores used the planned coding (A1, A2, W, S). Table 4-2 shows extra types of scores which received at least one tick by one assessor. Thirteen of these types could be distinguished. They were recoded after discussion with the assessors according to the categories in Table 4-2.

PMI-P code	Explanation	Recoded to
A1	(Tel) referred to a telephone call from	**A1**
A1(Tel)	another VTS post. The information was given not by radio but by telephone	**A1**
A2	(Tel) referred to a telephone call from	**A2**
A2 (Tel)	another VTS post. The information was given not by radio but by telephone	**A2**
S		**S**
W	(Visibility) referred to a warning of	**W**
W (Visibility)	reduced visibility	**W**
(Visibility)	(Time) referred to the time in the	**W**
W (Time)	scenario when a certain issue was	**W**
(Time) (Visibility)	communicated	**W**
(Time)	(Time) referred to the time in the scenario when a certain issue was communicated	**T**
<empty>	(Empty) referred to lines left blank intentionally to facilitate reading	**X**

Table 4-2 PMI-P score form recoding

While most items in the PMI-P score form could be recoded to the four formal categories **A1**, **A2**, **W**, and **S**, as presented in Table 4-2, several items remained that did not fit the scoring system. First, as a result of combining the scenario description and the score form, there were sections on the form that contained a description of the event to which the operator has to respond with an indication of the time. These events were recoded **T** for Time. 44 Items on the score form were labelled **T**. Second there were empty lines in the score form, as a visual support to the assessors. Sometimes assessors used these blank lines to indicate some sort of communication by the operator, without making clear exactly what had been communicated. These empty lines have been recoded **X**. 74 items on the score form were labelled **X**. Table 4-3 presents the scores for all 110 operators averaged over the two assessors.

The scores on **T** were low (average score per **T** item = 0.7). The scores on **X** were low as well (average score per **X** item = 3.6), but higher than those on **T** items. Closer inspection of the **X** items revealed that three of the lines labelled **X** on the score form out were responsible for 74% of all the scores. Lines 252, 259 and 273 on the PMI-P score form in Annex 4 scored 58, 82 and 61 times respectively (total score 201), which is almost 4% of all scores.

Item	Number of items	Total score	Average score	% of total scores
A1	105	3946	37.6	71.3
A2	30	520	17.3	9.4
S	7	188	26.9	3.4
W	37	581	15.7	10.5
T	44	30	0.7	0.5
X	74	270	3.6	4.9
All items	297	5535	18.6	

Table 4-3 Total and average scores per item for all operators (N=110) for different types of items

The scoring of **T** and **X** items can be explained by two reasons. First assessors sometimes scored communication on an event rather than individual communication items within an event, as the PMI-P score form prescribed. Assessors used the empty lines or the time line to score such communication. The reason for this type of scoring was that assessors sometimes knew that communication in an event had taken place, but didn't know exactly what had been said, or were unable to score what they had heard the operator say.
Another explanation for scoring outside the system was that assessors in several occasions used the wrong line to note their score, such as an empty line right above or below an item. It cannot be determined without doubt if these scores were meant to be given to the items above or below these **X** items. Unfortunately these scores have to be considered noise in the data.
In the further analysis the recoded **A1**, **A2**, **S** and **W** scores were used. **T** and **X** items were not considered further. The consequences for the total dataset is that from a total of 5,535 data points (for all data types, see Table 4-3) 300 data points were omitted (**T** and **X** types). This is a loss of 5.7% of the data. For future use, the PMI-P score form should be adjusted to prevent this type of incorrect scoring.

4.5 Reliability of PMI scores

4.5.1 Reliability of assessors on PMI-P scores

The inter-assessor reliability on the scores on PMI-P items was studied by comparing the assessors' scores on communication of the operators. Each operator was assessed independently by two assessors, who both scored the operator's communication on the PMI-P score form. Table 4-4 presents the average number of items scored for each operator and the Pearson correlation between the scores of the two assessors.

The correlations between assessor scores range from .685 to .871 on the four types of categories. Especially the correlation between **A1** scores is high. **A1** items are the most important items and represent reactive responses to a call from a ship with important, necessary information, information which has to be presented (see section 4.2.1). The correlation between assessors on **W** items on the other hand is much lower. The role of W items is further discussed in section 4.5.4.

Item type	Assessor	Mean	Std deviation	Correlation
A1	Assessor 1	34.63	5.807	.871[16]
	Assessor 2	36.22	5.553	
A2	Assessor 1	3.75	2.070	.852[16]
	Assessor 2	4.16	1.979	
S	Assessor 1	2.44	1.253	.811[16]
	Assessor 2	2.39	1.293	
W	Assessor 1	5.02	2.142	.685[16]
	Assessor 2	5.45	2.253	

Table 4-4 Correlation between assessors on PMI-P item scores

The correlations between assessor scores range from .685 to .871 on the four types of categories. Especially the correlation between **A1** scores is high. **A1** items are the most important items and represent reactive responses to a call from a ship with important, necessary information, information which has to be presented (see section 4.2.1). The correlation between assessors on **W** items on the other hand is much lower. The role of W items is further discussed in section 4.5.4.

4.5.2 Reliability of assessors on E and L data

The inter-assessor reliability on the scores on **E** and **L** data was studied by comparing the assessors' scores on communication of the operators for these data. Each operator was assessed independently by two assessors, who both scored the operator's communication on the PMI-P score form. Pearson correlations were calculated for the scores on items that were scored *Early* or *Late*. Table 4-5 presents the results.

Most items were scored as on time, and only items that were scored at least a minute earlier or later were scored **E** or **L** (see section 4.2.1). Therefore Table 4-5 shows that mean scores of **E** and **L** data are much fewer than scores on PMI-P score form data. The scores of both **E** and **L** in the **S** category are very few (N=7) compared to the other categories of items. Only one **S** item was scored **E** or L. Therefore the **E** and **L** scores on S items are excluded from the remaining analyses. Table 4-5 shows that correspondence between the other items range between .397 and .668. Correspondence between assessors was lower on **E** and **L** scores than it was on the PMI-P item scores, but it was still significant at the same probability level. The highest agreement can be found in the **A1** category for the **E** scores and in the **W** category for **L** scores.

[16] Correlation is significant at the 0.01 level (2-tailed).

		Early		
Item type	Assessor	Mean	Std deviation	Correlation
A1	Assessor 1	1.25	1.479	.642[17]
	Assessor 2	1.52	1.624	
A2	Assessor 1	.04	.188	.546[17]
	Assessor 2	.06	.245	
W	Assessor 1	.15	.354	.375[17]
	Assessor 2	.16	.372	
		Late		
Item type	Assessor	Mean	Std deviation	Correlation
A1	Assessor 1	.45	.644	.599[17]
	Assessor 2	.58	.783	
A2	Assessor 1	.09	.289	.397[17]
	Assessor 2	.11	.313	
W	Assessor 1	.29	.513	.668[17]
	Assessor 2	.33	.576	

Table 4-5 Correlation between assessors on items scores Early and Late

4.5.3 Reliability of assessors on behavioural scale data

The inter-assessor reliability on the scores on behavioural scales was studied by comparing the assessors' scores of operators for each scale. Each operator was assessed independently by two assessors, with a low score representing a good performance.

		N	Mean	Std deviation	Correlation
Finding solutions,	Assessor 1	109	1.8273	.46647	.497[18]
improvising	Assessor 2		1.7798	.62890	
Global overview	Assessor 1	110	1.6455	.62935	.637[18]
	Assessor 2		1.5273	.63114	
Setting priorities	Assessor 1	109	1.5636	.59858	.415[18]
	Assessor 2		1.5046	.60283	
Use of voice	Assessor 1	109	1.5688	.67174	.138
	Assessor 2		1.8624	.55223	
Flexibility	Assessor 1	109	1.7339	.50263	.359[18]
	Assessor 2		1.5963	.57926	
Coping with stress	Assessor 1	108	1.4037	.59503	.379[18]
	Assessor 2		1.4722	.58738	
Decisiveness	Assessor 1	109	1.7064	.53234	.427[18]
	Assessor 2		1.7545	.65226	
Motivation	Assessor 1	110	1.2909	.49484	.358[18]
	Assessor 2		1.2091	.43039	
Global effectiveness	Assessor 1	109	1.7636	.57353	.547[18]
	Assessor 2		1.7523	.64054	

Table 4-6 Correlation between assessors on behavioural dimensions

[17] Correlation is significant at the 0.01 level (2-tailed).
[18] Correlation is significant at the 0.01 level (2-tailed).

Correlation between the scores of assessors for each operator on the behavioural scales are presented in Table 4-6. Table 4-6 shows that correlations between the scores on the behavioural scales are lower than those on the PMI-P score form, but are still significant. The explanation for this is that the PMI-P score form uses a strict protocol and scores performance on the level of individual items. This produces a more consistent scoring than the subjective post-hoc assessment on behavioural dimensions.

4.5.4 Conclusion

The study in Rotterdam provided an excellent opportunity to collect data with PMI-P. All VTS operators of the Rotterdam Port Authority participated in the study. Therefore the data set contains all of the variability that can be found (in one test) in performance on this scenario for the total population. The data have been cleaned up and validated on a global level.

The general lay-out of the form and the use of categories leave room for confusion on the part of assessors. Combining the description of the scenario with the PMI-P score form has resulted in a score form with extra information on timing of events, and on sources of information to be used in communication with the operators. This has sometimes led to confusion in scoring operator performance, but the amount of confusion is not great enough to call the results into question as has been shown in Table 4.3 in section 4.4.

The PMI-P score form is based around individual items that need to be communicated, not around events in the scenario. Therefore events on the score form are not presented as a separate category of items. Assessors are not expected to score an event as a whole, but need to score individual items within an event. The completed score forms however display that in a considerable number of events, the assessors did score just the whole event and not the individual items. Apparently there are situations where the assessor judges that communication in the situation has taken place, without considering the separate elements that constituted the communication. Assessors sometimes have a need to score whole events in just one strike. This problem is exacerbated by the fact that **W** items (items that the operator has to perceive and react upon on his own initiative) often have to be scored on the form in just one line. This may encourage assessors to use this type of scoring (in just one line) in other events as well.

The analyses in this chapter have to determine if it is advisable to consider adding an extra category for scoring whole events to the PMI-P score form, or if the scoring of complete events should be made impossible on the PMI-P score form. This issue will be discussed in section 4.9.2.

According to this validation the method produces consistent results between the assessors on the PMI-P scores. The **E** and **L** data are less reliable, but can still provide useful

information. The whole PMI-P data set will be used in the subsequent sections of this chapter to study the issues relevant for situation awareness of VTS operators and RSA.
A differentiation between highly effective (*HE*) and ineffective (*IE*) operators based upon the behavioural rating scales will be used to differentiate between effective and ineffective performance. The selection of these groups of operators will be described in the next section.

4.6 Validation of the PMI-P score form

In the FLO research project all 110 VTS operators employed by the Rotterdam Port Authorities participated (MSR, 2001). This is a very large group of professional subjects for such an experiment in the maritime field. The results of the project therefore give an excellent indication how operators perform on the test. This however does not automatically imply that the test produces consistent results. The consistency of scoring was studied in the previous section. This section addresses the consistency of the scoring materials, in particular the PMI-P score form. Content validity was studied for both the use of the most important type of items used in the scoring system (**A1**) as for the scoring in agreement with the time line of the scoring form (*Early* or *Late*).

The content validation of PMI-P performed in the FLO project did not investigate individual items in detail. Overall scores of operators were assessed and compared with scores of other parts of the *PMI* method, to establish a validated method and a baseline measurement of VTS operator performance. The method was used to make statements on VTS performance on a high level of abstraction.

The focus of this chapter is the study of VTS operator situation awareness and how that relates to RSA in a scenario. To understand RSA and VTS operator situation awareness, an investigation is required on a more detailed level than that in the FLO project, on the level of individual events in the scenario.

4.6.1 Aims

The central question of this validation was whether the PMI-P score form, developed and used in the FLO study, was valid as a method for assessing VTS operator performance on the level of detail required to study RSA in the scenario and situation awareness in the performance of operators. In this validation the following issues were addressed.

- The most important items in the PMI-P score form were the **A1** items. These items contained necessary information (see section 4.2.1 for details). The PMI-P score form used in the FLO experiment had in total 105 different items of the **A1** category. The first question is whether all these items are "necessary" for effective performance and whether they are all equally important to assess effective VTS operator performance. This question addresses the issue whether all **A1** items refer to RSA, or if some of them refer to issues that are not required in the situation.

- The PMI-P score form was not only used to score items, but also provided an indicator of timing of communication in the form of **E** and **L** scores. The question studied in this section is whether these categories are used consistently in the PMI-P scores in the FLO experiment.

4.6.2 Methodology

4.6.2.1 Hypotheses

A1 items in the PMI-P score form are considered necessary information, i.e. information that an operator has to provide to the ships. In the scenario used in the FLO study 105 **A1** items were identified in total. Table 4-4 has shown that **A1** items are the ones scored most often for the operators and most consistently by the assessors. The question remains if the PMI-P score form contains the right amount of **A1** items and if all items that are labelled **A1** should indeed be labelled as such. On the PMI-P score form no distinction is made between different **A1** items, suggesting that they are all equally important.

Hypothesis 5

*The amount of **A1** response items on the PMI-P score form is well-chosen to enable scoring of effective VTS operator performance if all **A1** items are scored at least once, and if no **A1** items are scored much more often than others.*

Items that were scored earlier or later than expected were scored with a label **E** or **L**. If the PMI-P score form has been developed well, **E** and **L** scores will be spread randomly over items and not occur systematically at particular items in the PMI-P score form. However, if certain items, or even complete events, are scored systematically *earlier* or *later* than expected on the PMI-P score form in the scenario, this may be an indication that the score form is not correct, and that these items need to be put in another location on the form.

Hypothesis 6

*The PMI-P score form has been developed well if **E** and **L** scores are spread randomly over items and events and not systematically linked to particular items or events on the PMI-P score form.*

4.6.2.2 Dataset

The data used in this study were those on the communication of all operators on the traffic scenario scored on the PMI-P forms. The scores on the different data items **A1**, **A2**, **S**, and **W** were used in this part of the study. The **E** and **L** data were used to study Hypothesis 6. The data set consisted of the following:

- Traffic scenario used in the Rotterdam study
- PMI-P score form

- Performance scores of all operators on PMI-P data on **A1**, **A2**, **S**, and **W** scores by two independent assessors.
- Performance scores on **E** and **L** data of all operators.

4.6.2.3 Procedure

To test the hypotheses the data collected during the FLO study were used. The data were collected according to the criteria presented in section 4.2.1. The data were analysed using SPSS for statistical analysis, and MICROSOFT EXCEL spreadsheets for in-depth analysis of PMI-P scores on item data, and **E** and **L** data.

All PMI-P response data were entered in three separate MICROSOFT EXCEL spreadsheets; one for the item scores, one for the E scores and L scores. The operators were represented in the columns of the spreadsheet. For each operator there were two columns. Each column contained the scores of one of the assessors. The items were represented in the rows of the spreadsheet. Each row contained one communication item. The scores could be summarised over either operators or items.

4.6.3 Results

4.6.3.1 Validity of A1 items

In total the PMI-P score form consisted of 105 **A1** items scored by two assessors, giving a potential score of 210 in total. Some of these items were scored only once or twice, while other items were scored for all operators. Figure 4-3 presents in a histogram the distribution of this scoring.

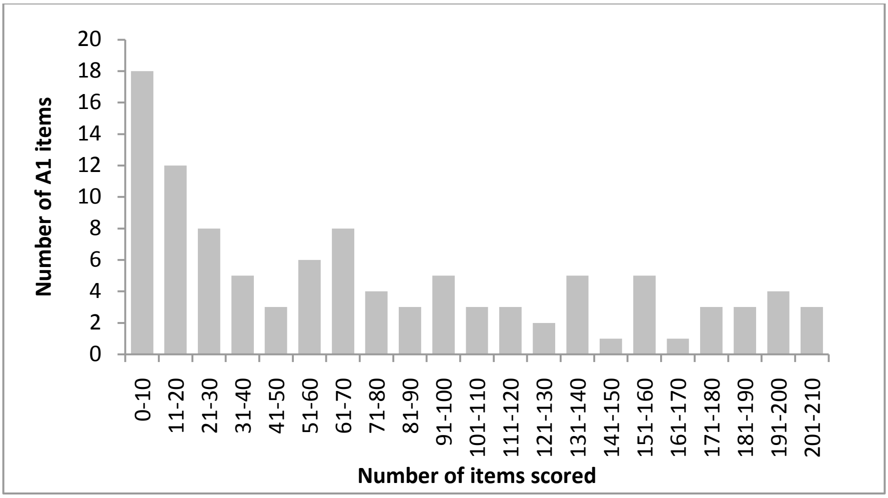


Figure 4-3 Scores on individual A1 items for all operators

The X-axis of Figure 4-3 shows the number of scores an **A1** item received (ranging from 1-209). The Y-axis displays the frequency that **A1** items were scored that number of times. As an example: 18 **A1** items were scored for 0-10 operators.

It is clear from Figure 4-3 that there is a large variance in the amount of times that **A1** items were scored. Several items have been scored once or twice, while others have been scored by all operators. One reason for the variance in scoring of items may be found in how communication is structured on the PMI-P score form. As an example, Table 4-7 presents one event from the scenario.

At 2.30 minutes in the scenario, the DONGESTROOM departed from the Shell-1 quay. There were two issues that the operator needed to tell the ship (items number 35 and 36, marked **A1**; communication about a barge 2500 meters, and a coaster 3000 meters upstream), and one issue that an operator could mention (item number 37, marked **A2**; about the seagoing KOBULETI). The operator needed to tell the KOBULETI about the departure of the DONGESTROOM (item 38 marked A1).

34	02.30	DONGESTROOM coming loose Shell-1 inward bound	Scored (%)
35	**A1**	Outward bound barge 2500 m. above the entrance NZH. There has been no communication with this ship	63
36	**A1**	Outward bound coaster (YEHYA) 3000 m. above the entrance NZH. Destination Sea	68
37	**A2**	Information: Departing KOBULETI has left Chevron. Bound for Sea	44
38	**A1**	Report departure DONGESTROOM to KOBULETI	9

Table 4-7 Part of the PMI-P score form at 2':30"

The last column of table 4-7 presents the scores for the different items. While the first two **A1** items (lines 35 and 36) are scored for a large number of operators, the third **A1** item (line 38) in the event is scored much less.
Table 4-8 presents the percentages of **A1** items scored in relation to their ranking within the events on the PMI-P score form. 52 **A1** items were the first item in the event mentioned on the score form. These items were scored for 51% of the operators on average. The other **A1** items, ranking lower on the score form were mentioned much less.

Rank	N	% Scores
1	52	51
2	30	36
3	12	21
4	10	25
5	1	7

Table 4-8 PMI-P scores A1 items ranked according to their location within an event

The following can be concluded from Table 4-8. Operators did not mention every item in an event. If they did mention something, it was more likely that they mentioned the events that were listed early in the list of responses. Apparently there is a hierarchy within the ranking of **A1** items on the PMI-P score form. In constructing the PMI-P score form, in many events the assessors put the most important item first.

4.6.3.2 Validity of E and L scores

To determine the validity of E and L scores these scores were studied both on the level of individual items and on the level of events.

4.6.3.2.1 E and L scores on individual Items

There are large differences between items regarding the number of times that they were communicated *Early* or *Late*. Table 4-5 in section 4.5.2 showed that operators do not score many **E** and **L** items. 69% of items were never scored **E** or **L**, 19% of items were scored **E** or **L** by one assessor, and only 12% of items were scored **E** or **L** by both assessors. It is therefore more interesting to look at the outliers than at the distribution in detail.
Outliers in **E** and **L** scores were defined as items that score above the 95% percentile of the distribution of scores (the 5% boundary always equals 0). The boundaries for the different item categories for **E** and **L** scores are presented in Table 4-9.

		Mean	Std Deviation	5th Percentile	95th Percentile
A1 first	Early scores	3.25	8.80	.00	19.70
	Late scores	.96	2.67	.00	5.70
A1 other	Early scores	2.55	7.02	.00	21.30
	Late scores	1.23	2.68	.00	8.10
A2	Early scores	.40	1.04	.00	3.35
	Late scores	.73	2.21	.00	6.50
W	Early scores	.92	2.50	.00	10.00
	Late scores	1.95	7.23	.00	8.00

Table 4-9 Means, standard deviations and 95% boundaries for E and L scores

Table 4-10 presents all of the 7 items that scored above 95% on either E or L. Two of the items score high on the **L**, the others score high on the **E**. Three of these items (marked *) are all from one particular event in the scenario. This particular event is looked at in more detail in the next section.

Of the remaining items the first one was scored **L** many times. This is the only **W** item from the list, which means that operators could address this item on their own initiative, based upon their own observation, without a request from a ship. This particular item refers to a vessel that is going up the river. The WATERBOOT 12 is a small boat sailing around in the harbour area delivering water to ships. This particular event describes that the boat is going down the river, staying on the north side without notification (the "wrong" side of the fairway). Six minutes later the boat will cross the river to the other side of the fairway, also without notification. Since this is a small boat that is familiar with the harbour, such behaviour is not uncommon. The particular event where the WATERBOOT is going down the river, staying on the north side, is noticed and communicated 131 times (this refers to scoring by both assessors, therefore scored for 59% of the operators), but remarked upon *Late* 55 times (for 25% of the operators). Apparently the operators do not find it necessary to immediately call the WATERBOOT about her intentions, but when she does not report herself soon enough, they take the initiative and ask for the information.

Type		Early	Late
W	Observation: WATERBOOT 12 is on the river, staying on North side until NG10	5	55[19]
A1 First	Information TANGER	0	21[19]
A1 First	Outward bound seagoing vessel (YEHYA) 1500 m. above the Eastern Head NZH (*)	77[19]	0
A1 Other	Unknown barge (MARIA) 300 m, behind that ship (*)	53[19]	3
A1 First	Outward bound seagoing vessel (YEHYA) 600 m. above the Eastern Head, going towards sea	29[19]	0
A1 Other	KOBULETI report back to YEHYA (*)	27[19]	15[19]
A1 Other	**Telephone call**: information to RWC (Regio Wachtchef)	24[19]	1

Table 4-10 Items that score E or L outside of the 95% boundaries

The other *Late* item presents information to the TANGER about the LILIANA who has just had an accident and is now anchoring. The TANGER needs to be informed about this. At the same time the operator also needs to inform the authorities about what has happened to the LILIANA. These items are directly following the item about the TANGER. Many operators inform the authorities first and then contact the TANGER. The time this takes causes the item on the TANGER to score *Late*.

The two remaining items are artefacts of the scoring form and refer to items that are located in the PMI-P score form several times, because they concern communication for which there are several occasions in the scenario. In both these events the item that is scored *Early* is preceded by the same item a few minutes earlier. The **E** score on the items refers to the fact that the item has been scored on the previous occasion.

4.6.3.2.2 E and L outlier scores on Events

The previous section has identified one event in the scenario where several items are scored **E** or **L** by a large number of operators. This event is described in Table 4-11.

[19] Item scoring above 95% on **E** or **L**

50	08:00	KOBULETI passing Shell 1 bound for sea	Early	Late
51	A1 First	Outward bound seagoing vessel (YEHYA) 3000 m. above the E-Head NZH	29	0
52	A1	Unknown barge (MARIA) 300 m, behind that ship	53	3
53	A1	One seagoing vessel inward 2000 m. below W-Head NZH destination MMVW	6	0
54	A1	KOBULETI report back to YEHYA	27	15
55	A1	KOBULETI report back to BONGO	7	1

Table 4-11 Event at with high number of items scored E or L

The KOBULETI calls the VTS operator passing the SHELL 1 harbour. At that phase in the scenario the traffic is relatively quiet and there is also no communication going on. The previous call is two minutes ago, while in most parts of the scenario there are calls at least every minute and most of the time more often. Apparently operators do not want to wait for the KOBULETI'S call and find this quiet period a good moment to show initiative, and start communication with the ship. This is shown by the large number of *Early* ratings, for the next five items (three of which exceed the 95th percentile).

The explanation for the deviating scores of this event therefore may be found in the situation. This is in itself not an indication that items were located in an incorrect location on the PMI-P score form. The items were located with regard to when the information is needed. Assessors located the items based upon RSA of the ships. Several operators communicate the items based upon their own workload: if there is a period in which there is not much to do, operators look around for other information they can already provide.

4.6.4 Discussion and conclusion

This section has presented a thorough examination of the PMI-P score form used in the FLO study. First the use of A1 items in the score form was critically examined. Next the scoring of data that were rated as being communicated *Early* or *Late* was studied.

4.6.4.1 Use of A1 items

A1 items are "Reactive responses to a call from a ship with important, necessary information, information which has to be presented" (see section 4.2.1). The PMI-P score form used in the FLO study contained 105 of such **A1** events. The completed score forms show that not all **A1** items are scored equally frequently. Operators score significantly better on **A1** items that are put first in an event on the PMI-P score form than on those placed lower in an event. Assessors have put what they consider the most important issues for communication first on the score form, and the operators communicate in agreement with this order. The lower ranking A1 issues should be studied carefully; maybe some of them should be **A2**'s. This is something that needs consideration in future research.
Another explanation for this is that the PMI-P score form consists of a list of observable communication items that an operator can perform during the scenario. In setting up the

PMI-P score form an attempt was made to make the list complete: all potential communication between VTS operator and ships and between VTS operator and others is presented as items in the PMI-P score form. There is communication that can take place at several locations (moments) during the scenario. This communication is necessary but has to take place only once. Therefore, even though the communication is necessary and therefore categorised as **A1**, it does not have to be scored at all locations in the PMI-P score form. In retrospect this issue could have been prevented if these similar items were concatenated in some form on the PMI-P score form and in the analysis.

Later sections of this thesis will differentiate in the analysis between the scores of **A1 First** and **A1 Other** items.

4.6.4.2 E and L scores

The issue studied in this section was whether **E** and **L** scores were related to specific items on the PMI-P score form. If this were to be the case, than that could be an indication of a problem with the PMI-P score form. Analysis of **E** and **L** data over PMI-P items revealed that there E and L scores were not spread over the data at random. There were indeed a few items on the PMI-P score form that received a high score of **E** and **L** responses.

A closer analysis of these data showed that for each of the items there was a clear explanation why some operators might be scored on the item *earlier* or *later* than the PMI-P score form indicated. This was not an indication that these items were located incorrectly on the PMI-P score form. The high-scoring elements were items that could indeed be communicated earlier or later than their indicated location on the score form, without disturbing the traffic process.

One of the items depended upon the perception of the operator and therefore could be noticed earlier or later. The other high scoring items were related to communication that could have taken place earlier according to the scenario and PMI-P scoring form. This depended on how the operator wanted to spread his workload, since these items were located in a relatively quiet part of the scenario. The analysis did not indicate that the PMI-P score form was not correct. However, since there is no absolute procedure that describes in what order communication needs to be handled, it is worthwhile to re-examine the PMI-P score form to see if some of these high scoring items need to be relocated or added once more to the score form.

4.7 Selection of Effective and Ineffective operators

4.7.1 Aims

The subsequent sections of this chapter use the data collected with the PMI-P score form to increase understanding of VTS operator performance and situation awareness.

It is useful to differentiate between effective and ineffective operators to get a better understanding of what constitutes effective performance and situation awareness. The behavioural rating scales described in section 4.2.2 provide an assessment of VTS operator performance that is not directly related to the score on PMI-P: it is a separate assessment of the same VTS operator performance. Ratings of operators on these scales will be used to create two groups of operators that can be compared on performance on PMI-P.

4.7.2 Development of a scale and selection of operators

Using the post-hoc rating scale of behavioural dimensions a selection was made of operators whose performance was, according to the assessors, Highly Effective (HE) or Ineffective (IE). The rating scale is a multi-dimensional scale of behavioural dimensions (see section 4.2.2). Discussions with the assessors who performed the scoring in the study revealed that their scores on some dimensions of the rating scale were more related to performance, while scores on the other dimensions were more related to what they perceived as the attitude of the operator towards the work. One dimension (use of voice) was related neither to performance nor to attitude. This dimension was one characteristic of VTS operator performance that is considered important in communicating correctly with traffic, but does not relate to the content of what is communicated. Four dimensions were considered to be more an assessment of attitude than of performance: "*flexibility*", *coping with stress*", "*decisiveness*", and "*motivation*".

- *Flexibility* was demonstrated when operators could easily adapt to changing situations, such as working in a different traffic area than they were used to (i.e. the area used in the scenario), responded correctly to sudden changes in the scenario, such as changing weather conditions. It was not a direct measurement of performance.
- *Coping with stress* referred to whether operators got excited during the exercise, especially during the part of the scenario where the accident happened. Their stress did not necessarily affect their communication to traffic.
- *Decisiveness* was demonstrated by operators who did not hesitate to act when the situation required. According to the assessors a decisive acting was not necessarily a correct action: operators sometimes acted incorrectly, but with determination.
- *Motivation* referred to the willingness of operators to perform their best at the tests. Some operators showed in their behaviour that they were not interested in participating in the test, while others were very motivated to participate.

It was felt that these four scales were not appropriate to use as measures of effective performance.

Three dimensions that according to the experts were scored on the basis of the observed performance were "*finding solutions, improvising*", maintaining a "*global overview*", and "*setting priorities*".

- *Finding solutions, improvising* meant that the operator came up with adequate solutions to solve problems in the traffic. This indicated that the operator had the required situation awareness (RSA) on level II and/or III, to know what the correct action was.

- *Global overview* was maintained by operators who were able to guide individual ships, while explicitly taking into account other traffic. Maintaining overview refers very directly to maintaining situation awareness on level II.
- S*etting priorities* referred to the ability of operators to address complex situations by solving most important problems first. This indicated that the operator understood what was going on in the traffic, and what was going to happen next. This required situation awareness (RSA) on level III.

The scale for "*global effectiveness*" was used by the assessors to provide an overall judgement of the performance of the operators. This scale was used to score whether operators kept a good balance between safety and efficiency throughout the simulation.

The three scales that signified performance "*finding solutions, improvising*", "*global overview*", and "*setting priorities*", and the overall performance scale "*global effectiveness*" were used to select operators who had performed effectively (without taking into account their attitude towards the work). Using these post-hoc rating scales a selection was made of operators whose performance was, according to the assessors, *Highly Effective (HE)* or *Ineffective (IE)*. Fifteen *Highly Effective (HE)* operators and fifteen *Ineffective (IE)* operators were selected, according to the following criteria.

1. For the selection of *HE* and *IE* operators the ratings on these four dimensions from both assessors were combined (see section 4.2.1 for details).
2. For the combined score, the ratings on the four dimensions mentioned were added together to create a new variable. The correlation between assessors on this new dimension was higher than the scores on the dimensions separately (.711; $p<0.01$), therefore the scores of both assessors were added creating a new variable called here *Situation Awareness Related Performance Rating* (SARPR).
3. Operator scores on this combined scale SARPR ranged from 8-23. Scores are shown in Figure 4-4. The operators with the lowest score on SARPR had the best performance and *vice versa*. Figure 4-4 shows that eight operators received a minimum (best) score of 8. On the other end of the scale none of the operators received a maximum (worst) score of 24.
4. To be selected as *HE* operators, the score on SARPR had to be **8** or **9** (A score of 8 was the best score that could be obtained. To achieve this, an operator had to receive the best score of **1** on each of the dimensions by both assessors. If an operator received a score of **2** on no more than one dimension by one assessor, an operator could still count as *HE*).
5. The *IE* operators were selected using the worst scores on SARPR. For the *IE* operators the score on SARPR was **17** or more. To make up the total number of 15 IE operators, one operator with a score of 17 was selected randomly out of the three operators with that score.

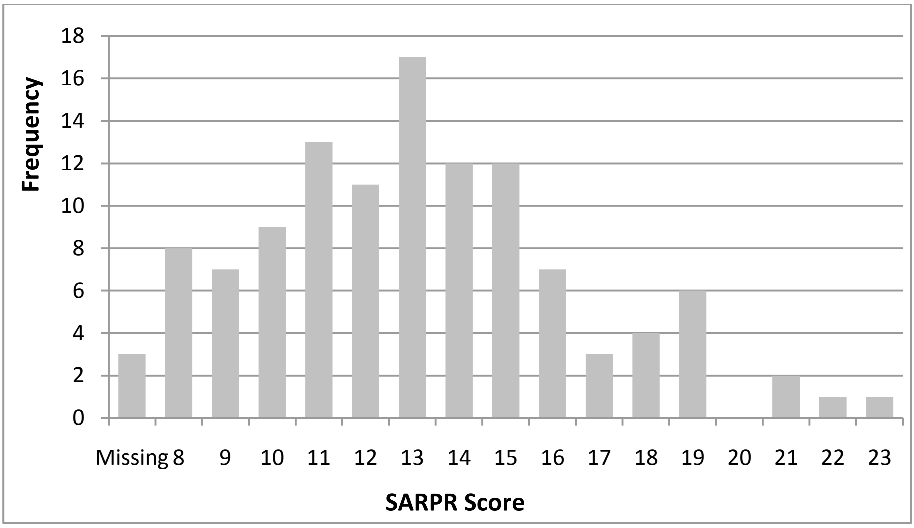


Figure 4-4 SARPR scores for all operators

The completed PMI-P performance assessment score forms of the selected 15 *HE* and 15 *IE* operators were used for the analysis of communication and situation awareness in the next sections.

4.7.3 Conclusion

It was possible, using the behavioural scores, to select a group of highly effective and ineffective operators. This was done by concentrating on the four performance related scales and combining them into a new scale (SAPR). This scale also achieved a much higher, and significant, inter-rater correlation than the individual scales (section 4.5.3). These groups of operators will be used in section 4.8 to determine how scores on PMI-P relate to effective performance and how effective performance relates to situation awareness.

4.8 Performance, Communication and situation awareness

4.8.1 Aims

In this section the FLO data set is used to study the relation between VTS operator performance and situation awareness. Section 4.8.3.1 analyses PMI-P data to study the differences in performance scores between IE and HE operators and relates that to their communication skills. Section 4.8.3.2 studies the differences between RSA in the scenario and VTS operator situation awareness as assessed in PMI-P. The aim of this section is to show how differences in VTS operator scores on PMI-P can be explained by differences in VTS operator situation awareness.

4.8.2 Methodology

4.8.2.1 Hypotheses

PMI-P uses communication to assess VTS operator performance. Therefore it may be the case that a score on PMI-P is nothing more than a measure of communication skills of operators, and not a measure of their understanding of traffic or their situation awareness. The first assumption to be tested is that this is not the case.

Hypothesis 7

PMI-P is regarded a valid instrument for assessing VTS operator performance if differences in PMI-P scores between HE and IE VTS operators cannot be explained by differences in communication skills and performance, but refer to something else.

If communication performance (the scores on A1, A2, S and W items) alone cannot account for all differences in scores on PMI-P between *HE* and *IE* operators, then the differences must be explained by something else in the VTS operator performance.

The PMI-P scoring form differentiates between different types of responses. Necessary communication is labelled **A1**, additional information is labelled **A2** etc. Differences in responses on the various response types measure if an operator knows what information is important in the situation and what is not.

That does not mean that the scoring of situation awareness has become straightforward using PMI-P. For instance: when a situation contains both **A1** and **A2** items, a score on an **A2** has to be evaluated in relation to the score on the **A1**. When an operator has RSA of the situation (with good level II situation awareness: comprehension) he will score the **A1** items required for that situation. When he has more situation awareness than required, he may also score several **A2** items. The scoring of **A2** items then indicates very good situation awareness. However, if the operator scores the **A2** items, but does not score the **A1** items, this may be interpreted as an indication that the operator has a wrong interpretation of the situation and therefore lacks situation awareness. The operator then displays situation awareness level I (perception of elements), but no good situation awareness level II (comprehension).

Hypothesis 8

PMI-P is regarded a valid instrument for assessing VTS operator performance if it demonstrates differences in Level I and Level II situation awareness between HE and IE operators.

4.8.2.2 Data set

The data used in this study were those on the communication of the Highly Effective (*HE*) and ineffective (*IE*) operators defined in section 4.7 on the traffic scenario scored on the PMI-P forms. The scores on the different data items **A1**, **A2**, **S**, and **W** were used in this part of the study. The data set consisted of the following:

- Traffic scenario used in the Rotterdam study
- PMI-P score form
- Performance scores of all operators (including the IE and HE operators) on PMI-P data on **A1**, **A2**, **S**, and **W** scores by two independent assessors.

4.8.2.3 Procedure

The data were analysed using SPSS for statistical analysis, and MICROSOFT EXCEL spreadsheets for analysis. Several formats are used to present the results, including different types of figures, charts and tables.

The completed PMI-P score form used for these analyses consisted of a large MICROSOFT EXCEL spreadsheet. The operators were represented in the columns of the spreadsheet. For each operator there were two columns. Each column contained the scores of one of the assessors. The items were represented in the rows of the spreadsheet. The rows followed the timeline of the scenario. Each row contained one communication item. The items were labelled with time of occurrence in the scenario and item category. Separate spreadsheets were used for the PMI-P scores. The scores could be summarised over either operators or items.

4.8.3 Results

4.8.3.1 PMI-P scores, communication and performance

First question to be answered is if differences in scores on SARPR (derived in section 4.7.2) used in the FLO study only reflect differences in communication skills and performance between operators. Is good communication the only factor that determines the differences in scores, or is there something else, such as better understanding of the traffic, more situation awareness, or more precisely more awareness of the issues required (RSA), and can it be proven to play a role in the scores that operators receive on the SARPR scale?

A first indication of the difference in performance between the two groups is the amount of items (of all types) communicated during the test. Table 4-12 shows the average amount of items scored by the two groups of operators and their standard deviations. Figures and tables show the data for 15 operators in each group (*HE* and *IE* operators), each scored by two assessors, therefore N=30.

	Grouping	N	Mean	Std. Deviation
Number of Items	IE	30	46	8.087
	HE	30	55	8.356

Table 4-12 PMI-P scores of HE and IE operators

Analysis shows that the *HE* operators communicated significantly more than the *IE* operators ($t = -4.663$; $p < 0.025$). At the same time, there is a considerable overlap between the two groups.

A closer look at the performance data shows that differences in amount of items scored are not distributed equally over the categories of items. Table 4-13 shows how the two groups of operators scored on different categories of items.

Item	Operator	N	Mean	Std. Deviation	Std. Error Mean
A1First	IE	30	18.50	3.848	.703
	HE	30	22.80	3.398	.620
A1Other	IE	30	14.83	3.842	.702
	HE	30	17.03	3.306	.604
A2	IE	30	4.50	1.757	.321
	HE	30	4.87	2.270	.414
S	IE	30	1.73	.828	.151
	HE	30	1.70	1.264	.231
W	IE	30	4.13	2.063	.377
	HE	30	5.80	1.955	.357

Table 4-13 PMI-P scores of HE and IE operators for different types of items

Table 4-13 shows that *HE* operators scored higher especially on **A1** and **W** items. For situation awareness these are the most important types, as was described in section 4.2.1. The **A1** items are the most important for understanding of the situation, level II situation awareness. This is even more so for the **A1** first items. They have been put first in the events because assessors agreed that they were the most important issues that needed to be communicated. So the A1-first items that assessors regarded as the most important, with the highest RSA in the situation, are scored best and with the largest differences by HE operators. It is interesting to notice that **W** items, issues that required operators to act upon their own initiative, which was the item related to Level III situation awareness, were scored better by *HE* operators as well.

The statistical significance of the differences in scores was tested using a T-test. The results in Table 4-14 show that differences between HE and IE operators on all three variables A1-First, A1-Other and W were significant. Table 4-14 also shows that there are no significant differences between HE and IE operators in the scores on the scores of the A2, and the S items. As is argued in section 4.2.1, these items are less interesting for situation awareness, and may even reflect that the operator was looking at issues in the situation he should not be focussing on.

The results so far are a clear indication that *HE* operators did not just communicate more, but that they were also more aware of relevant items within the scenarios than *IE* operators were.

	t	Df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Conf. Interval of the Difference	
						Lower	Upper
A1First	-4.588	58	.0003	-4.300	.937	-6.176	-2.424
A1Other	-2.377	58	.021	-2.200	.925	-4.053	-.347
A2	-.700	58	.487	-.367	.524	-1.416	.682
S	.121	58	.904	.033	.276	-.519	.585
W	-3.212	58	.002	-1.667	.519	-2.705	-.628

Table 4-14 Independent samples t-test on PMI-P scores of IE and HE operators

A closer look at the items scored reveals that the differences between what *HE* and *IE* operators score is even larger than Table 4-13 shows. Figure 4-5 presents how the 180 items were scored by the two types of operators. Four categories of responses are distinguished in Figure 4-5: Items scored by both *IE* operators and *HE* operators; items scored only by *IE* operators; items scored only by *HE* operators; and items scored by neither *IE* nor *HE* operators.[20]

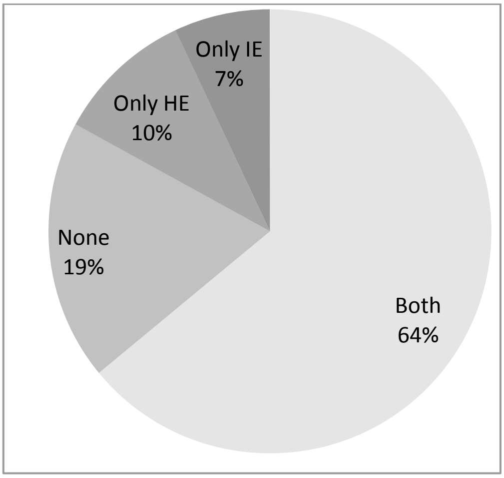


Figure 4-5 Distribution of items scored by HE and IE operators

[20] The categories are cleaned with respect to small differences in scores: Items with a score of either 0 or 1 are assigned to the category of not scored items. Items scored once refer to items where only one of the two assessors observed the communication, while the other did not. This was the case for 15 items that were subsequently labelled as not-scored items.

Figure 4-6 takes a closer look at the items that were only scored by HE or IE operators. Figure 4-6 shows that not only the amount of items that IE operators and HE operators scored differs, but that they also score different items. Figure 4-6 presents the distribution of the items scored only by IE operators and only by HE operators over the categories of items.

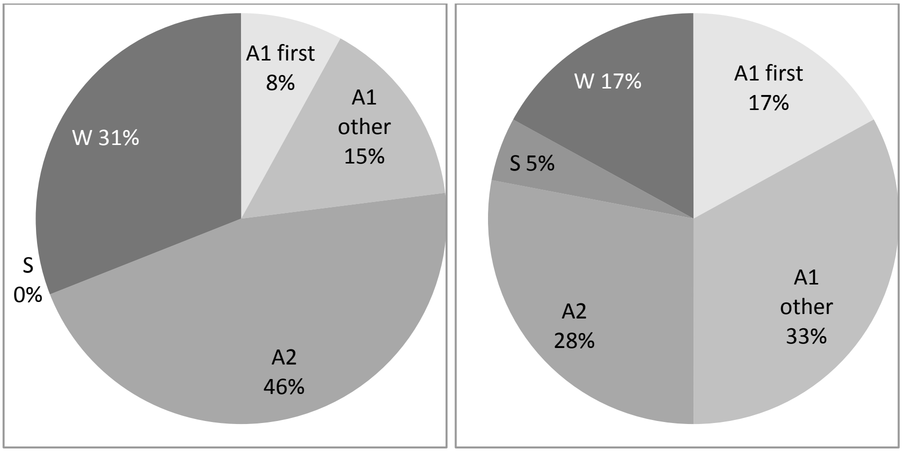


Figure 4-6 Distribution of data types scored only by IE operators (left) and HE operators (right)

Items scored by *IE* operators but not by *HE* operators are mainly **A2** (46%) and **W** items (31%). Items scored by *HE* operators but not by *IE* operators on the other hand are mainly **A1** items (A1 first 17%; A1 other 33%). Figure 4-6 demonstrates that *HE* operators score on the issues that are important for situation awareness; they are not just being more communicative.

Figure 4-7 shows a different way of presenting these results. For each of the four types of categories **A1**, **A2**, **W**, and **S** distribution of differences in scores on all individual items is shown in percentages. The horizontal axis in Figure 4-7 refers to whether an individual item was scored more often by *HE* or by *IE* operators. The vertical axis in the figure refers to the percentage of items of each category for which this is true. For example, out of 105 **A1** items, 1% of items was scored 16 times more often by *HE* operators than by *IE* operators.

Figure 4-7 shows that, even though *HE* operators did score more items of all categories, in each category there are several items that were scored more often for *IE* operators than by *HE* operators. Apparently the difference between the *HE* and the *IE* operators is not only the amount of items scored. There are also differences in which items are scored. The next section will take a closer look at these differences, to investigate if these differences are relevant to understanding VTS operator situation awareness.

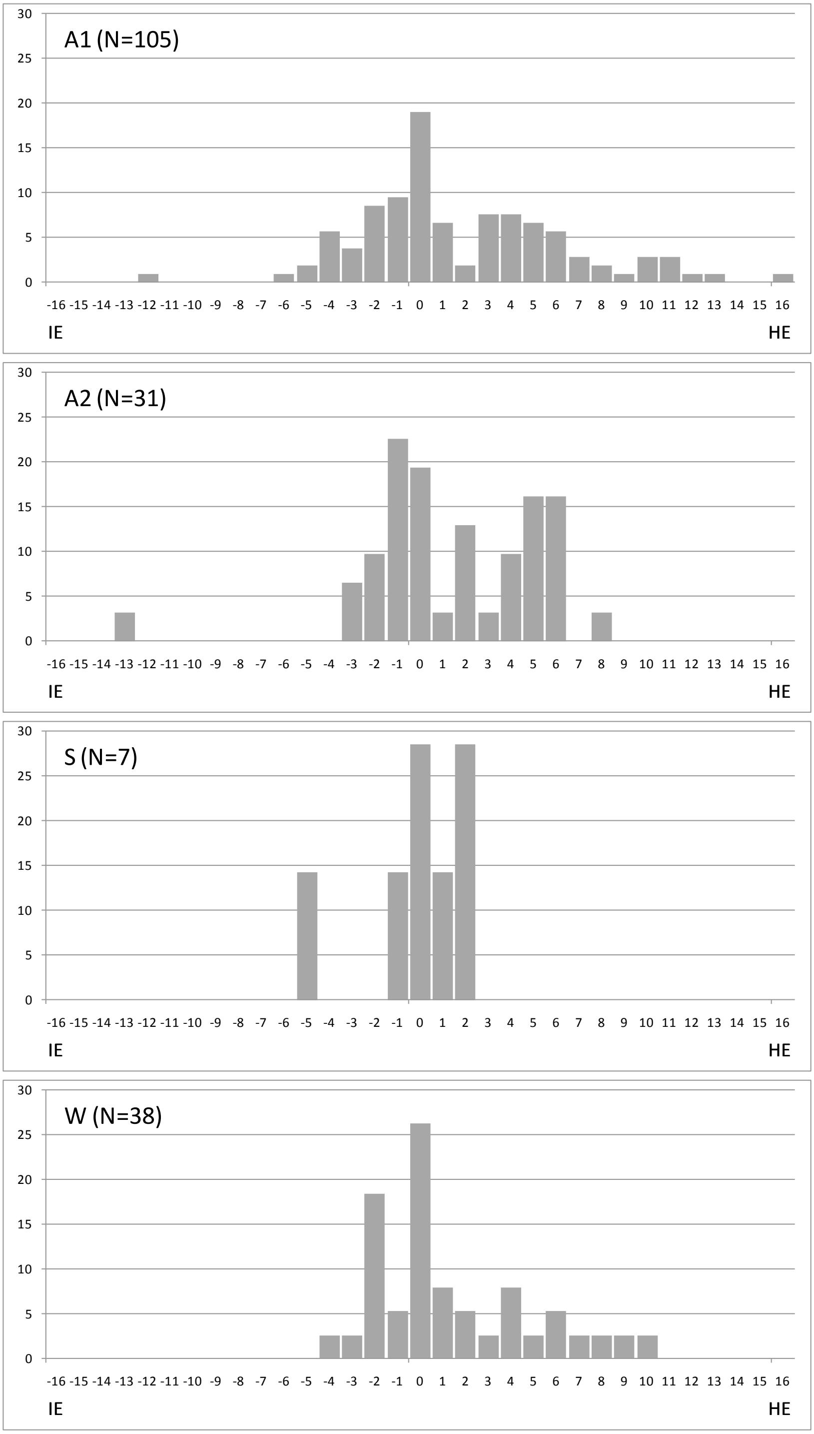


Figure 4-7 Distribution in % of differences in scores on categories of items by HE and IE operators

4.8.3.2 PMI-P scores and situation awareness

The difference between *HE* and *IE* operators is not only that the *HE* operators have more or better communication than *IE* operators. The previous section has shown that they also communicate about different items. Discussions with operators after they finished their test and discussions with the assessors about the results of the tests have provided several possible explanations for the differences in scores. This section will study three of these explanations. The first explanation is a difference in how operators think that the work needs to be done. There are two approaches possible here. Some operators believe that everybody on the water needs to have full information about everything. The other extreme is that information is provided only to those who need it. Most operators function somewhere between these two extremes, but talking too much is discouraged and is addressed specifically in the training of operators. There is general agreement that too much talking is pollution of the air and a distraction from the important issues.

The subsequent two sections address (lack of) situation awareness as the other two explanations for differences in scores between *HE* operators and *IE* operators. Section 4.8.3.2.1 addresses difference in Level I situation awareness, perception of elements. Section 4.8.3.2.3 addresses differences in Level II situation awareness, understanding of the situation. The examples in these sections refer to situations where operators are responding to calls from the traffic.

4.8.3.2.1 Full information for everyone

A first explanation for the differences in items scored during the scenario is that good VTS operators always make sure that in every traffic situation all parties involved share the same information. This means that if a good operator makes an agreement with one of the ships, he checks whether other ships involved have also got the message. A lesser performing operator may think that this step is not necessary. He may be of the opinion that it is the responsibility of the ships to listen to all communication and therefore that ships have to pick up information that relates to them, without explicit action from the VTS operator.

An example of such a situation is in minute 25 of the scenario. A small ship, the HAVENDIENST 17 reports to the VTS that it is leaving the WERKHAVEN. It sees the coaster YEHYA on the river and decides to go Starboard-Starboard with her, a standard manoeuvre. Since there is no direct communication between ships and such communication is normally run via the VTS operator, the HAVENDIENST 17 informs the VTS operator of this action. There are two communication items possible for the VTS operator.

- First communication item is that the VTS operator contacts and informs the YEHYA about the manoeuvre of the HAVENDIENST 17;

- Second communication item is that the operator can provide information about this event to other traffic.

The information in the first communication item is considered necessary for good VTS. This item is labelled **A1**. The information in the second communication item is considered additional information, since these ships are not directly involved in the manoeuvre. This item is labelled **A2**.

Table 4-15Table shows how this event was scored by the *HE* and *IE* operators. Scores refer to the number of times that one of the assessors considered an item scored. Therefore the maximum score for an item is 15 (operators) times 2 (assessors) = 30. The *HE* operators found it more important than *IE* operators to inform the YEHYA of the upcoming manoeuvre. The majority of the *HE* operators found it unnecessary to mention the manoeuvre to the other ships. Fewer of the *IE* operators found it necessary to inform the YEHYA, while at the same time they did consider it important to inform the two other ships, the HOO PRIDE and the KOBULETI, about the event.

25	HAVENDIENST 17 leaving the WERKHAVEN, going SB/SB with coaster (YEHYA)	IE	HE	Total	Difference
A1	Info YEHYA	8	18	26	10
A2	Info HOO PRIDE and KOBULETI	9	3	12	-6

Table 4-15 Grouped PMI-P scores at 25'

From these data it cannot be concluded that the *IE* operators had a lack of situation awareness. There may have been a difference in the interpretation of the job responsibilities. The assessors had a clear preference as to the required communication in this situation. They thought that the YEHYA needed to have situation awareness of the arrival of the HAVENDIENST 17 while the HOO PRIDE and the KOBULETI did not need situation awareness of that issue, and the majority of *HE* operators agreed. *IE* operators thought differently. They considered the information important for the HOO PRIDE and the KOBULETI as well.

It cannot be concluded from these data that the *IE* operators had a lack of situation awareness level I compared to HE operators. Both *IE* operators and *HE* operators did demonstrate their awareness that they needed to inform the other traffic about the pending manoeuvre of the HAVENDIENST 17 to go Starboard-Starboard with the YEHYA at main fairway. Both *IE* operators and *HE* operators informed other traffic about this issue, and the two items about this issue were scored 38 times in total for both *IE* operators and *HE* operators (see Table 4-15). Yet the important difference in their responses is that *HE* operators communicated this with the ship involved in the manoeuvre, the YEHYA, while *IE* operators informed other traffic. This difference in scoring does reflect a difference in situation awareness level II; it displays a difference in the model of communication strategy used by *IE* operators: Who should be informed about certain issues and who should not (or not yet) be informed? The omission to inform the main ship involved in the manoeuvre is significant in the rating of their performance. Informing other traffic is not enough. The VTS operator should be aware that the master of the YEHYA may not have heard or understood the request of the HAVENDIENST 17 and therefore the operator has to check with him. This requires deep understanding of the situation and Level II situation awareness.

Time	Type	IE	HE	Time	Type	IE	HE	Time	Type	IE	HE
01:00	A1	31	44	12:00	A1	14	33	22:00	A1	16	15
	A2	4	7		A2	2	1		A2	1	4
02:30	A1	22	45	13:00	A1	6	8	23:00	A1	38	31
	A2	16	13		A2	0	2		A2	0	2
05:00	A1	22	26	15:00	A1	21	24	24:30	A1	6	11
	A2	1	5		A2	2	0		A2	13	14
09:00	A1	11	8	15:30	A1	2	7	25:00	A1	8	18
	A2	4	4		A2	0	0		A2	7	3
10:30	A1	11	10	17:00	A1	34	38	31:30	A1	22	28
	A2	14	11		A2	2	1		A2	2	1
0:11:00	A1	41	62	20:00	A1	17	24				
	A2	4	6		A2	18	19				

Table 4-16 Scores of IE and HE operators on events containing both A1 and A2 items (N=30)

A closer examination of the traffic scenario showed there were 17 events which contained both **A1** and **A2** items, shown in Table 4-16. Table 4-16 presents the combined scores for the A1 and A2 items per event. The scores of *IE* operators and *HE* operators on these events show that in more than a third of these events similar patterns to the ones analysed in Table 4-15 could be observed (shaded grey in Table 4-16). In these events HE operators scored higher on the (combined) **A1** events, while IE operators scored higher on the (combined) **A2** events.

4.8.3.2.2 Lack of Level I Situation Awareness

A second explanation for differences in responses may be that *IE* operators do not exactly know what information is relevant to the ship, but at the same time do feel a need to present some information to the ship when they are called. They may have only partial Level I situation awareness and therefore provide some, but not all relevant information. An example of this can be observed in the following event. The event was built into the scenario to investigate this particular issue.

At this point in the scenario upstream in the area an incident is drawing the attention of the operator. At the moment the operator is dealing with the incident, he has to retain situation awareness of the other parts of the situation. When a call comes from another part of the traffic area, the operator has to respond correctly. In the scenario an example of such a situation is in minute 26:30. The IMKE, an inland barge, reports shifting from one quay at the corner of a harbour to another quay in the same harbour. At that point there are two ships relevant for the IMKE, on which she needs to be informed:

- First communication item is that the operator informs IMKE about the WATERBOOT 12 in the same harbour, going towards the quay that IMKE is about to leave;
- Second communication item is that the operator informs IMKE about the incoming HAVLYS.

Both items are labelled **A1**, meaning they are necessary information for the IMKE.

Table 4-17 shows how the *HE* and *IE* operators scored in this event. Scores refer to the number of times that one of the assessors considered an item scored.

26.30	IMKE coming loose at the ESSO in the corner shifting over to the TOTAL	IE	HE	Total	Difference
A1	Waterboat (WATERBOOT 12) across the SHELL1 destination ESSO	5	18	23	13
A1	Incoming HAVLYS now across the entrance	5	13	18	8

Table 4-17 Grouped PMI-P scores at 26':30"(N=30)

Table 4-17 shows that *HE* operators scored more items than *IE* operators about this issue. This conclusion is supported when these data are examined closely in Table 4-18. presents a distribution of the operators that provide information to the IMKE. More than half of the *IE* operators do not provide any information at all to the IMKE, and none of these operators provide all information. The *HE* operators on the other hand provide much more of the necessary information.

	Number of Items scored		
	0	**1**	**2**
IE operators	8	7	0
HE operators	4	6	5

Table 4-18 Frequency of HE and IE operators scores on communication at 26':30"

The situation with the IMKE occurred in a remote part of the traffic area, displayed on the other screen, a harbour on the downstream side of the area, far away from where the main action was at that time[21]. When they were forced to provide information about the traffic situation outside the incident area *IE* operators did not know what to respond. The data suggest that *IE* operators were focussing on the incident in another part of the traffic area and were ignoring what was going on elsewhere. *HE* operators showed more awareness of what was going on in the more remote parts of the traffic area, and could respond to situations there by providing more accurate information. In terms of situation awareness this is a problem of Level I situation awareness (perception of elements). *IE* operators seem to lack situation awareness of important elements in the situation outside their immediate focus area.

4.8.3.2.3 Lack of Level II situation awareness

A third explanation for responses to the scenario that can be used to explain the differences between *HE* and *IE* operator performance on PMI-P relates to their different interpretation of a given event. *IE* operators may be aware of certain elements in the event but misinterpret their importance. This is a Level II situation awareness problem (comprehension). An example of such an event can be found early in the scenario.

At minute 2:30 in the scenario the DONGESTROOM, a 90 meters long barge, reports that she is leaving the quay at Shell-1 in the North Sea Harbour (NZH), going up the river. There are several issues relevant in this event.

- There is a outward bound barge 2500 meters above the entrance of the NZH that has had no communication with the VTS operator;
- There is also an outward bound coaster 3000 meters above the entrance of the NZH, going towards sea.
- There is also the KOBULETI, a 151 meters long seagoing vessel with two tugs that has reported its departure from another quay in the same harbour 1:30 minutes before.

The information on the barge and the coaster are labelled **A1**, for necessary information to be given to the DONGESTROOM. Information on the KOBULETI is considered as additional information (**A2**) at this moment, since the chance of interaction between the DONGESTROOM and the KOBULETI is low.

02.30	DONGESTROOM coming loose Shell-1 inward bound	IE	HE	Total	Difference
A1	Outward bound barge 2500 m. above the entrance NZH. No communication	8	21	29	13
A1	Outward bound coaster (YEHYA) 3000 m. above the entrance NZH. Destination Sea	13	22	35	9
A2	Information: Departing KOBULETI has left Chevron. Bound for Sea	12	3	15	9

Table 4-19 Grouped PMI-P scores at 2':30"

[21] This was an example of the use of the active probing technique in the scenario. Operators were invited to demonstrate what they knew about that particular part of the traffic area by the call of the IMKE. The scenario did provide operators enough time to respond, without jeopardizing their actions with regard to the incident.

Table 4-19 shows how this event was scored by the *HE* and *IE* operators. Scores refer to the number of times that one of the assessors considered an item scored (i.e. N=30 for each group).

In the answers of the *HE* operators it is clear that they understand the situation completely. They present the information on the barge and the coaster, but largely ignore the information on the KOBULETI. The results of the *IE* operators show a different picture. They provide less often the relevant information to the DONGESTROOM, but instead present much more often the unnecessary information on the KOBULETI. Apparently the *IE* operators still have the recent communication with the KOBULETI from the previous event in their situation awareness. Mentioning the KOBULETI seems a consequence of that communication, but demonstrates a lack of comprehension of the significance of that ship for the DONGESTROOM, and therefore demonstrates a lack of Level II situation awareness. This is completely in agreement with the findings described in the previous section and the results from Table 4-16 also show that this difference in performance takes place quite often during the scenario.

4.8.4 Discussion and conclusion

In this section the data from the FLO study were used to investigate if the scores of the operators on PMI-P could be studied on a detailed level to better understand VTS operator performance and VTS operator situation awareness.

4.8.4.1 Communication and performance

First a study was conducted to investigate if differences in scores on the behavioural rating scales between *HE* and *IE* operators could be simply explained by comparing the amount of communication.

The study showed that there are differences in the amount of communication between the two groups of operators. *HE* operators on average communicated significantly more items than *IE* operators. But these differences alone were not enough to explain the differences in scores between the two groups. There were operators in the *HE* group who scored fewer items than operators in the *IE* group. Therefore the differences in SARPR scores must reflect something other than the mere amount of communication items scored on PMI-P.

4.8.4.2 Situation awareness

Table 4-13 showed that overall *HE* operators scored significantly more items than *IE* operators. Tables 4-14 and 4-15 revealed that these higher scores of *HE* operators were obtained in **A1** (both first and other), and **W** items, while there were no significant differences in the scores on in the **A2** and **S** categories. The **A2** and **S** items are the least important items to report, and reporting them is not always useful, and may even distract operators from more important issues. Operators that score **A2** items, and at the same time do not score **A1** items, show a lack of situation awareness. When *IE* operators omit to report necessary **A1** items but instead report unnecessary **A2** information, they not only show their own lack of situation awareness, but also draw the attention of ships towards unnecessary issues (outside RSA), thus clouding the situation awareness aboard ships as well.

Further examination of the PMI-P data revealed that not only do *HE* operators score more items than *IE* operators, but that they also score different items within categories. These differences were used to study Hypothesis 8, to investigate if there are differences in situation awareness between *HE* and *IE* operators.
Three potential explanations for differences in responses were explored. The first explanation is that differences in scores are not only a result of lack in situation awareness, but that other factors play a role as well. Factors such as communication style and in the interpretation of the task of a VTS operator, or plain talkativeness, may have led to different performance in providing all relevant information to all ships. Other explanations for differences in performance were shortcomings of Level I and Level II situation awareness of the *IE* operators.

Differences in performance as a consequence of a different communication style were studied by looking at situations where this might have played a role. An example of such a situation is described in section 4.8.3.2.1. In the event described *HE* operators scored higher on **A1** item, while *IE* operators scored higher on an **A2** item in the same event. From that situation it cannot be concluded that the operators lacked situation awareness Level I, but on a deeper level *IE* operators did have a wrong understanding of the situation; in other words they did lack situation awareness Level II. Comparable situations were discovered in 6 out of 17 events that contained both **A1** and **A2** items. Therefore the conclusion can be drawn that this is a significant difference between *IE* operators and *HE* operators.

Differences in scores that were caused by a lack of situation awareness Level I could only be studied in one event in the scenario, where this issue had been built in deliberately. The demonstration for this issue was in an event where the VTS operator had to provide information about a remote part of the VTS area, while focussing elsewhere. *HE* operators showed more awareness of what happened outside the area of their immediate focus, and could respond with more accurate information.

HE operators also demonstrated superior Level II situation awareness by providing the correct information in other situations. They mentioned **A1** items in the event and did not mention the **A2** event. IE operators, on the other hand, mentioned **A2** items just as much as they mentioned **A1** items, while they were providing an almost equal amount of items in total. This phenomenon can be observed at various points throughout the scenario.

4.9 Overall discussion

The conclusion of the work with SATEST in Chapter 3 was that situation awareness was a useful concept in describing the work of the VTS operator, but that the method for assessing situation awareness needed more work. The *freezing* paradigm of SATEST could not account for the ongoing dynamics of the traffic, nor for the effect of these dynamics on situation awareness. SATEST could demonstrate the situation awareness of an operator at a certain moment in time, but could not be used to analyse how that situation awareness had developed. In the discussion of SATEST the idea grew that communication might be used to analyse when operators became aware of conflicts and when they would solve them (see section 3.5).

The current chapter describes the development and application of PMI-P, a performance based method for assessing VTS operator situation awareness. The method builds upon the results from the previous chapter, and can be regarded as an additional approach for assessing situation awareness. It uses the same testing environment but is based upon a different testing paradigm.

4.9.1 Reliability of PMI-P

4.9.1.1 Reliability of assessors

PMI-P data are collected by scoring communication performance on a predefined score form. In the study described in this chapter, the data were collected by two independent assessors, VTS trainers, who were qualified to examine VTS operators and trainees. The inter-assessor correlation on the PMI-P scores was consistent on the **A1**, **A2** and **S** data with correlations between .81 and .87 (see section 4.5.1). Agreement between assessors was lower on the PMI-P scores on **W** data (correlation .65, see section 4.5.1). The **W**-data referred to proactive calls to a ship based upon the observation of the traffic by the VTS operator, without request from the ship. Scoring these items required more interpretation of the communication by the assessors than scoring the other items. Hence, the PMI-P score-form can be improved to support more consistent scoring of **W** items.

Inter-assessor agreement was also studied for the scores on **E** and **L** data in section 4.5.2. The correlation between scores on **E** and **L** data is lower than those on the other data. The most important explanation for the differences in correspondence between the different types of data can be found in the procedure followed in the experiment. In the preparations of the experiment, the focus was on the scoring of the **A1**, **A2 W**, and **S** data. This was practised with the assessors. The scoring of data as earlier or later (**E** or **L**) was originally added to the PMI-P score form to validate the score form itself. If many items were scored at a different moment than the score form indicated, this could be considered as an indication that they might need to be moved on the score form. The procedural rule applied for **E** and **L** scores was that this measurement could be used for items communicated one minute or more, earlier or later than the scenario required. Only in the analyses afterwards was the value of these **E** and **L** data for the performance of operators recognised. With stricter procedures the reliability of the **E** and **L** scores may be improved.

4.9.1.2 Reliability of assessment procedure

4.9.2 Validity of PMI-P score form

PMI-P can be used as a tool to assess VTS operator performance. This application is mainly of interest for organisations that perform training and examination of VTS operators, such as MSR. Validation of the PMI-P score form in sections 4.6 has shown that there are several improvements possible for future use of the current scenario and PMI-P score form. These improvements consist of: Consequent use of item categories, the introduction of a new category of items for complete events, and splitting **A1** items into two categories (in this study called "**A1 First**" and **"A1 Other**"). The scoring of **E** and **L** items on the score form should be more standardised.

The split of **A1** items into two categories justifies the question if it is also necessary to split other items as well, such as **A2**, **W** and **S** categories. Adding more categories makes it possible to differentiate better between items and by that refine the scores of operators even more. At the same time analysis is more difficult if there are not enough items within a category. This study has given no indication that this is either necessary or desirable.

The PMI-P score form used in this study is a combination of the description of the scenario and a form on which communication is scored. As a consequence the form contains both instructions for assessor communication to the operator and items that can be scored along one timeline. Because of this the lay-out of the form is sometimes confusing. New items seem to appear on the form, and they are scored by assessors, as was shown in Table 4-2 in section 4.4. An improved lay-out of the score form may prevent confusion. In line with the previous issue, assessors often scored indications of time on the PMI-P score form, which contained descriptions of events in the scenario, as a single item (see section 4.4). This scoring indicates that the operator has handled the situation, without stating explicitly what the operator has said. One solution to this problem is to modify the PMI-P score form to facilitate this process by having an item category for "*Events*". This does not seem to be a very good option however, since it makes scoring unclear. The same score that is given for a complete event could also be given to a particular item on the score form or to several items within the same event. Therefore scoring complete events will only introduce more noise in the data. It is better to adapt the score form to make scoring of complete events impossible.
Annex 5 presents a lay-out for a new PMI-P score form, with the improvements mentioned here.

4.9.3 PMI-P and situation awareness

What PMI-P has revealed about situation awareness comes as much from the development of the scenario and the PMI-P score form as it comes from the assessment itself.

The PMI-P score form captures the development of VTS operator situation awareness by describing all possible communications that an operator can have during the scenario run. The development of this form was conducted through elaborate discussions. For each separate event in the scenario all possible communications had to be determined. During this process the assessors themselves played the role of operators in the simulation to study the scenario and communication. The result of this development is the PMI-P score form used in the FLO study.

As assessment form for communication the PMI-P score form provides an excellent overview of what the assessors expect the situation awareness of VTS operators to be.

Section 4.8.3.2.2 describes how differences could be observed between *Highly Effective* and *Ineffective* operators in situation awareness Level I. The VTS simulator used two screens to display the traffic area. The operators had to be aware of the traffic in both parts of the area, and could not just focus on one part completely. *Highly Effective* operators were more capable of doing this.

Section 4.8.3.2.3 also studied differences in Level II situation awareness observed between *Highly Effective* and *Ineffective* operators. The PMI-P score form differentiates between situation awareness that operators need to have (**A1**) and situation awareness that is additional (**A2**). There is a clear hierarchy between the **A1** and **A2** items. **A1** items are RSA and need to be communicated. **A2** items represent task-related information that is not RSA. The **A2** items present additional information that clarifies the situation (see section 4.2.1) but may also distract from it and harm situation awareness of the ships captains. 46% of the items that are scored by *Ineffective* operators, but not scored by *Highly Effective* operators, are **A2** items. This illustrates that *Ineffective* operators understanding of the situation, their situation awareness Level II is less. When operators communicate **A2** items and at the same time forget to communicate **A1** items they are demonstrating a lack of situation awareness Level I or II. They either have missed important information (situation awareness Level I, perception) or they have misinterpreted the situation (situation awareness Level II, comprehension).

4.9.4 Evaluation against selection criteria

Chapter 2 defined criteria for selecting a method for VTS operator situation awareness assessment. This section will assess PMI-P on these criteria. The following criteria were mentioned:

1. Direct assessment of situation awareness
2. Method provides as much information on situation awareness as possible
3. On-line methods are preferred
4. Obtrusiveness of method has to be tested
5. Method should be suited for experiments

ad 1. PMI-P is not a method for the direct assessment of situation awareness. Instead the method assesses operator performance on communication. Analysis of this communication provides an indirect assessment of situation awareness. Communication shows the situation awareness of the operator on different levels. First an operator can only communicate issues of which he is aware (situation awareness Level I). Next the operator has to understand the situation to know what needs to be communicated (situation awareness Level II). The operator demonstrates Level II situation awareness in presenting the **A1** items first and **A2** items only when there are no more **A1** items that need to be communicated. Third when the operator understands the situation, he can predict what will happen (situation awareness Level III) and anticipate on that by providing information that ships need in their future. Detailed analysis of the scenario and operator scores such as the ones in section 4.8.3.2 can reveal information on all levels of situation awareness in operator performance on PMI-P.

ad 2. PMI-P measures VTS operator performance. The underlying assumption is that an operator cannot do anything without situation awareness of the traffic situation. Detailed analysis of the PMI-P score form and the scores of operators on PMI-P show how successful the method has been in linking operator performance to situation awareness. PMI-P focuses upon the time dimension of VTS operator situation awareness, as was stated as one of the objectives in section 4.1. To achieve this, the development of situation awareness throughout the scenario and the assessment of this development have played an important role in setting up the scenario and the events in that scenario. Detailed analyses of the results show that situation awareness at different levels can be inferred from the data. It is possible to follow the development of situation awareness over time throughout the scenario, by studying when items are communicated. Besides the scores on items on the PMI-P score form, data are collected about the timing of communication. Section 4.8.3.2.3 described how items scored *Early* or *Late* provided understanding of VTS operator Level III situation awareness.

ad 3. PMI-P is an on-line method.

ad 4. Solving the problem of obtrusiveness was one of the specific problems of SATEST that needed to be solved in PMI-P, and was done successfully. The obtrusiveness of PMI-P is not a problem, since operators do not notice that they are being assessed.

ad 5. PMI-P requires a significant effort in developing the PMI-P score form. Therefore the method is suited for larger experiments, but not so much for smaller tests.

Evaluation of PMI-P on the criteria from Chapter 2 illustrate that the application of criteria had somewhat shifted from the general assessment of situation awareness towards the specific time dimension in situation awareness. Further the methodology has shifted from a method for assessing VTS operator situation awareness towards a method that assesses VTS operator performance, taking into account situation awareness.

4.10 Conclusion

PMI-P is a powerful tool for assessing VTS operator performance. The method applies assessment of VTS operator communication to assess performance. Communication is scored on a PMI-P score form. This score form has been constructed using the concept of Required Situation Awareness about events in the scenario. For every event in the scenario the PMI-P score form contains information that the operator needs to communicate.

Differences in performance on PMI-P between operators can only partly be explained by differences in communication skills (section 4.8.3.2.1). The operator has to be aware of the traffic situation. He needs to know where the ships are and where they are going (situation awareness Level I), he needs to understand the implications for the routes of the ships and the conflicts that ships may have (situation awareness Level II), and he needs to be able to predict what will be the situation in a few minutes time or later (situation awareness Level III). Only an operator who is aware of the whole situation can provide the information that ships need to travel safely and efficiently through the VTS area. The difference between operators who perform well (HE) and operators who perform less (IE) is a reflection of their situation awareness as well as their communication skills. Without situation awareness operators cannot communicate effectively with the traffic.

Many experts in the field of VTS have participated in the development of this method. Both assessors and VTS operators have discussed the methods and the results. An external researcher who does not have access to these people, or does not want to engage in a dialogue with them, will never be able to develop a useful method in this domain. This has consequences that go beyond this particular thesis and will be discussed in chapter 5.

Understanding communication in PMI-P provides a solid handle on the assessment of VTS operator performance. This leads us to the question whether situation awareness is still needed as a concept in the performance of assessment. The short answer to this question is that in the development of PMI-P the required situation awareness has been used to understand the scenario and the required communication. The results from this study have not been fully adopted by the VTS field, but discussions I had recently indicate that there is interest in the field to further develop the results for training purposes. Section 5.3.3 in the next chapter will elaborate on this discussion.

CHAPTER 5

Discussion and Conclusion

The main objective of this study was to investigate what the assessment of situation awareness contributes to the understanding of VTS operator performance. To answer this question the concept of situation awareness and methods for assessing situation awareness were studied in chapter 2. Next two methods for assessing situation awareness of VTS operators were developed and tested in chapters 3 and 4. This chapter closes the circle with a discussion of what has been done and conclusions that can be drawn from this work. Section 5.1 discusses the concept of situation awareness and the relevance for VTS work. Section 5.2 discusses reliability, validity and usefulness of instruments that were developed to assess VTS operator situation awareness. Section 5.3 indicates directions for further research.

5.1 Situation awareness and VTS

5.1.1 Relevance of situation awareness for VTS work

Section 2.4 described how the importance of situation awareness for the work of the VTS operator has become recognised by maritime world in the past decade. In the second (1998) edition of the IALA VTS manual situation awareness was not mentioned. The third (2002) edition of the IALA VTS manual states that one of the considerations to determine a period of duty for a VTS operator/supervisor should be "the ability to develop and maintain situational awareness" (p.67). So not only has the concept been adopted, there also is understanding that it takes time to build situation awareness. Other IALA documents refer to the importance of tools to enhance situation awareness (IALA, 2002b) or tests that need to be designed to determine the ability of candidates for VTS to demonstrate spatial and situational awareness (IALA, 1998).

The relevance of situation awareness for VTS operators lies in the very nature of the work: operators have to maintain an overview of the traffic and have to provide navigational assistance to ships, which as stated in IMO Resolution A.857 (20) is especially important in difficult navigational or meteorological circumstances or in case of defects or deficiencies.

To be able to provide this assistance operators need situation awareness on all three levels as defined by Endsley, described in Section 2.2: They need to perceive the elements (Level I) and comprehend their meaning (Level II) and also they have to be able to predict the future (Level III). These levels were translated in terms of VTS tasks in section 2.4.2. In short: Perceiving the elements in VTS means that operators need to keep track of location, course and speed of all ships and all other objects in their designated traffic area. Comprehension of the situation means that the operator has to know which ships may have a conflict, either with other ships (risk of collision) or with the infrastructure (risk of stranding). Predicting refers to future position, course and heading of ships, and the consequences of those, such as conflicts that may develop. Having good situation awareness on all three levels is essential for the work of VTS operators.

5.1.2 Temporal aspect of situation awareness

Even though the importance of situation awareness for the VTS work has been very clear from the start of this study, the temporal nature of the concept has raised a few questions. This issue has played throughout this study.

Situation awareness is something that has to be maintained over a long period of time for instance a work shift. During this period the situation changes constantly and subsequently the mental model of the situation has to be constantly updated with new information becoming available. This temporal aspect of situation awareness was recognised early in the project and the SATEST experiments were set up with this in mind. Three breaks were selected in section 3.3.3 that would allow the researchers to study how situation awareness of VTS operators developed over the scenario. The results of this study were disappointing. Section 3.5.3 points out that the "*freeze*" paradigm applied in SATEST does not work well in the VTS domain with regard to studying this issue. The slow development of maritime traffic allows for the operator to employ different strategies in approaching ships and handling conflicts that can all be equally good, but lead to different assessment of situation awareness.

The second method for assessing VTS operator situation awareness (PMI-P) was developed in chapter 4 with the results from this work in mind. The assessment in PMI-P takes the different strategies into account that VTS operators employ in dealing with the traffic. Communications of operators with ships form the basis of this assessment. Timing of this communications provides an indication of temporal aspects of the operators' situation awareness. In this respect PMI-P addresses the temporal aspect of situation awareness better than SATEST.

5.1.3 Relevance and usability of RSA

Section 2.3.1 describes the debate on the term *situation* in situation awareness. To get a grip on the situation assessed in VTS, the concept of Required Situation Awareness (RSA) was introduced in section 4.1. The term Required Situation Awareness (RSA) is used to describe the task environment and in particular the scenario used in testing in terms of the situation awareness that is necessary to correctly handle the situation. This RSA has been operationalised in the PMI-P score form in Section 4.2.1. The description has been compared to actual VTS operator performance in Section 4.6 and beyond.

A small study was conducted to develop a quantitative model of RSA (Wiersma, 2000). According to this model the RSA for the situation at any moment in time (*t*) may be calculated by adding together the RSA of each event RSA (E) active at that particular moment. Thus $RSA(t) = \Sigma_1^n RSA\ E_n(t)$. With this model it was possible to describe RSA over the whole length of a scenario. However, this process was very time consuming and the expected outcomes were only relevant for that particular scenario. The method provided information about where in a scenario the RSA was high. However this information could also be obtained by counting the amount of items (ships) active at a certain moment, an analysis that was much easier.

In publications on situation awareness research studies, there usually is no clear indication if and how researchers have established an image of RSA. Questions asked are reported, but not a total representation of RSA development throughout an experiment or scenario. Also as a tool for communication RSA has its use in the development of scenarios and in the tools for analysis of scores based upon these scenarios. This makes the concept useful only in assessment contexts that make use of a scenario for testing situation awareness. For these scenarios it can be discussed which items or issues are relevant at each particular moment in the scenario or even if all elements are equally important as Endsley stated in section 3.5.3. During the Rotterdam trials operators that they did not regard reproduction of the complete traffic situation as important for evaluating a situation. Getting the most important elements right was considered more important than getting all the elements (see section 3.4.2.3.2).

5.2 Reliability, validity and usefulness of instruments

The reliability and internal validity of the methods developed in this thesis were discussed in the chapters that presented these methods (for SATEST see Sections 3.4.2.3, 3.4.3.4; and 3.5; for PMI-P see Sections 4.5, 4.6, 4.8.4 and 4.9). This section will wrap up the discussion by once more reviewing the methods and addressing the issues that have remained for discussion.

The first method developed, SATEST used the Freeze paradigm to test situation awareness, like many other methods for assessing situation awareness. Despite the small number of operators whose data could be used in the analyses, several methods for analysis were developed, both qualitative and quantitative.

A qualitative analysis based upon ranking of the score forms was performed for the data in Rotterdam. This ranking was performed on the basis of performance on the scoring of a break, without proper criteria for what this good performance consisted of. Even then the ranking, which was repeated a month later, yielded fairly consistent results, as was shown in section 3.4.2.2.8. Ranking filled in score forms can provide good understanding of the performance of operators on individual breaks. Ranking of operator scores between breaks was not possible during the Rotterdam experiments. There was no tool for comparison of different traffic situations. An objective evaluation of breaks can be obtained by the use of the RSA concept, introduced in Chapter 4, section 4.1.2. With RSA criteria can be developed that can support the understanding of situations and that can facilitate comparison of different breaks.

A quantitative analysis was performed on the data of both Rotterdam and Helsinki. The results in Rotterdam were merely scored for the number of ships reported, which is a too coarse method of analysis. In Helsinki more complex modes of analysis were tried, but due to the small number of operators the experiments did not yield statistically testable results. The analyses did demonstrate interesting phenomena that need to be taken into account in the further development of the method. Often operators did not report a conflict in situations, where there was a conflict present. This underreporting could be caused by two mechanisms. First it could be that operators had missed the conflict, and not scoring the conflict was a correct indication of failure to recognise the situation by the operator. Other operators did recognise the conflict, acted upon it and thus solved the conflict. These operators did not regard the situation as a conflict any more, but there situation awareness of the situation and performance were excellent. In the scoring however, this was not reflected, since they too did not score a conflict. Training of operators in the filling in the score form may solve this problem. To improve quantitative analysis requires the scenario to be analysed using RSA.

The improvements of the SATEST method described so far require more complex analysis for better understanding and objective interpretation of scores. Parallel to that another potential use of the method is in training. During training sessions the SATEST method can be applied in simplified form. The situation can be sketched on paper and the filled in results can be compared the actual situation on the screen. Differences between the two can be discussed immediately and the trainer can indicate important aspects that the trainee might have missed. This simple application of the method was considered very useful when this work was presented on conferences, but it has not become part of the official VTS operator training methodology (see section 3.5.4).

The second method developed, PMI-P used communication performance of operators to test situation awareness, an indirect assessment. All Rotterdam VTS operators were tested in this study. The study showed that PMI-P was a good method to assess operator communication. The relation between the assessment of communication and operator situation awareness was not always clear, but detailed analysis of the results showed that they can be used for that.

The scenario used in the Rotterdam PMI-P study cannot be used again to test operators in Rotterdam. Operators remember test scenarios, as was already shown during the SATEST trials (section 3.4.2.3.3). The scenario used in the PMI-P study had an accident and therefore is even more memorable. For other operators this scenario offers a very good test situation. The results of the Rotterdam operators provide an excellent base line measurement of performance.

Figure 2-7 on page 31 presented the schematic overview of situation awareness measures that was developed in Section 2.5.2. Figure 5-1 presents an adapted figure that shows the measures for which assessment methodologies were developed in this thesis. SATEST, described in Chapter 3 uses recollections from memory, PMI-P provides information on operator performance. The model in Figure 5-2 provides a good handle for a discussion on these methods.

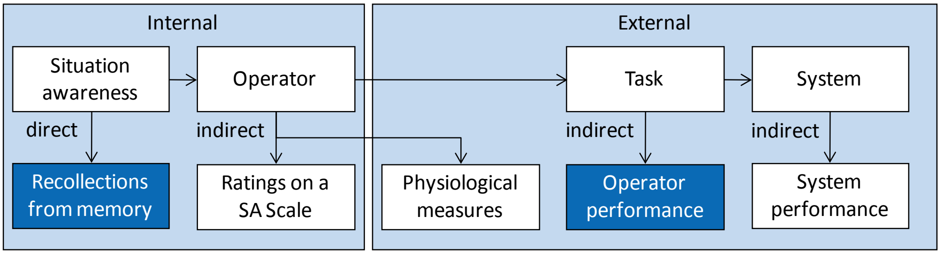


Figure 5-1 Situation awareness measures used to develop assessment methodologies

The relation between situation awareness and the measures is modelled in Figure 5-1 in two steps: The number of links (arrows in Figure 5-1) between situation awareness and the measure, and the quality of each link. More links means more transformations, each indicating loss of information. The quality of the link refers to how much information on situation awareness is contained within each link. Based upon this model SATEST provides more direct information on situation awareness than PMI-P. At the same time measures to the right in Figure 5-1 provide more information on a global level (tasks, system).
This distinction is clearly reflected in the analyses in chapter 3 and 4: SATEST analysis provides detailed information on situation awareness, that needs to be translated to the task level, PMI-P provides information on the task performance level, that needs to be interpreted in terms of situation awareness. A choice of measure therefore depends on the position of the researcher and on the specific research question.
The work on SATEST has shown that it is fairly simple to assess situation awareness in a VTS context. The method can be applied in a simple form that allows direct discussion on the results. At the same time, it is complex to use this method as an experimental tool that provides valid results in practise (see Section 3.5.4). Testing of operators requires considerable numbers of subjects and time, because scenarios develop slowly and can only be used once (because operators remember them). Comparison between different scenarios or even different moments within scenarios is difficult. The conclusion is that SATEST is an excellent tool for training purposes, when operators can be confronted with their own performance on SATEST, but that its use for experimental scientific research is limited. It can be used to test the effect of alternative designs of HMIs on VTS operator

situation awareness, but the number of alternatives under investigation should be narrowed down to a few.

PMI-P uses situation awareness in a more implicit manner. The method has been used successfully in establishing a basic performance assessment of VTS operator performance. The study in chapter 4 shows that PMI-P provides more consistent results than rating scales currently in use in assessing VTS operator performance (Section 4.5.3). Because of this consistent, objective manner of scoring performance, the method would fit well in training and certification of VTS operators.

Both methods have in common several elements: they make use of a simulator to present a traffic scenario, they use experts in the development of scenarios in specifying the and RSA in them and they use experts in the evaluation of the results.

In the experiments in this thesis the simulator of MSR in Rotterdam was mostly used. Only in the Helsinki experiment with SATEST was another simulator developed. In the different experiments operators were asked if they considered the simulator a fair representation of their actual traffic. The answer to this question was always positive. Apparently simulators can present a traffic image that resembles reality enough to perform tests as have been done in this thesis.

The use of experts for both developing and evaluating tests throughout this study requires caution. Experts are experienced operators and trainers of operators. In the assessment of VTS operator situation awareness their ideas and philosophies of what good performance is are leading. Operators are assessed according to the standards set by these experts, and performance that resembles the approach of these experts is considered good. However, different approaches are possible in addressing traffic. The proactive approach customary in the Rotterdam area was not very much appreciated in Helsinki. Operators in Finland were expected to take a more reactive attitude. Only when asked by the ships did operators provide information. There are several reasons why these differences exist: They can be based in the regional implementation of the legal framework for the VTS work, or can be related to national or regional cultures, which may impose a more active or reactive approach. Even the origins of the regional VTS, stemming from lighthouse authorities or from coast guard, may result in a more passive or more active approach towards interacting with the traffic.

As a result different strategies for communicating with traffic can be observed in different areas. It is therefore necessary to take the regional context into account formulating "correct" communication behaviour.

At present, in the absence of an external criterion for VTS operator performance, the use of experts is the only way to go forward. However, the results of the study have to be externally validated. One option for this might be to adapt incident analysis to incorporate **A1** and **A2** items in communication analysis.

5.3 Directions for further research

5.3.1 Further development of SATEST and PMI-P

The next step in the development of SATEST can go in two directions: more elaborate analysis or simplification. As stated in section 5.2, improvement of the analysis of SATEST requires further development of the RSA concept. If RSA is used to evaluate the scenario, both criteria for ranking can be developed and the results can be used for quantitative scoring.

The next step in the simplification of SATEST is to develop the method as a smaller scale research tool. First steps in this direction have already been taken in the EMBARC project. In this project the introduction of AIS[22] in VTS was studied. In a test bed a small portable VTS simulator was developed that could show traffic with and without AIS information. In a number of scenarios the effects of AIS on the traffic image were demonstrated. After each scenario operators were requested to fill in a simplified SATEST score form. The results were compared to the actual traffic situation and the differences were used to discuss AIS for VTS. The demonstrations were done in various ports around Europe (Hoebee, et al. 2005, Wiersma, & Padje, 2005).

For experiments with a new scenario some improvements can be made in the PMI-P method, especially in the categories of items, in the presentation of the score form and in the relation between communication performance and situation awareness.

PMI-P used four categories of items to score communication (section 4.2.1). **A1** and **A2** items were communication items referring to necessary (**A1**) or additional (**A2**) information, presented after a call from a ship. **S** referred to the standard procedure call, while **W** items referred to communication initiated by the operator without a call from the traffic. In the in-depth analyses (section 4.6.4.1) the **A1** items were split into two categories **A1first** and **A1other**. The first A1 item in an event was usually the most important one to mention. These items were mentioned most. In future scenarios it should be considered to use these three categories of **A** items. Splitting of other categories, such as **W** items does not seem to be necessary or useful.

The PMI-P score form use in Rotterdam was a combination of the scenario and the form on which responses should be noted. The data needed to be cleaned up before they could be analysed in section 4.4. Annex 5 presents a score form with better visual support for scoring.

Section 4.8 presented the relation between VTS operator performance and situation awareness. This relation was not used explicitly *a priori* in setting up the scenario, but was implicit in the development of the scenario and the communication. Even though the situation awareness, necessary to guide the traffic through the scenario, was discussed during the development of the scenario, it was not made an explicit part of the PMI-P

[22] The **Automatic Identification System** (**AIS**) is a short range coastal tracking system used on ships and by Vessel Traffic Services (VTS) for identifying and locating vessels by electronically exchanging data with other nearby ships and VTS stations. Information such as unique identification, position, course, and speed can be displayed on a screen or an ECDIS (http://en.wikipedia.org/wiki/Automatic_Identification_System, visited 26-08-2009)

score form. If the objective of further study is increase understanding of situation awareness in VTS performance, it is recommended to make this explicit in the description of the scenario and link it to specific items on the scores form in the manner discussed in section 4.8.3.

5.3.2 From presentation of information to presentation of reliability in VTS displays

Another issue that requires further investigation concerns the presentation of information on VTS displays. Information presentation was the issue that started this research a decade ago. At present information presentation still is an issue for further investigation, but the focus of the debate has changed.

Modern VTS stations are equipped with displays that present different types of data on ships on single monitors. Data on ships that are this information can be displayed separately or integrated. In the first case information is presented in layers, and operators can select the layers that they want to see. In the second case algorithms are used to process raw data and the results of these calculations are presented on the screen.

Both of these approaches are adopted by designers of VTS displays and preferences for each of the approaches can be found when discussing the equipment with operators of different ports or different nationalities.

Both the approaches have advantages and disadvantages. They are the result of the fact that the data that are collected are often incomplete, and contain errors or noise. These errors and noise can result from false radar reflection or bad radar coverage, as well as from unreliable AIS data, as a result of misplaced antennae or incorrectly filled in data sheets. As a result the data presented on the traffic displays are sometimes blurred, the dimensions of ships are sometimes displayed incorrectly, at other times ships disappear from the screen completely or are displayed at locations where they are not.

Presenting the raw data in layers and leaving the interpretation of the data to the operator has advantages. Operators usually know for their traffic area where radar reflections change the radar image or where radar coverage is bad. With experience operators learn how to interpret the data anomalies, which become part of situation awareness. Moreover when data from different sources are conflicting, both are displayed, indicating to the operator that closer attention is required. The disadvantage of this approach is that interpretation of the raw data is an extra task for the operator that requires attention. Another disadvantage is that all this information may clog up the screen.

Presenting processed information on the traffic displays solves the disadvantages mentioned in the previous paragraph. The screen looks cleaner and the operator can direct his attention to the traffic. The disadvantage of presenting processed data is that the operator does not know how the results have been achieved or how certain the presented locations are. If the algorithms make a mistake in selecting the position of a ship when radar and AIS data disagree, the ship may be positioned in the display in an incorrect location or with incorrect dimensions.

An operator has to be constantly aware of how reliable the information is that is presented on his displays. Research is needed on the question how this dimension of reliability of information should be displayed on a VTS monitor. Assessment of VTS operator situation awareness, and then especially how an operator keeps track of not only the data presented but also on the quality of the data can provide valuable understanding that can help the designer present the information in an optimal manner.

The SATEST method described in Chapter 3 presents the best approach for such an assessment. Different HMI lay-outs can be tested with quality of information presented in different ways. In the SATEST score form queries can be added to indicate the quality of data that are reproduced. Operators can run one or more scenarios with the different displays Wiersma, 2004).

In the assessment a choice can be made between a study in the form of a full experiment, which would require a large number of operators and a series of tests, or a study on a smaller scale, where a limited number of operators are tested on the same scenario with different HMI displays, followed by a discussion with them on the advantages or disadvantages of the different lay-outs. The second method is more likely to provide quick answers and preferences, while the first method provides scientific and more objective information against higher costs.

5.3.3 Situation awareness and training of VTS operators

Situation awareness can be trained effectively, as has been shown in studies in several domains (Endsley, 2000c; Vincenzi, Hays & Seamon 2000; Saus et al., 2006). Even though the importance of situation awareness for the work of VTS operators is well recognised, this does not mean that it is generally trained or used as a training tool in the field.

While rounding of the work for this thesis in the Fall of 2009 I had discussions with representatives of the NNVO, the Dutch VTS Training Foundation[23], and MSCN, Maritime Simulation Centre Netherlands[24]. In these discussions it be clear that, although the concept of situation awareness is well known and its importance for VTS work is understood, it is at present not part of the VTS training program. At present (December 2009) in the Netherlands situation awareness of VTS operators is neither trained explicitly, nor is situation awareness explicitly used as a tool to support VTS operator training.

The persons I was speaking to were not familiar with the work that had been done described in this thesis, or the publications about the work. In the past years I had not taken time to present this work to the field of VTS operator training and apparently so had no one else. The interest in the application of the work described in this thesis has now

[23] The Nationale Nautische Verkeersdienst Opleiding (NNVO) is the registered institute in The Netherlands to facilitate for the VTS training of Dutch candidate and certified VTS operators.

[24] MSCN is part of the Maritime Research Institute Netherlands (MARIN) and is the other official organisation besides MSR Rotterdam that provides training to VTS operators.

been aroused the publication of this thesis may provide a starting point to translate the work into training practise.

5.3.4 Epilogue from situation awareness to security awareness: a new challenge

This thesis has confined itself to the study of situation awareness in the domain of maritime traffic control. However, hundreds of researchers have shown that the concept can be applied in many domains. Of special interest for a Safety Science Group is the application of situation awareness in relation to safety and security.

In recent years I have been involved in the domain of security, setting up and delivering academic education programs in that field. The security domain has a very practical in orientation and generally holds the credo "*safety is a science, security is an art*". At present there is not much basic and scientific research going on in the field of security. The education programs aims to filling this gap.

One of the issues that has come up frequently in discussions about security is the issue of security awareness. Security awareness can be defined as:

> *The degree to which an individual (employee, client, visitor, etc.) is consciously aware of the relation between his actions and his (personal or organisational) security.*

Security professionals often refer to three characteristics of security awareness: knowledge, attitude and behaviour (Thomson & Solms, 1998). The issue is best studied in the information security domain, even though the relevance to other security domains is obvious. Security awareness training programs are provided frequently without a clear definition of what security awareness is and without a method to assess the effect of the training programs on the security awareness (Malskat, 2009).

The concept of situation awareness provides useful understanding for the security field. Situation awareness differs from security awareness in the respect that situation awareness refers to the actual understanding of the task environment, while security awareness is often regarded to have a more permanent character, transferring over different tasks. Yet at the same time, security awareness can be regarded a specific type of situation awareness, with respect to a certain aspect or *sub-task*: maintaining personal and organisational security while performing one's main job.

One thesis project in the MoSSM[25] program studied security awareness using the framework provided in chapter two of this thesis (Malskat, 2009). The research questions were how security awareness can be measured and how the results of these measurements can be used to control security awareness in the organisation of the course member.

In the case of the thesis project a security training was provided to personnel in the organisation of the student. Objective of the training was to make people aware that the human being is both the strongest and the weakest link in all security issues.

[25] Master of Security Science and Management, TopTech TUDelft

In the discussions about the assessment of the effect of the training, especially Figure 2-7 (page 31 of this thesis) proved useful in the development of a methodology for assessing security awareness. This figure illustrates how different measures of situation awareness (security awareness) relate to the awareness. The further (more steps) away the measure is from situation awareness, the weaker the link between the measure and situation awareness. The different measures provide handles in the environment that the researcher can take hold of. The appropriate measure depends on the role and position of the researcher and on the type and amount of information that the researcher can obtain.

In the thesis project the effect of this training was assessed using security incident reports. This was an appropriate measure in terms of collection, considering the staff position of the student. Incident reports can be seen as a system performance measure. According to Figure 2-7 the relationship between the reports and situation awareness is weak. However, in the definition of security awareness used in that study filing a report is regarded as an act that requires security awareness, and the reporting rate increased after security awareness training, therefore security awareness increased. The incident reports were considered indirect but adequate measure of security awareness in the context of the study. Closer examination of the incident reports may reveal whether the reports themselves contain information regarding the security awareness at the time of the incident. The description of the situation in the reports may provide an indication whether the incident was a result of lack of security awareness of an employee. This does not only provide information on how often security breaches take place, but it may also provide valuable information on the types of scenarios that lead to security incidents. Such a study moves the analysis further left in Figure 2-7, towards the task and the operator performance. This line of research was proposed as further line of inquiry in the study of Malskat.

The study of Malskat proves that the model developed in Chapter 2 on situation awareness measures is useful in the discussion on the development of measures, and that the approach towards assessing situation awareness presented in this thesis can be used to study security awareness. Further development in this direction is an interesting line of future research.

References

Adam, E.C. (1994). Tactical Cockpits? The coming revolution. In Gilson, R.D., Garland, D.J. & Koonce, J.M. (Eds.), *Situation awareness in complex systems.* (pp. 101-110). Daytona Beach: Embry-Riddle Aeronautical University Press.

Adams, M.J., Tenney, Y.J., & Pew, R.W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*(1), 85-104.

Banbury, S., & Tremblay, S. (Eds.). (2004). *A cognitive approach to situation awareness: theory and application.* Hampshire: Ashgate.

Bisantz, A.M., & Vicente, K.J. (1994). Making the abstraction hierarchy concrete. *International Journal of Human-Computer Studies, 40*(1), 83-117.

Bolstad, C.A. (1991). Individual pilot differences related to situation awareness *Proceedings of the Human Factors Society 35th annual meeting* (Vol. 35, pp. 52-56): HFS.

Bolstad, C.A., & Hess, T.M. (1996). Situation awareness and older workers. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 289-294). Daytona Beach: Embry-Riddle Aeronautical University Press.

Butter, R., Regelink, H., Baldauf, M., Wiersma, J.W.F., Granholm, G., & Coles, P. (1998). *COMFORTABLE Work Package 02 "Recognition, Risk, Evaluation & Assessment".* Rotterdam.

Butter, R. & Wiersma, J.W.F. (2002). Development of a Perfomance Monitoring Instrument for Vessel Traffic Service Operators. In: Waard, D. de, Brookhuis, K. A., Moraal, J. & Toffetti, A. (Eds.), Human Factors in Transportation, Communication, health and the Workplace (pp. 387-396). Nederland: Shaker Publishing BV.

Byrne, E.A. (1996). Role of volitional effort in the application of psychophysiological measures to situation awareness. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 147-154). Daytona Beach: Embry-Riddle Aeronautical University Press.

Carmody, M.A. (1993). *Task-dependent effects of automation: The role of internal models in performance, workload, and situation awareness in a semi-automated cockpit.* Unpublished Dissertation, UMI Dissertation Service 9416605, Texas Tech University.

CEC (1996). *Transport research EURET Maritime transport: Tools to access VTS and to increase the efficiency of VTS (TAIE)* Rotterdam.

Charness, N. (1996). Expert performance and situation awareness. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 35-42). Daytona Beach: Embry-Riddle Aeronautical University Press.

Chubb, G.P. (1996). Displays to enhance air combat situational awareness. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 345-350). Daytona Beach: Embry-Riddle Aeronautical University Press.

Companion, M.A. (1994). Situational awareness in emergency management systems: An overview. In Gilson, R.D., Garland, D.J. & Koonce, J.M. (Eds.), *Situation awareness in complex systems.* (pp. 283-289). Daytona Beach: Embry-Riddle Aeronautical University Press.

COST 301 (1987) *Final Report: Annexe to the main report Volume 1: Concepts and definitions.* Camberley, UK: Easams Ltd. On behalf of the Commission of the European Communities.

Croft, D.G., Banbury, S., Butler, L.T., & Berry, D.C. (2004). The role of awareness in situation awareness. In Banbury, S. & Tremblay, S. (Eds.), *A cognitive approach to situation awareness: theory and application* (pp. 82-103).

Dennehy, K.E. (1996). *The development of a rating scale to measure the situation awareness of student civil pilots.* Unpublished Dissertation, UMI Dissertation Service 9709092, Cranfield University.

Dennehy, K.E. (1997). *Cranfield situation awareness scale users manual* (Report No. 9702). Cranfield: College of Aeronautics.

Dominguez, C. (1994). Can SA be defined? In Vidulich, M., Dominguez, C., Vogel, E. & McMillan, G. (Eds.), *Papers and annotated bibliography. Tech. Rep. No. AL/CF-TR-1994-0085* (pp. 5-15). Wright Patterson Air Force Base, OH: Armstrong Laboratory.

Durso, F.T., Truitt, T.R., Hackworth, C.A., Crutchfield, J.M., Nikolic, D., Moertl, P.M., et al. (1996). Expertise and chess: A pilot study comparing situation awareness Methodologies. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 295-304). Daytona Beach: Embry-Riddle Aeronautical University Press.

East, K.P. (1996). *Workload, performance and the development of situation awareness: An information processing-based approach.* Unpublished Thesis, UMI Dissertation Service ADGMM09674, University of Guelph.

Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement *Proceedings of the Human Factors Society 32th annual meeting* (Vol. 32, pp. 97-101): HFS.

Endsley, M.R. (1990). *Situation awareness in dynamic human decision making: Theory and measurement.* Unpublished Dissertation, University of Southern California.

Endsley, M.R. (1991). Situation awareness in dynamic systems *International Ergonomics Association Proceedings 1991* (pp. 801-803): IEA.

Endsley, M.R. (1994). Situational awareness: some reflections and comments. In Gilson, R.D., Garland, D.J. & Koonce, J.M. (Eds.), *Situation awareness in complex systems.* (pp. 315-317). Daytona Beach: Embry-Riddle Aeronautical University Press.

Endsley, M.R. (1995a). Measurement of Situation Awareness in Dynamic-Systems. *Human Factors, 37*(1), 65-84.

Endsley, M.R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Endsley, M.R. (1996). Direct measurement of situation awareness in simulations of dynamic systems: Validity and use of SAGAT. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 107-114). Daytona Beach: Embry-Riddle Aeronautical University Press.

Endsley, M.R. (2000a). Theoretical underpinnings of situation awareness: A critical review. In Endsley, M.R. & Garland, D.J. (Eds.), *Situation awareness analysis and measurement* (pp. 3-32). MahMaw, NJ: LEA.

Endsley, M.R. (2000b). Direct measurement of situation awareness: Validity and use of SAGAT. In Endsley, M.R., & Garland, D.J. (Eds.). (2000). *Situation awareness analysis and measurement* (pp. 147-173). MahMaw, NJ: LEA.

Endsley, M.R. (2000c). Training for situation awareness. In Endsley, M.R. & Garland, D.J. (Eds.), *Situation awareness analysis and measurement* (pp. 349-365). MahMaw, NJ: LEA.

Endsley, M.R., & Garland, D.J. (Eds.). (2000). *Situation awareness analysis and measurement.* MahMaw, NJ: Lawrence Erlbaum Associates.

Endsley, M.R., & Robertson, M.M. (1996). Team situation awareness in aviation maintenance. *Proceedings of the Human Factors and Ergonomics Society, 2*, 1077-1081.

Flach, J.M., Mulder, M., & Paassen, M.M. v. (2004). The concept of the situation in psychology. In Banbury, S. & Tremblay, S. (Eds.), *A cognitive approach to situation awareness: theory and application* (pp. 42-60). Hampshire: Ashgate.

Flach, J.M. (1995). Situation awareness: Proceed with caution. *Human Factors, 37*(1), 149-157.

Gaba, D. M., Howard, S.K., & Small, S.D. (1995). Situation Awareness in Anesthesiology. *Human Factors, 37*(1), 20-31.

Garland, D. J., Endsley, M.R., Andre, A.D., Hancock, P.A., Selcon, S.J., & Vidulich, M.A. (1996). *Assessment and measurement of situation awareness.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society.

Gilson, R.D., Garland, D.J., & Koonce, J. M. (1994). *Situation awareness in complex systems.* Daytona Beach: Embry-Riddle Aeronautical University Press.

Hardy, D.J., & Parasurama, R. (1997). Cognition and flight performance in older pilots. *Journal of Experimental psychology: Applied, 3*(4), 313-348.

Hoebee, W, Wiersma, J.W.F., Padje, W. van 't & Holtz, I. (2005). *Human factor aspects of AIS in the VTS environment.*, Conference paper International conference on port-maritime development and innovation (pp. 1-8). Rotterdam: Port of Rotterdam.

Hogg, D.N., Folleso, K., Strand-Volden, F., & Torralba, B. (1995). Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics, 38*(11), 2394-2413.

IALA (1993). *IALA Vessel Traffic Services Manual* St Germaine-en-Laye: IALA.

IALA (1998). *IALA Recommendation V-103 on standards for training and certification of VTS personnel.* St Germaine-en-Laye: IALA.

IALA (2001). *IALA guidelines on the assessment of training requirements for VTS Existing VTS personnel, candidate VTS operators, and revalidation of VTS operator certificates* St Germaine-en-Laye: IALA.

IALA (2002a). *IALA Vessel Traffic Services Manual* St Germaine-en-Laye: IALA.

IALA (2002b). *Recommendation on the Provision of Shore Based Automatic Identification Systems (AIS)*. St Germaine-en-Laye: IALA.

IALA (2003). *The aspects of the Training of VTS Personnel relevant to the introduction of Automatic Identification System*. Saint Germain en Laye.

Jarvis, D. (1997). *Second workshop COMFORTABLE Task 27*. Rotterdam: MSR.

Jones, D.G. (1996). *Effect of cue type on situation awareness.* Unpublished Dissertation, Texas Tech University, UMI Dissertation Service 9711315.

Jones, D.G., & Endsley, M.R. (1996). Sources of situation awareness errors in aviation. *Aviation Space and Environmental Medicine, 67*(6), 507-512.

Judge, C.L., & Bushman, J.B. (1992). Situation awareness: Modelling, measurement and impacts *Proceedings of the Human Factors Society 36th annual meeting* (Vol. 36, pp. 40-42): HFS.

Kaber, D.B. (1996). *The effect of level of automation and adaptive automation on performance in dynamic control environments.* Unpublished Dissertation, Texas Tech University, UMI Dissertation Service 9638398.

Klein, G. (1996). Studying situation awareness in the context of decision-making incidents. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 177-182). Daytona Beach: Embry-Riddle Aeronautical University Press.

Laughery, K.R., & Wogalter, M.S. (1997). Warnings and risk perception. In Salvendy, G. (Ed.), *Handbook of Human Factors and Ergonomic* (pp. 1174-1197). New York: Handbook of Human Factors and Ergonomic.

Malskat, J.A.H. (2009). *Beveiligingsbewustzijn meten! (in Dutch)* Unpublished Masters thesis, Delft Top Tech, Delft.

McGowan, A.M., & Banbury, S. (2004). Evaluating interruption-based techniques using embedded measures of driver anticipation. In Banbury, S. & Tremblay, S. (Eds.), *A cognitive approach to situation awareness: theory and application* (pp. 176-192). Hampshire: Ashgate.

McGuinness, B. (1995). Situational awareness measurement in cockpit evaluation trials. In AGARD-AMP (Ed.), *Situation awareness: Limitations and enhancement in the aviation environment* (pp. 7.1-8): AGARD-CP-575.

Metalis, S.A. (1993). Assessment of pilot situational awareness: measurement via simulation *Proceedings of the Human Factors and Ergonomics Society 37th annual meeting* (Vol. 37, pp. 113-117): HFES.

MSR (2001). *Ontwikkeling van het Prestatie Monitoring Instrument voor Vessel Traffic service operators. MSR Report R&D/98/046,Rotterdam.*

NNVO (2009). Leerplan Basisopleiding (BO) praktijkvaardigheid verkeersdienstsimulator (VDS), Rotterdam

O'Neil, W. A. (2000). IALA VTS 2000 Symposium, Keynote Address. Singapore: IMO.

Pew, R.W. (1994). An introduction to the concept of situation awareness. In Gilson, R.D., Garland, D.J. & Koonce, J.M. (Eds.), *Situation awareness in complex systems.* (pp. 17-23). Daytona Beach: Embry-Riddle Aeronautical University Press.

Pew, R.W. (1996). The state of situation awareness measurement: Circa 1995. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 7-16). Daytona Beach: Embry-Riddle Aeronautical University Press.

Prince, C., Salas, E., & Stout, R.J. (1996). Situation awareness: team measures, training methods. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 135-140). Daytona Beach: Embry-Riddle Aeronautical University Press.

Pritchett, A., & Hansman, R.J. (2000). Use of testable responses for performance-based measurement of situation awareness. In Endsley, M.R. & Garland, D.J. (Eds.), *Situation awareness analysis and measurement* (pp. 198-210). MaHwah, NJ: lea.

Regelink, H., Coles, P., & Jarvis, D. (1999). *COMFORTABLE Final Report.* Rotterdam: MSR.

Regelink, H., & Jarvis, D. (1999). *End-user involvement in the design and testing of VTS tools.* Rotterdam: MSR.

Robertson, M.M., & Endsley, M.R. (1996). A methodology for analyzing team situation awareness in aviation maintenance. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 313-320). Daytona Beach: Embry-Riddle Aeronautical University Press.

Rodgers, M.D., Mogford, R.H., & Strauch, B. (2000). Post hoc assessment of situation awareness in air traffic control incidents and major aircraft accidents. In Endsley, M.R. & Garland, D.J. (Eds.), *Situation awareness analysis and measurement* (pp. 73-112). MaHwah, NJ: lea.

Rotterdam, P. o. (2009). *Port Statistics.*

Sarter, N.B., & Woods, D.D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*(1), 5-19.

Saus, E.-R., Johnsen, B.H., Eid, J., Riisem, P.K., Andersen, R., Thayer, J.F. (2006) The effect of brief situational awareness training in a police shooting simulator: An experimental study. *Military Psychology,* 18 (SUPPL.), pp. S3-S21.

Scerbo, M.W., Freeman, F.G., Mikulka, P.J., Schoenfeld, V., Eischeid, T.M., Krahl, K., et al. (1999). *Hazardous States of Awareness - What are they and how do we measure them?* . Paper presented at the International Symposium on Aviation Psychology, 10th, Columbus, OH.

Schank, R., & Abelson, B. (1977). *Scripts, plans, goals and understanding : an inquiry into human knowledge structures*. Hillsdale, N.J.: Erlbaum.

Schenk, T.L. (1994). Situational awareness in emergency services incident command. In Gilson, R.D., Garland, D.J. & Koonce, J.M. (Eds.), *Situation awareness in complex systems*. (pp. 291-293). Daytona Beach: Embry-Riddle Aeronautical University Press.

Selcon, S.J., & Taylor, R.M. (1990). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew system design. In AGARD-AMP (Ed.), *Situational awareness in areospace operations* (pp. 5.1-8): AGARD-CP-478.

Selcon, S.J., Taylor, R.M., & Koritsas, E. (1991). *Workload or situational awareness?. TLX vs. SART for aerospace systems design evaluation.* Paper presented at the Proceedings of the Human Factors Society.

Small, S.D. (1996). Measurement and analysis of situation awareness in anesthesiology. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 123-128). Daytona Beach: Embry-Riddle Aeronautical University Press.

Smith, K., & Hancock, P.A. (1995). Situation Awareness Is Adaptive, Externally Directed Consciousness. *Human Factors, 37*(1), 137-148.

Smyth, C.C., Malkin, F.J. and DeBellis, W.B. (1990). Counterair situation awareness display for army aviation. In AGARD-AMP (Ed.), *Situational awareness in areospace operations* (pp. 10.11): AGARD-CP-478.

Stanton, N.A., Chambers, P.R.G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science, 39*(3), 189-204.

Stern, J. A., Wang, L., & Schroeder, D. (1996). Physiological measurement techniques: What the heart and eye can tell us about aspects of situational awareness. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 155-162). Daytona Beach: Embry-Riddle Aeronautical University Press.

Strauch, B. (1996). Post-hoc assessment of situation assessment in aircraft accident/incident investigations. In Garland, D.J. & Endsley, M.R. (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 163-170). Daytona Beach: Embry-Riddle Aeronautical University Press.

Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew system design. In AGARD-AMP (Ed.), *Situational awareness in areospace operations* (pp. 3.1-17): AGARD-CP-478.

Taylor, R.M. (1996). *Cognitive approach to human-centered systems design.* Paper presented at the Proceedings of SPIE - The International Society for Optical Engineering.

Taylor, R.M., & Selcon, S.J. (1990). Cognitive quality and situational awareness with advanced aircraft attitude displays. *Proceedings of the Human Factors Society*, 26-30.

Thomson, M.E., & Solms, R. van (1998). Information security awareness: educationg your users effectively. *Information management & computer security, 6*(4), 167-173.

Vidulich, M., Dominguez, C., Vogel, E., & McMillan, G. (1994). *Situation awareness; Papers and annotated bibliography* (No. Tech. Rep. No. AL/CF-TR-1994-008). Wright Patterson Air Force Base, OH: Armstrong Laboratory.

Vidulich, M.A. (2000). *The relationship between mental workload and situation awareness.* Paper presented at the Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium'.

Vidulich, M.A., & Hughes, E.R. (1991). Testing a subjective metric of situation awareness *Proceedings of the Human Factors Society 35th annual meeting* (Vol. 35, pp. 1307-1311): HFS.

Vidulich, M.A., & McCoy, A.L. (1995). Attentional control and situational awareness in a complex air combat simulation. In AGARD-AMP (Ed.), *Situation awareness: Limitations and enhancement in the aviation environment* (pp. 18.11-15): AGARD-CP-575.

Vincenzi, D.A., Hays, R.T., Seamon, A.G. (2000) *Measuring situation awareness in training systems: A multivariate approach.* Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium', pp. 361-364.

Vincenzi, D.A., Mouloua, M., & Hancock, P.A. (Eds.). (2004). *Human Performance, situation awareness and AutomationII: Current Research and Trends: v.I & v. II (HPSAA II)*: Psychology Press.

Westrenen, F.C. van (1999). *The maritime pilot at work.* . Dissertation, Delft University of Technology, Eburon:Delft.

Wickens, C.D. (1992). *Engineering psychology and human performance. Second edition.* New York: Harper Collins.

Wickens, C.D. (2000). *The when and how of using 2-D and 3-D displays for operational tasks.* Paper presented at the Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium'.

Wiersma, J.W.F. (2000). Situation Awareness in Vessel Traffic Service Operation. In: Endsley, M.R. & Kaber, D.B. (Eds.), *Human Performance, situation awareness and automation: user-centered design for the new millenium* (pp. 280-285). Missisippi State University. Wiersma, J.W.F. (2004). Principles of design and Human Factors in transportation. In Waard, D. de, Brookhuis, K.A. & Weikert, C.M. (Eds.), *Human factors in design* (pp. 13-25). Maastricht: Shaker Publishing bv.

Wiersma, J.W.F. (2004). Principles of design and Human Factors in transportation. In Waard, D. de, Brookhuis, K.A. & Weikert, C.M. (Eds.), *Human factors in design* (pp. 13-25). Maastricht: Shaker Publishing bv.

Wiersma, J.W.F. (2005). Assessing VTS operator situation awareness. In Jagtman H.M. & Wiersma, J.W.F., (Eds.), *The research agenda of risk and design anno 2005* (pp. 25-31). Delft: TUD.

Wiersma, J.W.F., Butter, R. & Padje, W. van 't (2000). A Human factors approach to assessing VTS operator performance. In Bovy P.H.L. (Ed.), *Transport, Infrastucture and Logistics* (pp. 1-16). Delft: Delft University Press.

Wiersma, J.W.F. & Butter, R. (2002). Situation awareness in maritime traffic control: a comparison of two methods. In: Waard, D. de, Brookhuis, K.A., Moraal, J. & Toffetti, A. (Eds.), *Human Factors in Transportation, Communication, Health, and the Workplace* (pp. 377-385). Nederland: Shaker Publishing BV.

Wiersma, J.W.F., Glansdorp, C., & Stoop, J. (1995). Traffic management systems: A safety help or hindrance? In Liu B. & Blosseville, J.M. (Eds.), *Transportation systems: theory and application of advanced technology* (Vol. 1. A postprint volume from the IFAC Symposium, Tianjin, PRC, 24-26 August 1994., pp. 501-506). London: Pergamon.

Wiersma, J.W.F., & Heijer, T. (1996). *Human Factors Models and Methodologies for Vessel Traffic Service Operator Performance.* (No. PL95-198 E.C. 4th Framework Programme Transport "To COMFORT VTS Management" (COMFORTABLE) Project 6.3.4, Task 27 WP1). Delft: Delft University of Technology.

Wiersma, J.W.F., & Hooijer, J. (1997). Situation awareness base-line for Vessel Traffic Service Operators *From Experience to Innovation.* (pp. 469-471). Helsinki: Finnish Institute of Occupational Health.

Wiersma, J.W.F., Jarvis, D., & Granholm, C. (2000). *VTS operator performance and acceptance of a short-term path prediction display in a low density traffic area.* Paper presented at the Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Association, 'Ergonomics for the New Millennium'.

Wiersma, J.W.F., Mastenbroek, N., & Wulder, J.H. (1999). SATEST: The 1997 Rotterdam Experiment on VTS operator performance. In Inoue, K. & Nishida, S. (Eds.), *7th IFAC/IFIP/IFOR/IEA symposium on analysis, design and evaluation of man-machine systems. Kyoto, 9/16/98* (pp. 281-286). Oxford: Elsevier.

Wiersma, J.W.F., & Mastenbroek, N. (1999). Measurement of Vessel Traffic Service Operator Performance. *Al & Society, 12*, 78-86.

Wiersma, J.W.F., & Padje, W. van 't (2005). Does Automatic Identification System information improve efficiency and safety of Vessel Traffic Services. In Waard, D. de, Brookhuis, K.A. Egmond, R. van & Boersema B. (Eds.), Human factors in design, safety and management (pp. 97-104). Maastricht: Shaker.

Wilson, G.F. (1996). Psychophysiological assessment of SA? In D.J. Garland & M.R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 141-146). Daytona Beach: Embry-Riddle Aeronautical University Press.

ANNEX 1

Reviewed literature on situation awareness

	Type of paper			Field of application.							Method used							
	Theory	Definition SA	Methodology	Experiments	Maritime	Aviation	ATC	Road taffic	Military	Other	Performance	Subjective	Post hoc and process	Physiological	Query	Freeze	Spatial location	Other
1	x	x				x			x									
2	x	x								x								
3	x									x								
4				x		x					x							
5				x		x			x			x						
6	x		x							x		x						
7				x				x			x							
8	x	x	x							x								
9	x	x																
10	x	x	x			x	x		x									
11	x	x	x	x		x			x						x	x		
12	x	x								x								
13		x		x		x					x							
14				x		x					x							
15	x		x			x			x		x							
16	x	x	x	x		x			x	x								x
17	x	x	x				x							x				
18	x							x										
19	x	x	x	x		x												
20	x		x			x			x									x
21	x	x																
22	x		x	x		x			x									
23				x						x					x	x		
24	x	x								x								
25	x	x	x	x		x												
26	x		x			x					x	x						
27	x		x							x	x				x			
28				x			x						x					
29		x		x		x				x					x			
30	x	x	x	x			x											
31			x	x		x			x		x							
32	x	x	x			x									x	x		

	Type of paper			Field of application.							Method used							
	Theory	Definition SA	Methodology	Experiments	Maritime	Aviation	ATC	Road taffic	Military	Other	Performance	Subjective	Post hoc and process	Physiological	Query	Freeze	Spatial location	Other
33		x	x	x		x			x						x	x		
34	x	x	x	x		x												
35	x	x								x								
36	x	x																
37	x	x	x			x							x					x
38	x		x	x		x			x	x		x			x	x		
39	x	x								x								
40	x		x	x											x	x		
41	x	x	x															
42	x	x	x	x						x								x
43	x	x	x			x				x								
44			x	x			x								x	x		
45	x	x	x							x								
46	x	x				x			x									
47	x						x							x				x
48	x	x	x							x								
49	x	x																
50	x	x								x								
51				x		x			x								x	x
52				x		x			x							x	x	
53	x		x			x			x			x						
54	x	x								x								
55	x	x					x											
56	x		x							x								x
57	x					x							x					
58	x		x							x								
59			x	x				x										x
60	x		x	x				x			x						x	x
61	x	x	x	x			x					x						
62				x		x			x		x	x						
63	x	x					x			x								
64	x	x	x							x					x	x		

	Type of paper			**Field of application.**							**Method used**							
	Theory	**Definition SA**	**Methodology**	**Experiments**	**Maritime**	**Aviation**	**ATC**	**Road taffic**	**Military**	**Other**	**Performance**	**Subjective**	**Post hoc and process**	**Physiological**	**Query**	**Freeze**	**Spatial location**	**Other**
65	x	x					x											
66	x						x											
67	x	x								x								
68	x					x												
69	x	x				x												
70	x	x				x			x									
71	x	x	x			x			x			x						
72	x	x	x	x			x											
73	x	x		x		x												x
74	x	x	x							x								
75	x	x	x	x						x								
76		x		x	x				x				x					x
77	x	x	x											x				
78	x		x			x			x									
79	x		x						x				x					
80		x				x					x	x						
81				x		x			x		x							x
82	x	x				x				x								
83	x						x											
84	x	x	x			x			x									
85		x		x		x			x			x						
86	x	x	x															
87	x	x				x			x									
88	x	x	x			x						x			x	x		
89		x		x			x				x				x	x		
90	x		x	x		x	x						x					x
91	x	x		x			x											
92				x	x				x		x	x						
93				x		x					x							x
94	x			x		x												
95		x		x		x					x				x			
96	x					x					x							x

	Type of paper			Field of application.							Method used							
	Theory	Definition SA	Methodology	Experiments	Maritime	Aviation	ATC	Road taffic	Military	Other	Performance	Subjective	Post hoc and process	Physiological	Query	Freeze	Spatial location	Other
97	x	x	x			x												x
98	x	x								x								
99	x	x	x															
100				x		x									x			x
101	x		x			x									x			
102	x		x				x				x							
103	x	x								x								
104	x	x					x											
105	x		x				x				x							
106	x		x							x								
107	x		x				x											
108	x		x															
109	x		x								x							
110	x		x			x				x								
111	x						x											
112				x			x						x					x
113	x	x																
114	x	x	x			x												
115	x	x				x												
116																		
117	x	x								x								
118				x		x						x						
119	x	x				x			x									
120				x		x			x		x	x						
121			x	x		x			x			x						
122	x	x	x							x			x					
123	x		x	x						x					x	x		
124	x	x					x											
125	x					x			x	x								
126			x	x			x				x	x			x			x
127				x		x			x		x							
128				x						x				x				

	Type of paper			Field of application.							Method used							
	Theory	Definition SA	Methodology	Experiments	Maritime	Aviation	ATC	Road taffic	Military	Other	Performance	Subjective	Post hoc and process	Physiological	Query	Freeze	Spatial location	Other
129	x		x	x		x	x			x			x		x			
130	x		x			x	x						x					
131		x		x		x			x									x
132	x	x	x			x						x						
133	x		x									x						
134				x		x			x			x						
135	x		x			x			x			x						
136	x	x								x								
137		x	x	x		x						x						
138				x						x					x	x		x
139			x	x		x			x		x							
140	x		x															x
141	x		x	x		x			x		x	x						
142		x		x		x			x						x	x		
143				x				x			x	x		x				x
144				x				x			x							
145		x																
146	x	x									x							
147	x		x			x								x				

1. Adam, E.C. (1994) Tactical Cockpits - The coming revolution. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 101-110.
2. Adams, M.J., Tenney, Y.J. and Pew, R.W. (1995) Situation awareness and the cognitive management of complex ssytems. *Human factors* 37(1) 85-104.
3. Adams, S.K., Babin, S.P. and Sabri, Z.A. (1996) The virtual patient - an application of situation awareness for system design and training in intensive care. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 329-338.
4. Beer, J.M.A., Gallaway, R.A. and Previc, F.H. (1996) Aircraft recognition thresholds and manual attitude control: Individual performance links that might be important to situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 321-328.
5. Bell, H.H. and Waag, W.L. (1996) Using observer Ratings to assess situational awareness in tactical air environments. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 93-100.
6. Bennett, F.S. (1996) Situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 415-422.
7. Beusmans, J. Aginsky, V., Harris, C., and Rensink, R. (1996) Analyzing situation awareness during wayfinding in a driving simulator. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 245-252.
8. Billings, C.E. (1994) Situation awareness in complex systems: A commentary. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 321-325.
9. Billings, C.E. (1996) Situation awareness measurement and analysis: A commentary. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 1-5.
10. Blanchard, R.E., Endsley, M., Mogford, R., Smolensky, M. and Rodgers, M. (1993) Situation awareness - transition from theory to practise. In: *Proceedings of the Human Factors and Ergonomics Society 37th annual meeting*, HFES-37 pp. 39-42.
11. Bolstad, C.A. (1991) Individual pilot differences related to situation awareness. In: *Proceedings of the Human Factors Society 35th annual meeting*, HFS-35 pp. 52-56.
12. Bolstad, C.A. and Hess, T.M. (1996) Situation awareness and older workers. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 289-294.
13. Bowers, C.A. Braun, C. and Kline, P.B. (1994) Communication and team situational awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 305-311.
14. Breda, L van, Veltman, H.A. and Bronkhorst, A.W. (1997) Three-dimensional displays in the cockpit to improve situation awareness. In: Soekkha, H.M. ed. *Aviation safety.* Proceedings of the International Aviation Safety conference 1997, IASC-97, VSP, Utrecht, pp. 729-742.
15. Brickman, B.J., Hettinger, L.J., Roe, M.M., Stautberg, D., Vidulich, M.A., Haas, M.W. and Shaw, R.L. (1996) An assessment of situation awareness in an air combat simulation: The global implicit measure approach. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of*

situation awareness. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 339-344.

16. Busquets, A.M. Parrish, R.V. Williams, S.P. and Nold, D.E. (1994) Comparison of pilots' acceptance and spatial awareness when using EFIS vs. pictorial display formats for complex, curved landing approaches. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 139-167.
17. Byrne, E.A. (1996) Role of volitional effort in the application of psychophysiological measures to situation awareness . In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 147-154.
18. Cameron, D.L. (1996) Evaluation of "RLMS" automotive rear lighting . In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 253-258.
19. Carmody, M.A. (1993) *Task-dependent effects of automation: The role of internal models in performance, workload, and situation awareness in a semi-automated cockpit*. Dissertation Texas Tech University., UMI Dissertation Service 9416605.
20. Carretta, T.R. Perry, D.C. and James Ree, M. (1994) The ubiquitous three in the prediction of situational awareness: round up the usual suspects. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 125-137.
21. Charness, N. (1996) Expert performance and situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 35-42.
22. Chubb, G.P. (1996) Displays to enhance air combat situational awareness . In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 345-350.
23. Collier, S.G. and Follesø, K. (1996) SACRI: A measure of situation awareness for nuclear power plant control rooms. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 115-122.
24. Companion, M.A. (1994) Situational awareness in emergency management systems: An overview. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 283-289.
25. Dennehy, K. (1997) *Cranfield situation awareness scale users manual*. College of Aeronautics Report No.9702, Cranfield.
26. Dennehy, K.E. (1996) *The development of a rating scale to measure the situation awareness of student civil pilots*. Dissertation Cranfield University, UMI Dissertation Service 9709092.
27. Durso, F.T., Truitt, T.R., Hackworth, C.A., Crutchfield, J.M., Nikolic, D., Moertl, P.M., Ohrt, D. and Mannaing, C.A. (1996) Expertise and chess: A pilot study comparing situation awareness Methodologies . In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 295-304.
28. Durso, F.T., Truitt, T.R., Hackworth, C.A., Ohrt, D., Hamic, J.M., Crutchfield, J.M., and Manning, C.A. (1996) Factors characterizing en route operational errors: Do they tell us anything about situation awareness?. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of*

situation awareness. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 189-196.

29. Dyck, J.L. and Gilson, R.D. (1994) Situation awareness in marginal weather conditions. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 251-259.
30. East, K.P. (1996) *Workload, performance and the development of situation awareness: An information processing-based approach.* Thesis, University of Guelph. UMI Dissertation Service ADGMM09674.
31. Eidelkind, M. Moffett, R., Arendt, D. and McKee,C. (1996) Give me situation awareness, or give me death. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 407-414.
32. Endsley, M.R. (1990) A methodology for the objective measurement of pilot situation awareness. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 pp. 1.1-9.
33. Endsley, M.R. (1990) Predictive utility of an objective measure of situation awareness. In: *Proceedings of the Human Factors Society 34th annual meeting*, HFS-34 pp. 41-45.
34. Endsley, M.R. (1990) *Situation awareness in dynamic human decision making: Theory and measurement.* Dissertation University of Southern California.
35. Endsley, M.R. (1991) Situation awareness in dynamic systems. In: *International Ergonomics Association Proceedings 1991*. IEA 1991 pp. 801-803.
36. Endsley, M.R. (1994) Situational awareness: some reflections and comments. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 315-317.
37. Endsley, M.R. (1995) A taxonomy of situation awareness errors. *Human Factors in Aviation Operation.* Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), Trinity College, Dublin, Ireland, 1994. Vol. 3, Brookfield, VT, Avebury Aviation, 1995, pp. 287-292.
38. Endsley, M.R. (1995) Measurement of situation awareness in dynamic systems. *Human Factors* 37 (1) 65-84.
39. Endsley, M.R. (1995) Toward a theory of situation awareness in dynamic systems. *Human Factors* 37 (1) 32-64.
40. Endsley, M.R. (1996) Direct measurement of situation awareness in simulations of dynamic systems: Validity and use of SAGAT. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 107-114.
41. Endsley, M.R. (1996) Theoretical underpinnings of situation awareness: A critical review. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 17-24.
42. Endsley, M.R. and Kiris, E.O. (1995) The out-of-the-loop performance problem and level of control automation. *Human Factors* 37(2) 381-394.
43. Endsley, M.R. and Robertson, M.M. (1996) Team situation awareness in aviation maintenance. In: *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, HFES-40 pp. 1077-1081.
44. Endsley, M.R. and Rodgers, M.D. (1996) Attention distribution and situation awareness in Air Traffic Control. In: *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, HFES-40 pp. 82-85.
45. Endsley, M.R., Klein, G. Woods, D.D., Smith, Ph.J. and Selcon, S.J. (1995) Future directions in cognitive engineering and naturalistic decision making. In: *Proceedings of the Human Factors and Ergonomics Society 39th annual meeting*, HFES-39 pp. 450-453.

46. Entin, E.B., Serfaty, D. and Entin, E.E. (1996) Modeling and measuring situation awareness for target-identification performance. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 233-244.
47. Evans, A, Donohoe, L, Kilner, A., Lamoureux, T., Atkinson, T, MacHendrick, H. and Kirwan, B. (1997) ATC, Automation, and human factors: Research Challenges. In: Soekkha, H.M. ed. *Aviation safety*. Proceedings of the International Aviation Safety conference 1997, IASC-97, VSP, Utrecht, pp. 131-150.
48. Firth, J.L. (1995) Neurological dimensions of situational awareness. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment*. AGARD-CP-575 pp. 26.1-5.
49. Flach, J.M. (1995) Situation awareness: Proceed with caution. *Human Factors* 37(1) 149-157.
50. Flach, J.M. (1996) Maintaining situation awareness when stalking cognition in the wild. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 25-34.
51. Fracker, M.L. (1989) Attention allocation in situation awareness. In: *Proceedings of the Human Factors Society 33rd annual meeting*, HFE-33 pp. 1396-1400.
52. Fracker, M.L. (1990) Attention gradients in situation awareness. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 pp. 6.1-10.
53. Fracker, M.L. and Vidulich, M.A. (1991) Measurement of situation awareness: A brief review. In: *International Ergonomics Association Proceedings 1991*, IEA 1991 pp. 795-797.
54. Gaba, D.M. and Howard, S.K. (1995) Situation awareness in anesthesiology. *Human Factors* 37(1) 20-31.
55. Garland, D.J. and Hopkin, V.D. (1994) Controlling automation in future air traffic control: The impact on situational awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 179-197.
56. Garland, D.J., Endsley, M.R., Andre, A.D., Hancock, P.A., Selcon, S.J. and Vidulich, M.A. (1996) Assessment and measurement of situation awareness. In: *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, HFES-40 pp. 1170-1173.
57. Granada, T.M., McClure, D. and Fogarty, J.W. (1991) The development of an altitude awareness program: An integrated approach. In: *Proceedings of the Human Factors Society 35th annual meeting*, HFS-35 pp. 47-51.
58. Green, M., Odom, J.V. and Yates, J.T. (1996) Measuring situational awareness with the "ideal observer" . In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 351-358.
59. Gugerty, L. and Tirre, W. (1996) Cognitive correlates of explicit and implicit measures of situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 267-273.
60. Gugerty, L. and Tirre, W. (1996) Comparing explicit and implicit measures of situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 259-266.
61. Hendy, K.C. (1995) Situation awareness and workload: Birds of a feather? In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment*. AGARD-CP-575 pp. 21.1-7.

62. Hettinger, L.J., Brickman, B.J. Roe, M.M., Nelson, W.T. and Haas, M.W. (1996) Effects of virtually-augmented fighter cockpit displays on pilot performance, workload, and situation awareness. In: *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, HFES-40 pp. 30-33.
63. Hitchcock, L. (1994) Yours, mine, and ours: Some observations on the metaphysics of situation awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 3-16.
64. Hogg, D.N., Follesø, K., Strand-Folden, F and Torralba, B. (1995) Development of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms. *Ergonomics* 38(11) 2394-2413.
65. Hopkin, V.D. (1994) Situational awareness in air traffic control. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 170-178.
66. Hoshstrasser, B.D. and Small, R.L. (1994) The impact of associate systems technology on an air traffic controller's situational awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 227-236.
67. Hosley, D.L. (1994) Situational awareness in aerospace operations personnel training. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 263-268.
68. Hvidt, C. (1990) Situational awareness in aerospace operations only a pilots challenge? In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 pp. KA.1-4.
69. Javaux, D. and Figarol, S. (1995) Distributed situation awareness: a concept to cope with the challenges of tomorrow. *Human Factors in Aviation Operation*. Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP), Trinity College, Dublin, Ireland, 1994. Vol. 3 Brookfield, VT, Avebury Aviation pp. 293-298.
70. Jensen, R.S. (1997) The boundaries of aviation psychology, human factors, aeronautical decision making, situation awareness, and crew resource management. *International Journal of Aviation Psychology,* 7(4) 259-267
71. Jenssen, S.E. (1995) Will the mission workload profile allow effective situation awareness? In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment*. AGARD-CP-575 pp. 22.1-5.
72. Jones, D.G. (1996) *Effect of cue type on situation awareness*. Dissertation Texas Tech University, UMI Dissertation Service 9711315.
73. Jones, D.G. and Endsley, M.R. (1996) Sources of situation awareness errors in aviation. *Aviation, space and environmental medicine*. 67(6) 507-512.
74. Judge, C.L. and Bushman, J.B. (1992) Situation awareness: Modelling, measurement and impacts. In: *Proceedings of the Human Factors Society 36th annual meeting*, HFS-36 pp. 40-42.
75. Kaber, D.B. (1996) *The effect of level of automation and adaptive automation on performance in dynamic control environments*. Dissertation Texas Tech University, UMI Dissertation Service 9638398.
76. Kaempf, G.L., Klein, G., Thordsen, M.L. and Wolf, S. (1996) Decision making in complex naval command and control environments. *Human Factors* 38(2) 220-231.
77. Kennedy, R.S. and Ordy, J.M. (1996) Situation awareness: A cognitive neuroscience model based on specific neurobehavioral mechanisms. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 49-56.

78. King, R.E. and Flynn, C.F. (1995) Development of techniques to identify individuals with superior potential for situational awareness. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment.* AGARD-CP-575 pp. 5.1-6.
79. Klein, G. (1996) Studying situation awareness in the context of decision-making incidents. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 177-182.
80. MacLeod, I.S., Taylor, R.M. and Davies, C.L. (1996) Perspectives on the appreciation of team situational awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 305-312.
81. Marshak, W.P., Kuperman, G., Ramsey, E.G. and Wilson, D. (1987) Situational awareness in map displays. In: *Proceedings of the Human Factors Society 31st annual meeting*, HFS-31 pp. 533-535.
82. McCoy C.E. and Kollross Woleben, J. (1994) Individual differences in weather situation awareness and assessment. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 239-249.
83. McCoy, C.E., Orasanu, J., Smith, P.J., VanHorn, A., Billings, C., Denning, R., Rodvold, M., and Gee, T. (1996) Situational awareness at different levels of abstraction: The distributed cooperative problem-solving domain of ATCSCC-airline operations. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 197-202.
84. McGuinness, B. (1995) Situational awareness measurement in cockpit evaluation trials. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment.* AGARD-CP-575 pp. 7.1-8.
85. McMillan, G.R., Bushman, J., Lynn, C. and Judge, A. (1995) Evaluating pilot situational awareness in a operational environment. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment.* AGARD-CP-575 pp. K1.1-6.
86. Meister, D. (1996) Experimental analysis and measurement of situation awareness: A commentary. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 43-48.
87. Menu, J.P. and Amalberti, R. (1990) Les déterminants de l'appréciation de la situation tactique et le développement de systèmes d'aided ergonomiques. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 pp. OV-F.1-9.
88. Metalis, S.A. (1993) Assessment of pilot situational awareness: measurement via simulation. In: *Proceedings of the Human Factors and Ergonomics Society 37th annual meeting*, HFES-37 pp. 113-117.
89. Mogford, R.H. (1994) Mental models and situation awareness in air traffic control. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 199-207.
90. Mogford, R.H. (1996) Accuracy estimation in situation awareness research. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 183-188.
91. Mogford, R.H. (1997) Mental models and situation awareness in Air Traffic Control. *International Journal of Aviation Psychology*, 7(4) 331-341
92. Morrison, J.G., Kelly, R.T. and Hutchins, S.G. (1996) Impact of naturalistic decision support on tactical situation awareness. In: *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, HFES-40 pp. 199-203.

93. Mosier, K.L. and Chidester, T.R. (1991) Situation awareness and situation awareness in a team setting. In: *International Ergonomics Association Proceedings 1991*. IEA 1991 pp. 798-800.
94. Naish, P.L.N. (1990) The simulation of localized sounds for improved situation awareness.. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 p. 12-1.
95. Olmos, O., Liang, C.-C., andWickens, C.D. (1996) Construct validity of situation awareness measurement related to display design . In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 219-226.
96. Orasanu, J. (1996) Evaluating team situation awareness through communication. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 283-288.
97. Orasanu, J. (1997) Situation awareness: Its role inflight crew decision making. At: http://olias.arc.nasa.gov/publications/OSU_Orasanu/AvPs-SA.txt.html.
98. Pew, R. (1994) An introduction to the concept of situation awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 17-23.
99. Pew, R.W. (1996) The state of situation awareness measurement: Circa 1995. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 7-16.
100. Prince, C., Salas, E. and Stout, R.J. (1996) Situation awareness: team measures, training methods. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 135-140.
101. Prince, C., Salas, E., Bowers, C. and Jentsch, F. (1996) Team situation awareness research: Many paths to a destination. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 129-134.
102. Pritchett, A.R. Hansman, R.J. and Johnson, E.N. (1996) Use of testable responses for performance-based measurement of situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 75-82.
103. Prothero, J.D., Parker, D.E., Furness III, T.A., and Wells, M.J. (1996) Towards a robust, quantitative measure for presence. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 359-368.
104. Rantanen, E. (1994) The role of dynamic memory in air traffic controllers' situation awareness . In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 209-215.
105. Rantanen, E.M. and Butler, J.S. (1996) Measurement of air Traffic controllers' situation awareness and performance during simulator training. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 203-210.
106. Redding, E.F. (1994) Key situation awareness factors for orbiter flight crew and pad ground crew emergency egress and escape decisions during space shuttle terminal countdown operations. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 269-279.

107. Redding, R.E., Cannon, J.R., Lierman, B.C., Ryder, J.M., Purcell, J.A. and Seamster, T.L. (1991) The analysis of expert performance of the en route Air Traffic Control curriculum. In: *Proceedings of the Human Factors Society 35th annual meeting*, HFS-35 pp. 1403-1407.
108. Reid, G.B. (1996) SA Measurement: Lessons learned from workload. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 101-106.
109. Reising, J.M. (1996) Performance measures and situational awareness: How strong the link?. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 65-68.
110. Robertson, M.M. and Endsley, M.R. (1996) A methodology for analyzing team situation awareness in aviation maintenance. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 313-320.
111. Rodgers, M.D. and Duke, D.A. (1994) SATORI: Situation assessment through the re-creation of incidents. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 217-225.
112. Rodgers, M.D., Mogford, R.H. and Mogford L.S. (1996) Air traffic controller awareness of operational error development. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 171-176.
113. Salas, E. Stout, R.J. and Cannon-Bowers, J.A. (1994) The role of shared mental models in developing shared situational awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 297-304.
114. Sarter, N.B. and Woods, D.D. (1991) Situation awareness: a critical but ill-defined phenomenon. *International Journal of Aviation Psychology*, 1(1) 45-57
115. Sarter, N.B. and Woods, D.D. (1995) How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors* 37(1) 5-19.
116. Sarter, N.B., Woods, D.D., Braune, R.J., Palmer, E., Rogers, W.H., Wickens, C.D. Wiener, E.L. and Boje, E. (1990) Situation awareness in the advanced commercial cockpit. In: *Proceedings of the Human Factors Society 34th annual meeting*, HFS-34 pp. 21-25.
117. Schenk, T.L. (1994) Situational awareness in emergency services incident command. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 291-293.
118. Selcon, S.J. and Taylor, R.M. (1990) Evaluation of the situational awareness rating technique (SART) as a tool for aircrew system design. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 pp. 5.1-8.
119. Selcon, S.J. and Taylor, R.M. (1991) Decision support and situational awareness. In: *International Ergonomics Association Proceedings 1991*. IEA 1991 pp. 792-794.
120. Selcon, S.J., Hardiman, T.D, Croft, D.G. and Endsly, M.R. (1996) A test-battery approach to cognitive engineering: To meta-measure or not to meta-measure, that is the question! In: *Proceedings of the Human Factors and Ergonomics Society 40th annual meeting*, HFES-40 pp. 228-232.
121. Selcon, S.J., Taylor, R.M. and Koristas, E. (1991) Workload or situational awareness?: TLX vs. SART for aerospace systems design evaluation. In: *Proceedings of the Human Factors Society 35th annual meeting*, HFS-35 pp. 62-66.

122. Sferrazza, J. and Wilson, M.B. (1996) Modeling situation awareness: Using the influence diagram. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 383-406.
123. Small, S.D. (1996) Measurement and analysis of situation awareness in anesthesiology. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 123-128.
124. Smith, K. and Hancock, P.A. (1995) Situation awareness is adaptive, externally directed consciousness. *Human Factors* 37 (1) 137-148.
125. Smyth, C.C., Malkin, F.J. and DeBellis, W.B. (1990) Counterair situation awareness display for army aviation. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 p. 10.1.
126. Sollenberger, R.L. and Stein, E.S. (1996) A Simulation Study of Air Traffic Controller Situational Awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 211-218.
127. Steiner, B.A. and Camacho, M.J. (1989) Situation awareness: Icons vs. alphanumerics. In: *Proceedings of the Human Factors Society 33rd annual meeting*, HFE-33 pp. 28-32.
128. Stern, J. A., Wang, L., and Schroeder, D. (1996) Physiological measurement techniques: What the heart and eye can tell us about aspects of situational awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 155-162.
129. Stokes, A.F. and Wilt, D. (1996) FliteScript: A multimedia test to index situational schemata in pilots. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 227-232.
130. Strauch, B. (1996) Post-hoc assessment of situation assessment in aircraft accident/incident investigations. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 163-170.
131. Sullivan, C. and Blackman, H.S. (1991) Insights into pilot situation awareness using verbal protocol analysis. In: *Proceedings of the Human Factors Society 35th annual meeting*, HFS-35 pp. 57-61.
132. Taylor, R.M. (1990) Situational awareness rating technique (SART): The development of a tool for aircrew system design. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations*. AGARD-CP-478 pp. 3.1-17.
133. Taylor, R.M. (1996) Experiential measures: Performance-based self ratings of situational awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 83-92.
134. Taylor, R.M. and Selcon, S.J. (1990) Cognitive quality and situational awareness with advanced aircraft attitude displays. In: *Proceedings of the Human Factors Society 34th annual meeting*, HFS-34 pp. 26-30.
135. Taylor, R.M. and Selcon, S.J. (1991) Subjective measurement of situational awareness. In: *International Ergonomics Association Proceedings 1991*. IEA 1991 pp. 789-791.
136. Taylor, R.M. and Selcon, S.J. (1994) Situation in mind: Theory, application and measurement of situational awareness. In: Gilson, R.D., Garland, D.J. and Koonce, J.M. eds. *Situation awareness in complex systems*. Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 69-77.
137. Taylor, R.M., Shadrake, R. Haugh, J. and Bunting, A. (1995) Situation awareness, trust and compatibility: Using cognitive mapping techniques to investigate the relationships between important cognitive systems variables. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment*. AGARD-CP-575 pp. 6.1-14.

138. Uytterhoeven, G., Vlaminck, M. de and Javaux, D. (1996) Situation awareness evaluation for an operator support system in a nuclear power plant. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 275-282.
139. Venturino, M., Hamilton, W.L. and Dvorchak, S.R. (1990) Performance measures of merit for tactical situation awareness. In: AGARD Aerospace Medical Panel eds. *Situational awareness in areospace operations.* AGARD-CP-478 pp. 4.1-5.
140. Vidulich, M.A. (1996) The role of scope as a feature of situation awareness metrics. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 69-74.
141. Vidulich, M.A. and Hughes, E.R. (1991) Testing a subjective metric of situation awareness. In: *Proceedings of the Human Factors Society 35th annual meeting*, HFS-35 pp. 1307-1311.
142. Vidulich, M.A., McCoy, A.L. and Crabtree, M.S. (1995) Attentional control and situational awareness in a complex air combat simulation. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment.* AGARD-CP-575 pp. 18.1-5.
143. Ward, N.J., Fairclough, S. and Humphreys, M. (1996) The effect of task automatisation in the automotive context: A field study of an autonomous intelligent cruise control system. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 369-374.
144. Ward, N.J., Parkes, A. and Lindsay, P. (1996) The effect of automotive head-up displays on attention to critical events in traffic. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 375-382.
145. Wickens, C.D. (1995) Situation awareness: impact of automation and display technology. In: AGARD Aerospace Medical Panel eds. *Situation awareness: Limitations and enhancement in the aviation environment.* AGARD-CP-575 pp. K2.1-13.
146. Wickens, C.D. (1996) The tradeoff of design for routine and unexpected perforrnance: Implications of situation awareness. In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 57-64.
147. Wilson, G.F. (1996) Psychophysiological assessment of SA? In: Garland, D.J. and Endsley, M.R eds. *Experimental analysis and measurement of situation awareness.* Embry-Riddle Aeronautical University Press, Daytona Beach Fl. USA., pp. 141-146.

ANNEX 2

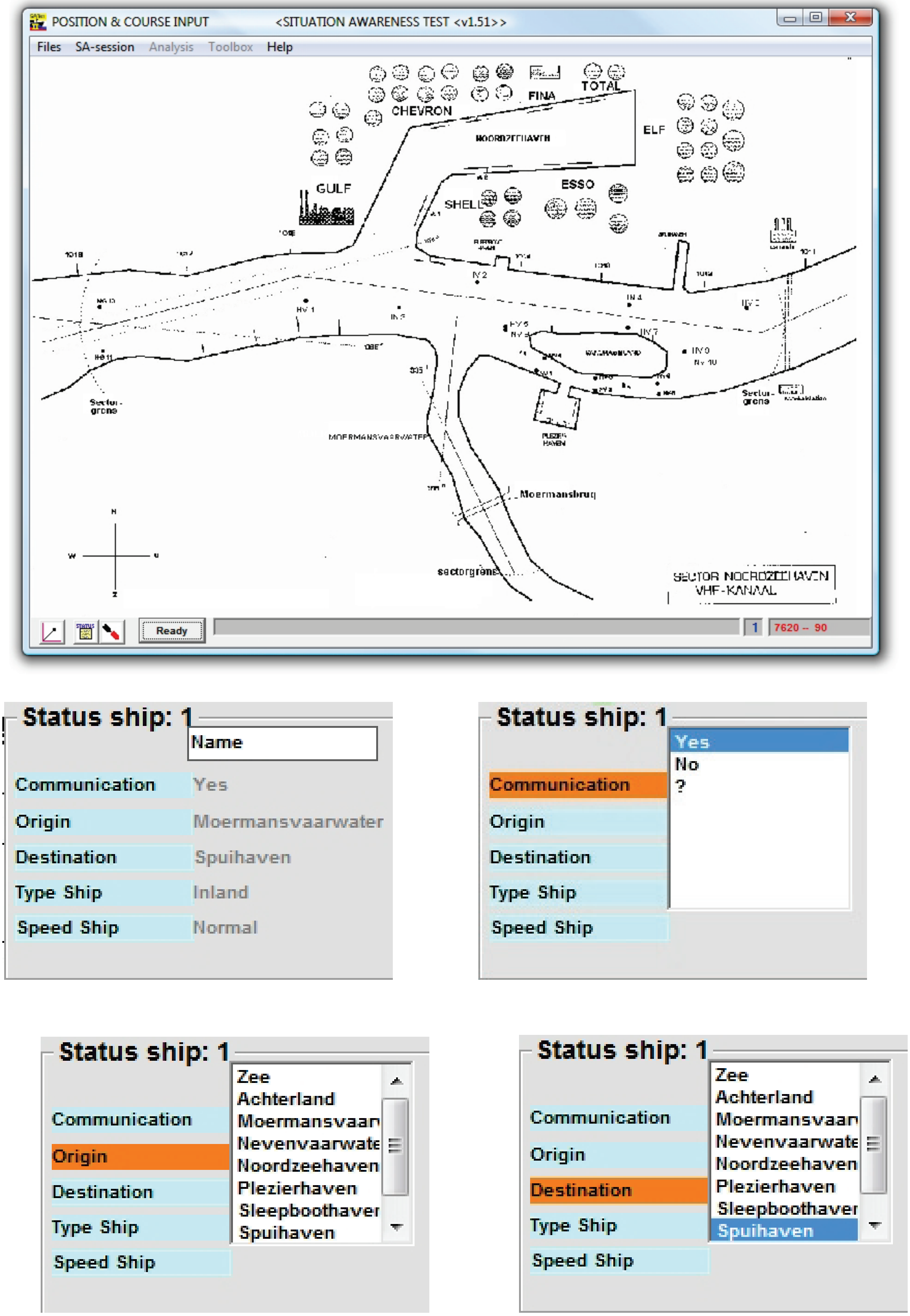
SATEST protocols and score forms

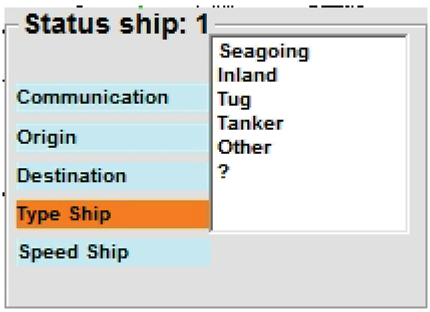
TU Delft

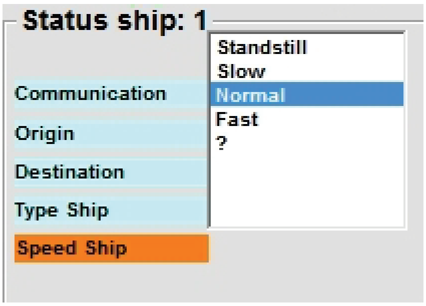
Situation Awareness
Score formulier 1.2

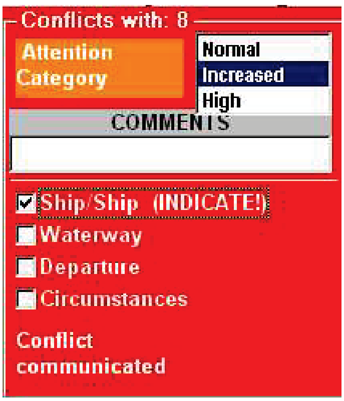
Schipnummer	*Trial*	*Proefpersoon*	*Datum*
			27 februari 1997

1	Communicatie schip	nee	ja	
2	Naam of label schip			
3	Van	Zee	Noordzeehaven	
		Achterland	Plezierhaven	
		Moermansvaarwater	Sleepboothaven	
		Nevenvaarwater	Spuihaven	
4	Bestemming	Zee	Noordzeehaven	
		Achterland	Plezierhaven	
		Moermansvaarwater	Sleepboothaven	
		Nevenvaarwater	Spuihaven	
5	Lading/Type schip	Zeevaart	Sleepboot	
		Binnenvaart	Tanker	
6	Snelheid	langzaam	normaal	snel
7	Bijzonderheden of afspraken			
8	Verdient extra aandacht	nee	ja	
9	Bevindt zich in conflict situatie	nee	ja	
10	Kan/zal in conflict situatie komen	nee	ja	
11	Waarom conflict	Schepen	Gebied	
		Afvaart	Omstandigheden	


POSITION & COURSE INPUT
<SITUATION AWARENESS TEST <v1.51>>
Files SA-session Analysis Toolbox Help
TOTAL
FINA
CHEVRON
ELF
GULF
ESSO
SHELL
Moermansbrug
sectorgrens
SECTOR NOORDZEEHAVEN
VHF-KANAAL
Ready
7620 -- 90
Status ship: 1
Name
Communication
Yes
Origin
Moermansvaarwater
Destination
Spuihaven
Type Ship
Inland
Speed Ship
Normal
Status ship: 1
Yes
No
?
Communication
Origin
Destination
Type Ship
Speed Ship
Status ship: 1
Zee
Achterland
Moermansvaar
Nevenvaarwate
Noordzeehaven
Plezierhaven
Sleepboothaver
Spuihaven
Communication
Origin
Destination
Type Ship
Speed Ship
Status ship: 1
Zee
Achterland
Moermansvaar
Nevenvaarwate
Noordzeehaven
Plezierhaven
Sleepboothaver
Spuihaven
Communication
Origin
Destination
Type Ship
Speed Ship


Status ship: 1
Communication
Origin
Destination
Type Ship
Speed Ship
Seagoing
Inland
Tug
Tanker
Other
?


Status ship: 1
Communication
Origin
Destination
Type Ship
Speed Ship
Standstill
Slow
Normal
Fast
?


Conflicts with: 8
Attention Category
Normal
Increased
High
COMMENTS
Ship/Ship (INDICATE!)
Waterway
Departure
Circumstances
Conflict communicated

HC-03a	Scenario & instructeursmeldingen	18 schepen	Oefeningtijd: 02:00 uur	Tij: Eb

MELDINGEN INSTRUCTEUR SECTOR NOORDZEEHAVEN (3A)

02.00 uur	ZP Montelena uit de Sleepboothaven naar Noordzeehaven Esso binnenkant westzijde. Voormelding vertrek Vinga Polaris van Chevron
02.02 uur	ZP Montelena wacht tot de Liguria gepasseerd is
02.05 uur	Vinga Polaris vertrekt. Subro Valour vraagt afvaart i.v.m. aanpassen Spuihaven
02.06 uur	Liguria vaart de Noordzeehaven binnen
02.07 uur	ZP Montelena vaart achter de Liguria de Noordzeehaven binnen
02.08 uur	Hoo Creek afvarend naar zee (in Engels)
02.10 uur	Subro Valour gaat naar de Noordwal en vraagt stuurboord/stuurboord met afvarende hoo Creek voor aanpassen Spuihaven
02.13 uur	Benwil dwars van de Sleepboothaven vraagt of hij door kan komen
02.14 uur	Caraibe inkomend NG 11 noord van de lichtenlijn met bestemming Handelshaven
02.15 uur	Subro Valour keert over stuurboord en steekt achteruit de Spuihaven in. Palva op vertek Noordzeehaven Esso naar zee. Alice PG gaat de Brise over bakboord oplopen
02.17 uur	Van Uden 22 vertrekt van Mineral Hoboken naar Moermansvaarwater Alice PG vraagt naar doorvaarthoogte Moermansbrug
02.19 uur	Hoo Creek loopt onbekend schip (Viroin) over de zuid op
02.20 uur	Caraibe gaat terug naar de zuid
02.21 uur	Vinga Polaris meldt zich dwars Shell 1 steiger
02.22 uur	Maria Elisabeth afvarend aan de zuidwal met bestemming Plezierhaven
02.23 uur	Palva vertrokken
02.24 uur	Alice PG meldt zich door de brug. Jan senior afvarend met bestemming Mineral Hoboken. * Jan Senior wil aan de zuid varen
02.25uur	Vinga Polaris wacht op de afvarende Hoo Creek
02.26 uur	Van Uden 22 vaart het Moermansvaarwater in
02.29 uur	Viroin onbekend schip steekt thv de Sleepboothaven achter de inkomende Caraibe over naar het Moermansvaarwater
02.30 uur	Brise meldt zich door de brug. Alice PG in de monding Moermansvaarwater en doet Stuurboord/Stuurboord met ongemeld schip
02.31 uur	* Onbekend schip (Viroin) steekt over naar de zuid
02.33 uur	Deutschland meldt zich door de brug en is bestemd voor het achterland via Nevenvaarwater. Inkomende Artisgracht dwars van NG 11 bestemd voor de Handelshaven
02.34 uur	* Onbekend schip (Viroin) loopt voor de Brise over naar Moermansvaarwater. * Vinga Polaris vraagt om positiemelding
02.35 uur	De afvarende Jan Senior gaat de Zuid opzoeken voor de inkomende artisgracht
02.36 uur	* Caraibe vraagt om positiemelding
02.37 uur	Brise in de monding Moermansvaarwater gaat voor de artisgracht naar de noord
02.38 uur	Artisgracht Stuurboord/Stuurboord met Jan Senior. Deutschland in de monding en gaat via Nevenvaarwater
02.40 uur	Voormelding vertrek Subro Valour uit Spuihaven
02.43 uur	Maria elisabeth afgemeerd. Brise vaart de noordzeehaven binnen. Artisgracht komt terug naar de zuid
02.44 uur	Uitvarende Palva uit de Noordzeehaven thv Shell1

Met * gemerkte meldingen zijn alleen ter informatie voor de instructeur

NAUTISCHE BIJZONDERHEDEN SECTOR NOORDZEEHAVEN

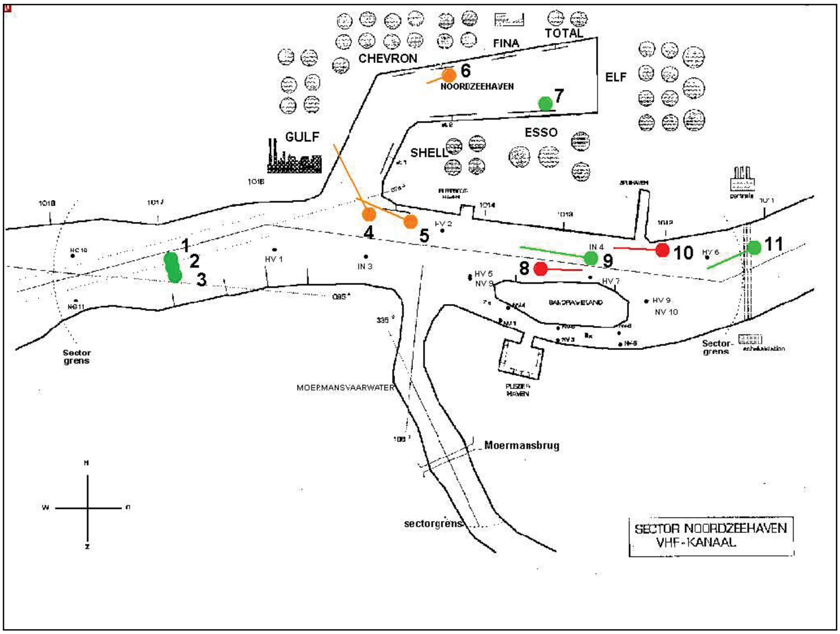
Tusse boei NG 11 en HV 1 ligt de bulkcarrier Mineral hoboken met langszij kraan Reiger voor overslag in binnenschepen; langzaam passerren i.v.m. golfslag en zuiging

Conflict1		
Description	Subro Valour wants to pass (sea going vessel) Hoo Creek Starboard/Starboard	
Before	At 5 minutes	stop scenario after call of Subro Valour and answer of the VTS operator
During	At 8 minutes	stop scenario after call Hoo Creek
After	At 10 minutes	stop scenario when Subro Valour indicates to go SB/SB with Hoo Creek the situation is over.

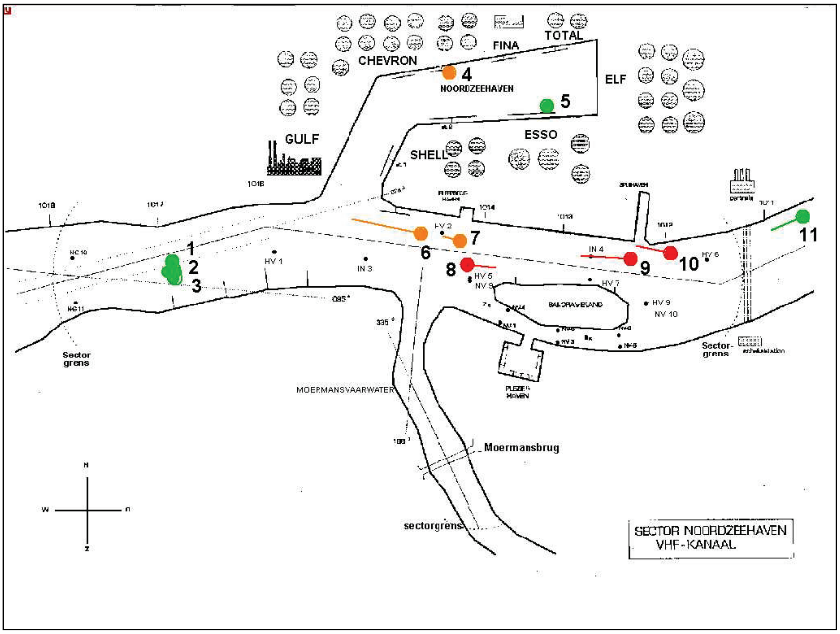
Conflict 2		
Description	The Caraibe comes in conflict with de Vinga Ppolaris when she wants to cross to the river to enter de Handelshaven.	
Before	At 14 minutes	stop scenario when Caraibe calls to report that she is going to sail on the North side
During	At 17 minutes	stop scenario after call Van Uden 22
After	At 20 minutes	stop scenario when conflict is over

Conflict3		
Description	A Ship that has not been in contact with the VTS (Viroin) crosses the river towards Moermansvaarwater in front of Brise	
Before	At 35 minutes	stop scenario call Jan Sr. Reporting it is going South
During	At 37 minutes	stop scenario after call Brise and report of unreported ship
After	At 40 minutes	Stop scenario when situation is resolved

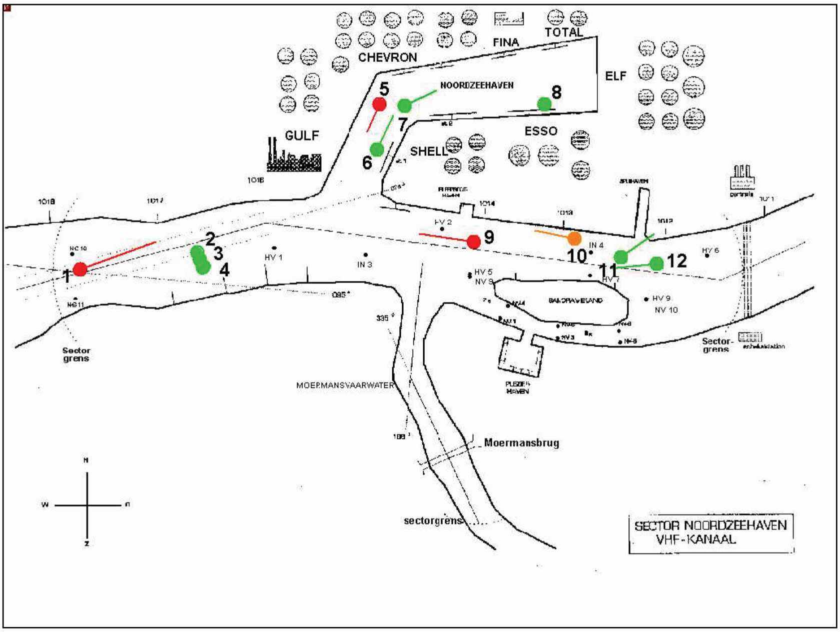
Normative pictures of traffic at breaks



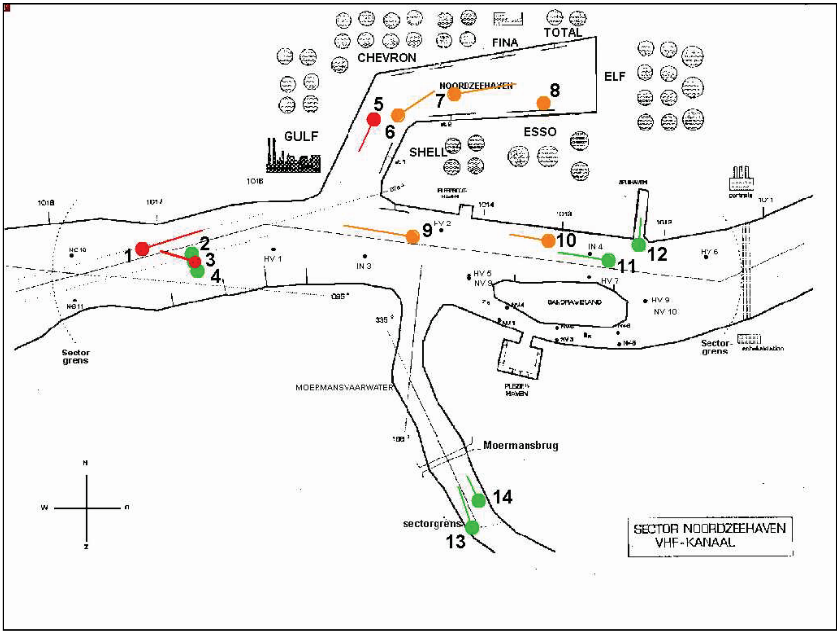
Break 1-1



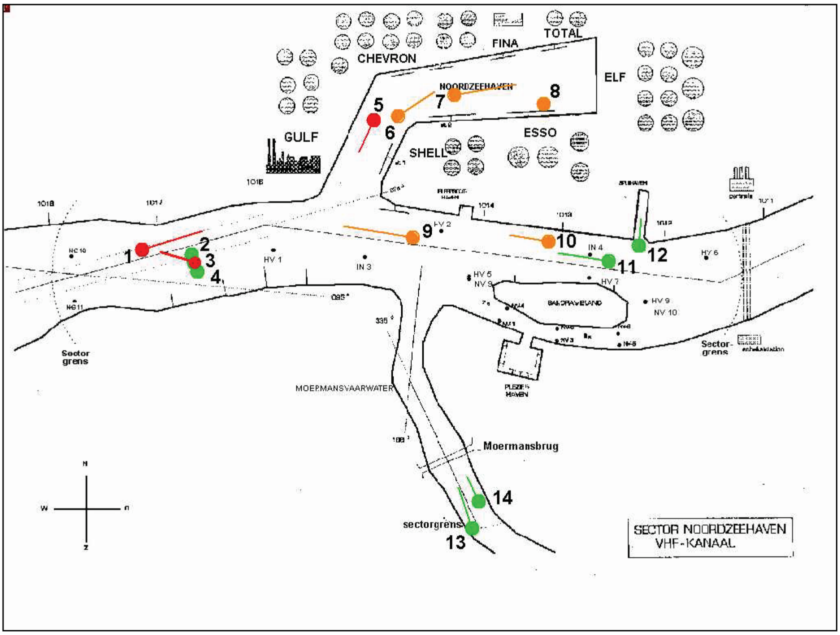
Break 1-2



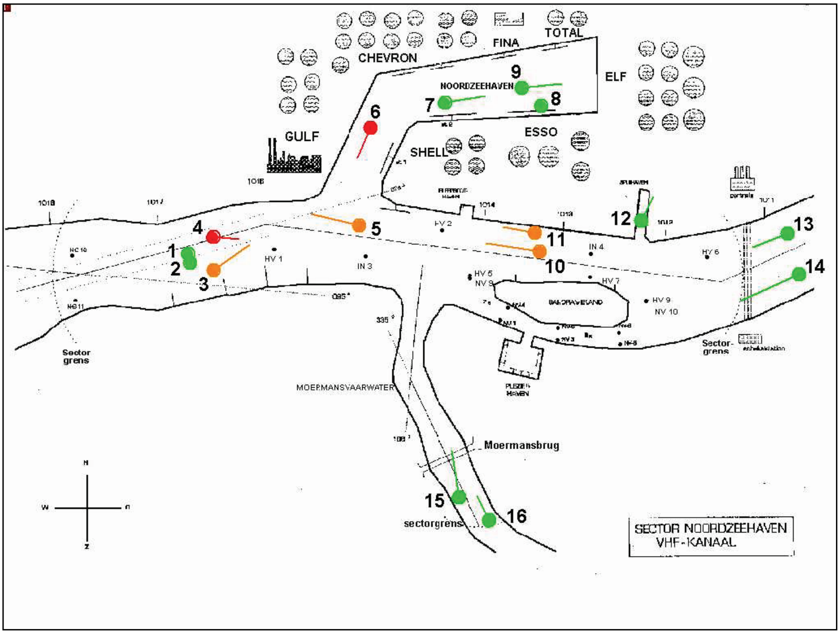
Break 1-3



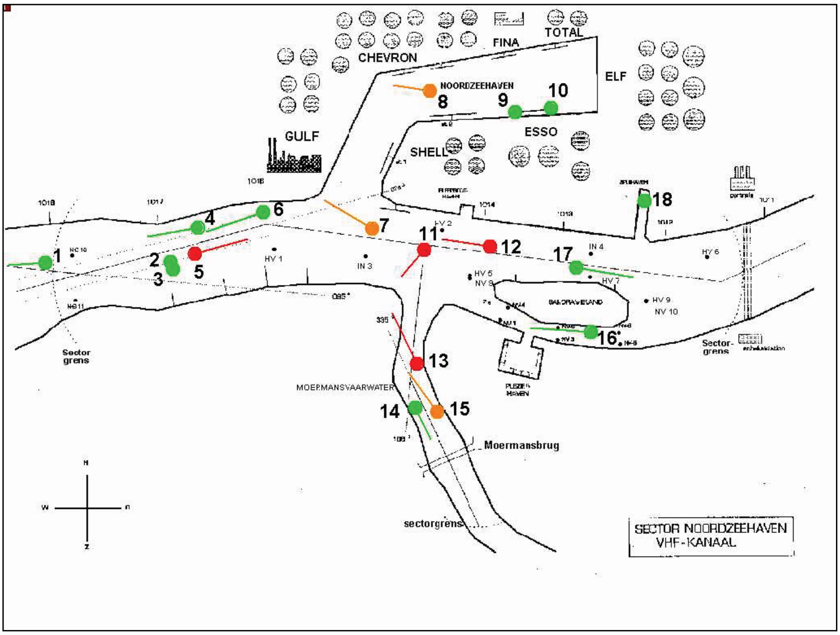
Break 2-1



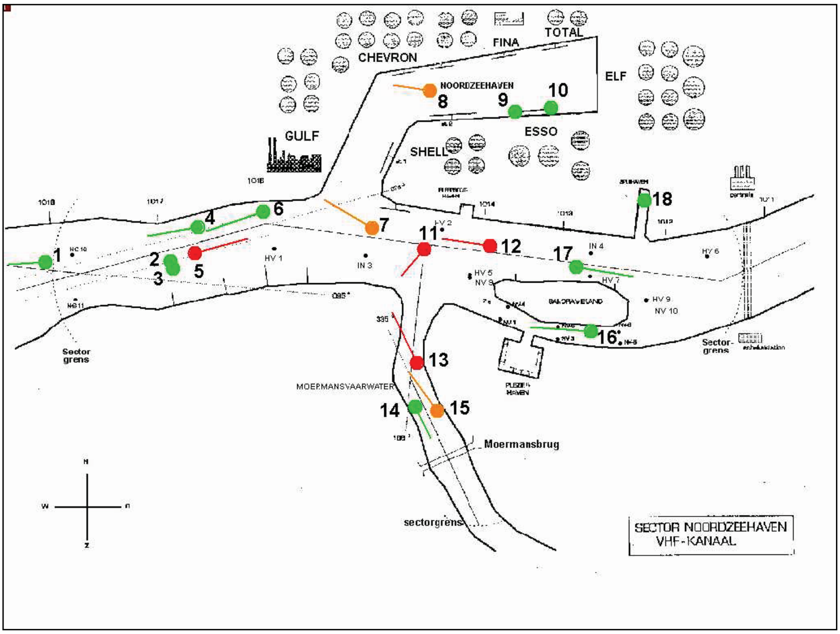
Break 2-2



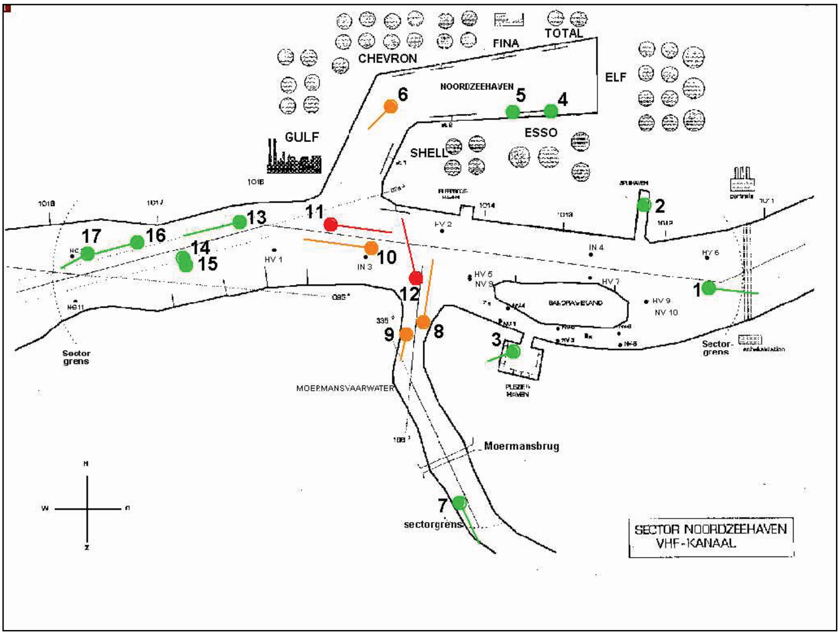
Break 2-3



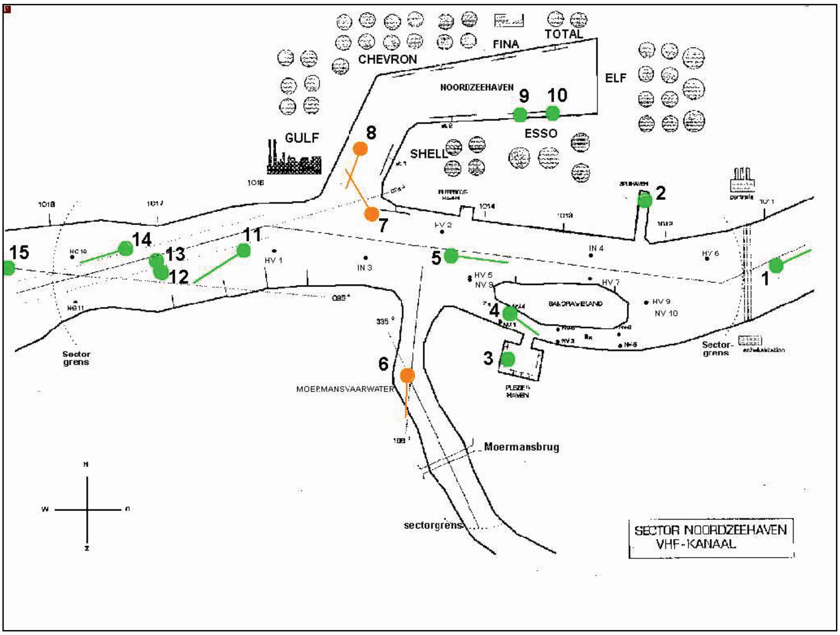
Break 3-1



Break 3-2



Break 3-3



Break 4

SATEST Protocol Rotterdam

Inleiding

Het belang van het werk van VTS operators wordt alom erkend. Er wordt dan ook veel geld geïnvesteerd om dit werk beter, effectiever en veiliger te kunnen uitvoeren. Een aanzienlijk deel van dit geld wordt gestoken in projecten waarin nieuwe technische hulpmiddelen worden ontwikkeld, die de operator moeten ondersteunen. Het is echter moeilijk om te bepalen of een specifiek apparaat nu ook daadwerkelijk helpt om het werk beter uit te voeren. Het nut ervan is lang niet altijd voor iedereen duidelijk. Mensen zijn gewend aan bepaalde dingen, ieder heeft zo zijn eigen voorkeuren en manier van werken enzovoorts. Nieuwe apparaten worden daarom altijd met scepsis bekeken.

Een goede manier om ervoor te zorgen dat een apparaat een verbetering oplevert is om de mensen die er mee moeten werken daarover te laten meepraten. De Workshop van 28 april jl. in Rotterdam was hiertoe een eerste aanzet. Onderzoekers moeten geconfronteerd worden met de kritische vraag van gebruikers: "Wat moeten wij in de praktijk hiermee?". Alleen op deze manier worden producten ontwikkeld waar u in de praktijk echt iets aan heeft.

Een andere manier is om te kijken of een apparaat een verbetering oplevert is het kijken of het gedrag van een operator verandert wanneer hij ermee werkt. Het experiment waaraan u nu meewerkt is een eerste inspanning om het gedrag van operators vast te leggen. Dit moet eerst worden gedaan. Daarna kunnen in vervolgexperimenten veranderingen in het gedrag worden bestudeerd die het resultaat zijn van andere apparatuur en extra hulpmiddelen.

Dit onderzoek wordt uitgevoerd door Marine Safety Rotterdam en de Technische Universiteit Delft, vakgroep Veiligheidskunde. Het onderzoek wordt betaald door de Europese Unie, het Gemeentelijk Havenbedrijf Rotterdam en de onderzoekende groepen zelf.

Meten van gedrag

Het is zeer lastig om te meten wat een VTS operator nu precies doet. Het deel dat waarneembaar is, is de communicatie met anderen. Dit kan worden vastgelegd. Het belangrijkste deel van het werk echter, speelt zich af in het hoofd van de operator. Die moet bepalen welke schepen in de gaten moeten worden gehouden; wie moeten er op de hoogte worden gebracht en van wat; waar conflicten dreigen en wanneer moet dat worden gecommuniceerd. De communicatie zelf is dus slechts een onderdeel van dit geheel.

In dit experiment proberen de onderzoekers een beeld te krijgen van die andere informatie, die normaal aan het oog onttrokken blijft. Deze informatie bevindt zich in het hoofd van de operator. Het is zijn bewustzijn van de (verkeers)situatie. In onderzoekstermen gaat wordt dit *Situation Awareness* genoemd. De methode om deze situation awareness van VTS operators te meten is speciaal voor dit onderzoek ontwikkeld door MSR en de TUDelft. Hiervoor is overleg geweest met praktijkmensen, om juist die elementen naar boven te halen, die voor het werk van de VTS operator relevant zijn.

Het is zeker niet het doel van dit experiment om u te beoordelen. Het gaat er in dit onderzoek niet om of u beter of slechter presteert dan uw collega's. Wij hoeven dan ook niet uw naam leeftijd etc. te weten. Om toch uw gegevens bij elkaar te houden krijgt u van ons een proefpersoonnummer. Op deze wijze kunnen wij wel achterhalen of uw test gegevens compleet zijn, maar is het ons niet bekend van welke kandidaat de testgegevens zijn. Dit nummer garandeert uw anonimiteit bij het experiment.

Het experiment

Het experiment bestaat uit de volgende elementen:

Uitleg

De instructie die u nu aan het lezen bent legt u het hoe en waarom uit van dit experiment. Indien u na het lezen van dit papier vragen hebt kunt u deze stellen aan de proefleider.

Selecteren proefpersoonnummer

U kunt zelf een nummer in de lijst aankruisen en dit vervolgens op uw formulieren invullen. Achter dit nummer kunt u aangeven hoeveel jaren werkervaring u heeft. U kunt hiervoor kiezen uit drie categorieën:

1. Minder dan één jaar werkervaring als VTS operator;
2. Eén tot en met drie jaar werkervaring als VTS operator;
3. Meer dan drie jaar werkervaring als VTS operator.

Deze informatie over uw werkervaring hebben wij nodig om de gegevens uit dit onderzoek op een juiste wijze te kunnen interpreteren.

Opzet experiment

In het onderzoek wordt gebruik gemaakt van de VTS simulator van MSR. U bent de VTS operator, terwijl een instructeur de overige rollen op zich neemt, net zoals in de trainingssessies bij MSR normaliter gebeurt. De simulatie speelt zich af in het fictieve gebied "Sector Noordzeehaven".

Op een vooraf bepaald moment wordt de simulatie stopgezet. U wordt dan voor een computer gezet waarop nogmaals de sector wordt afgebeeld. U moet hier dan zo volledig mogelijk de verkeerssituatie beschrijven. Met de muis kunt u op de afbeelding schepen positioneren en hun koers en snelheid aangeven. Geeft u van alle schepen nauwkeurig de positie aan. Vervolgens wordt u verzocht van alle schepen informatie te geven, informatie over het schip en informatie over (mogelijke) conflicten waarbij het schip betrokken is of kan raken.

Als u klaar bent met invullen gaat u weer verder met de simulatie. Het verkeer wordt hervat tot aan de volgende stop. In totaal wordt de simulatie vier keer gestopt om de verkeerssituatie te beschrijven.

Scoreformulieren

U kunt dit experiment op twee manieren uitvoeren, met behulp van de computer of op papier. Onze voorkeur gaat uit naar het gebruik van de computer, omdat dit de verwerking van de

resultaten sterk vereenvoudigd. Personen die het gebruik van de computer een onoverkomelijk bezwaar vinden kunnen dit onderzoek ook gedeeltelijk of in zijn geheel op papier uitvoeren.
In dit experiment moet u twee soorten antwoorden geven. Eerst geeft u op het computerscherm de positie aan van de schepen. U plaatst de schepen door op de linker muisknop te drukken. Wanneer u de cursor van de muis verplaatst kunt u de koers en snelheid van het schip aangegeven. Deze positie wordt vastgelegd wanneer u nogmaals op de linker muisknop drukt. Het schip wordt automatisch genummerd. U krijgt nu een venster te zien waarop u informatie over het schip kunt invullen. Wanneer u met de linker muisknop op de weergegeven onderwerpen klikt verschijnt er een venster met mogelijke opties of een leeg veld waarop u zelf informatie moet invoeren. De juiste optie kunt u weer aanklikken met de linker muisknop. Het informatiescherm verdwijnt wanneer u erbuiten klikt met de muis.
Indien u een schip op de verkeerde plek heeft gelegd, kunt u het weer verwijderen door er met de rechter muisknop op te klikken. Informatie in de invulvakken kunt u vervangen door eenvoudig andere informatie te kiezen.
Indien u alle schepen heeft weergegeven kunt u conflicten aangeven. Dit doet u door met de muis op de volgende knop te klikken.

Indien u nu op een schip klikt verandert de kleur in rood. U kunt nu aangeven met welke schepen het schip in conflict is en of hierover communicatie met het schip is geweest. Ook een conflict tussen drie of meer schepen kan worden aangegeven door op alle betrokken schepen te klikken. U kunt schakelen tussen het conflictscherm en het gewone scherm door nogmaals op de knop te klikken. Indien u alle conflicten hebt ingevoerd bent u klaar.
Indien u het te lastig vindt om al deze informatie met de computer in te voeren kunt u ook gebruik maken van de papieren versie van het experiment. Na het invoeren van de schepen via het computerscherm kunt u de informatie over de schepen en conflicten aangeven op de voorgedrukte formulieren. U gebruikt daarbij voor elk schip één formulier. De juiste antwoorden geeft u aan door ze te omcirkelen (zie voorbeeld).

3	Van	Zee	Noordzeehaven	Achterland
		Moermansvw.	Plezierhaven	Sleepboothaven
		Nevenvaarw.	Spuihaven	?

Try-out

Dit experiment begint met een try-out. De test wordt gestart en na een paar minuten stilgezet. U wordt dan in de gelegenheid gesteld om de vragen op de formulieren in te vullen. Als u klaar bent met invullen loopt u samen met de proefleider de antwoorden door. Indien u nog vragen heeft kunnen deze beantwoord worden. U kunt nu ook beslissen of u voor het experiment de computer wilt gebruiken of dat u toch liever papier gebruikt.

Het experiment

Nadat all onduidelijkheden zijn opgelost kunt u beginnen met het echte experiment. Dit zal ongeveer vijf kwartier in beslag nemen.

Schedule 1			
0’	Start	Start of scenario	
5’	Break 1	Moderate	Early
20’	Break 2	Simple	Late
37’	Break 3	Complex	Middle
45’	Break 4	End of scenario	

Schedule 2			
0’	Start	Start of scenario	
8’	Break 1	Moderate	Middle
14’	Break 2	Simple	Early
40’	Break 3	Complex	Late
45’	Break 4	End of scenario	

Schedule 3			
0’	Start	Start of scenario	
10’	Break 1	Moderate	Late
17’	Break 2	Simple	Middle
35’	Break 3	Complex	Early
45’	Break 4	End of scenario	

Helsinki Scenario information

Scenario 1, starting time 10.00 am

SHIP Name (No)	Quay	Arrival	Departure	Destination
AILI (L11)	EPL	10.30	------	---------
EEVA (L12)	EK7	11.00	------	---------
HANNA (L13)	OLA	------	10.00	Sköldvik
INKA (L14)	LJ4	------	10.00	Tallinn
JENNI (L15)	LJ5	------	10.30	StPetersburg
KAISA (L16)	LEL	n. 11.30	------	---------
LAINA (L17)	EM3	------	9.50	Tallinn

SHIP Name (No)	Ini-file	Track	Time lag (s)
AILI (L11)	HarmSis	1300t	-------
EEVA (L12)	FlathSis2	1300t	780
HANNA (L13)	LaajasUlos	1440t	200
INKA (L14)	LänsisHarm	1555t	-------
JENNI (L15)	JtksHarm	1550t	500
KAISA (L16)	VuoSis3	1455t	-------
LAINA (L17)	SrkHarm3	0065t	-------

AILI, small cruiser vessel, length about100m
EEVA, car-passenger-ferry, length about 200m
HANNA, small product tanker, length about150m
INKA, car-passenger-ferry, length about 160m
JENNI, small container, length about 150m
KAISA, motor barge (sand, gravel etc), length about 50m
LAINA, catamaran, length about 40m, velocity is restricted to 20 knots

Scenario 2, starting time 18.00

SHIP Name (No)	Quay	Arrival	Departure	Destination
MAIJA (L21)	EO2	-------	18.00	Stockholm
NELLI (L22)	EM3	-------	18.00	Tallinn
OUTI (L23)	LM9	-------	18.00	Hannover
PAULA (L24)	LS8	-------	18.00	Rotterdam
RAILI (L25)	EK6	-------	18.05	Cruising
SALLI (L26)	LEL	abt19.30	-------	-----------
TARU (L27)	LJ4	-------	18.00	Tallinn

SHIP Name (No)	Ini-file	Track	Time lag (s)
MAIJA (L21)	KatNokkU1	1200t	-------
NELLI (L22)	SrkHarm3	0060t	-------
OUTI (L23)	LänsisKata	1525t	-------
PAULA (L24)	JtksHarm	1550t	200
RAILI (L25)	KatNokkU2	1200t	300
SALLI (L26)	VuoSis3	1455t	-------
TARU (L27)	LänsisHarm	1550t	-------

MAIJA,car-passenger-ferry, length about 200m
NELLI, hydrofoil vessel, length about 40m, velocity restricted to 20kn
OUTI, container, length about 180m, DEEP CHANNEL
PAULA, container, length about 150m
RAILI, car-passenger-ferry, length about 180m
SALLI, motor-barge, length about 50m
TARU, car-passenger-ferry, length about 160m

Description of traffic at breaks

Scenario 1 Break 1

(Points of particular interest are marked *)

- ◊ HANNA* and AILI* are meeting north from Kuustamiekka. The ships are on a conflicting course.
- ◊ JENNI has just left berth.
- ◊ LAINA has just passed Harmaja pilot station. Once the ship has passed this point it is leaving the VTS area and is no longer of interest.
- ◊ KAISA is right on track.
- ◊ INKA is on track just before Husunkivi.

- ◊ The vector display shows PB/PB passing of HANNA and AILI, whereas the predictive display already shows SB/SB passing.

Scenario 1 Break 2

(Points of particular interest are marked *)

- ◊ The situation between HANNA and AILI has been solved. Both continue their journey.
- ◊ INKA is on track just passed Husunkivi.
- ◊ KAISA is right on track.
- ◊ KAISA*, INKA* and HANNA* are meeting south of Kustaanmiekka. Arrangement of traffic on the area must be made.
- ◊ JENNI is continuing her journey right on track.
- ◊ LAINA is on the south end and is about to leave the visible area.

Scenario 1 Break 3

(Points of particular interest are marked *)

- ◊ AILI is entering port.
- ◊ HANNA* has passed Kustaanmiekka and has to choose a route.
- ◊ INKA* has to decide on which side to pass Harmaja pilot station (east) and is going SB of the track to make room for EEVA to pass.
- ◊ KAISA is just continuing her journey.
- ◊ JENNI* has gone too far straight ahead and needs to start her turn PB.
- ◊ EEVA has entered the area on the south. EEVA has a small course deviation from the standard track, to make room for INKA to pass.
- ◊ (LAINA has left the area.)

- ◊ The predictive display shows that INKA is making room for EEVA and that JENNI has started to turn. The vector display does not show these manoeuvres yet.

Scenario 2 Break 1

(Points of particular interest are marked *)

- ◊ OUTI* is out of traffic lane
- ◊ TARU*, RAILI* and SALLI* have to make an arrangement of going order south side of Kustaanmiekka
- ◊ MAIJA and PAULA are on track
- ◊ NELLI has passed Harmaja pilot station and is leaving traffic area

- ◊ The predictive display shows OUTI's deviation from track.

Scenario 2 Break 2

(Points of particular interest are marked *)

- ◊ MAIJA* has to make a choice of route on which side to pass Harmaja pilot station.
- ◊ The meeting situation of TARU*, RAILI* and SALLI* south of Kustaanmiekka continues.
- ◊ OUTI is returning to track
- ◊ PAULA is continuing her journey.

- ◊ The predictive display shows MAIJA's choice for the east side of Harmaja pilot station. The vector display does not show this choice yet.
- ◊ The predictive display shows TARU turning south.

Scenario 2 Break 3

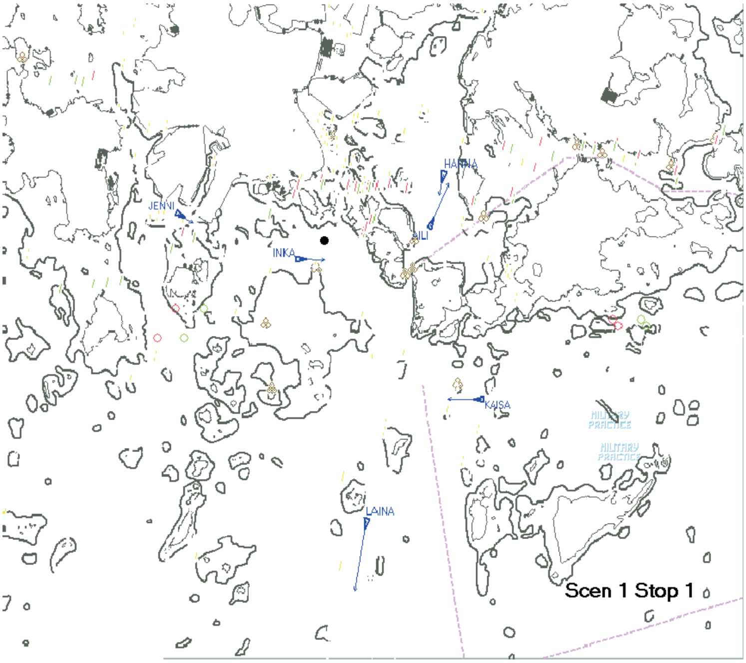
(Points of particular interest are marked *)

- ◊ OUTI is back on track and continuing her journey.
- ◊ PAULI* and SALLI* are passing
- ◊ RAILI* has to make a choice of route on which side to pass Harmaja pilot station.
- ◊ TARU is continuing her journey
- ◊ MAIJA has just passed Harmaja pilot station.

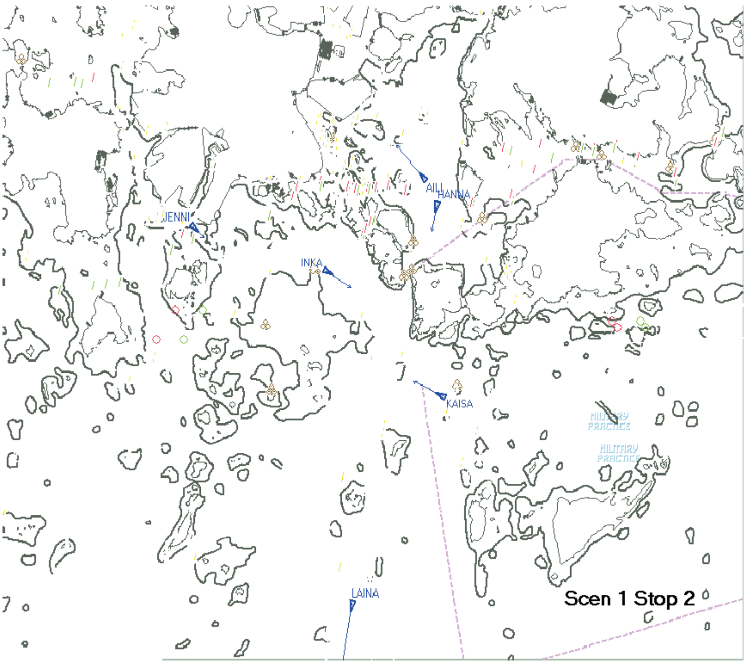
- ◊ The predictive display shows the choice of route of RAILI

Normative pictures of traffic at breaks

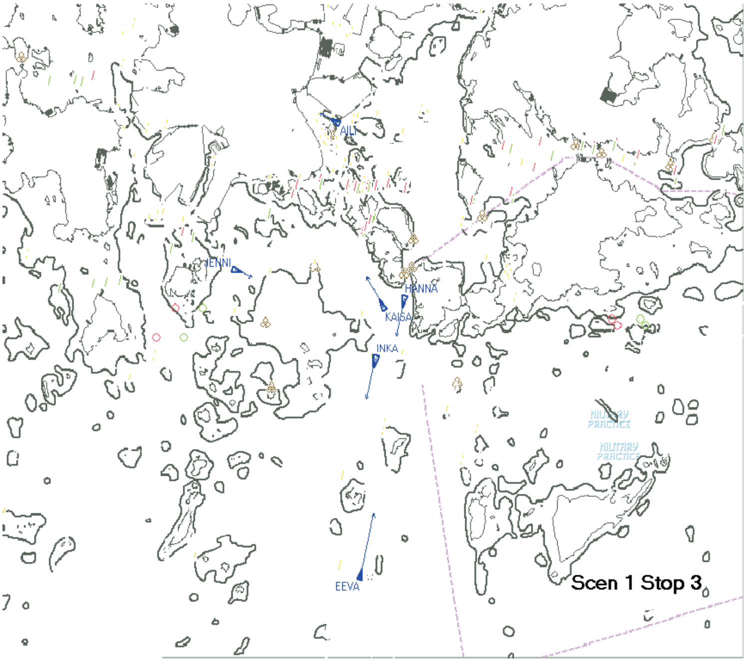
Scenario 1 break 1



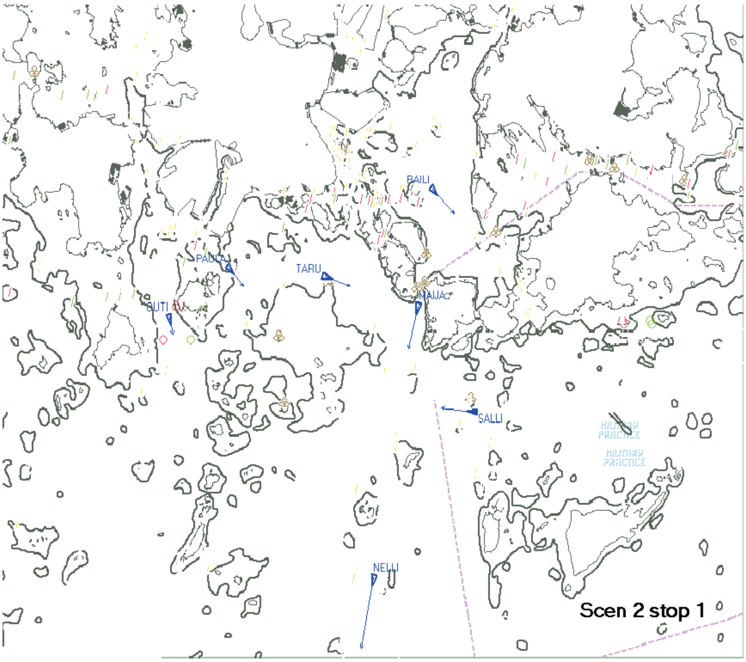
Scenario 1 break 2



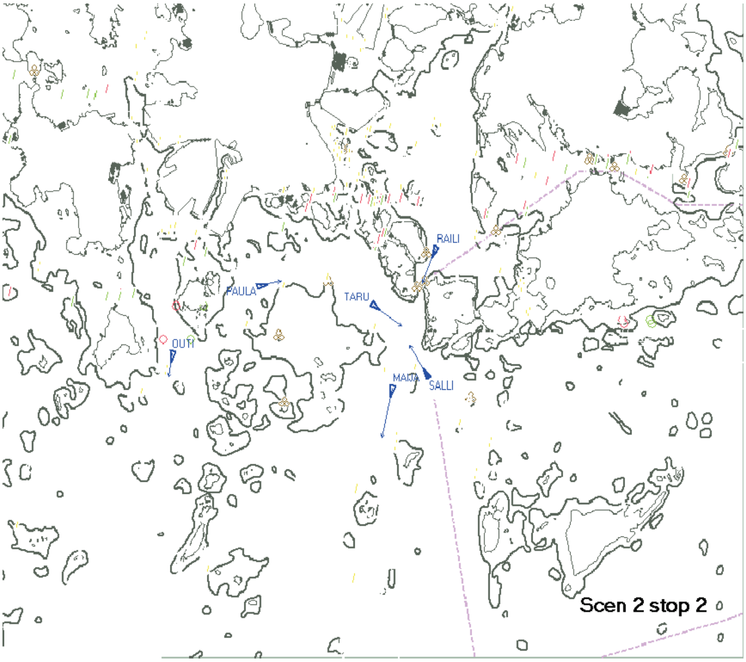
Scenario 1 break 3



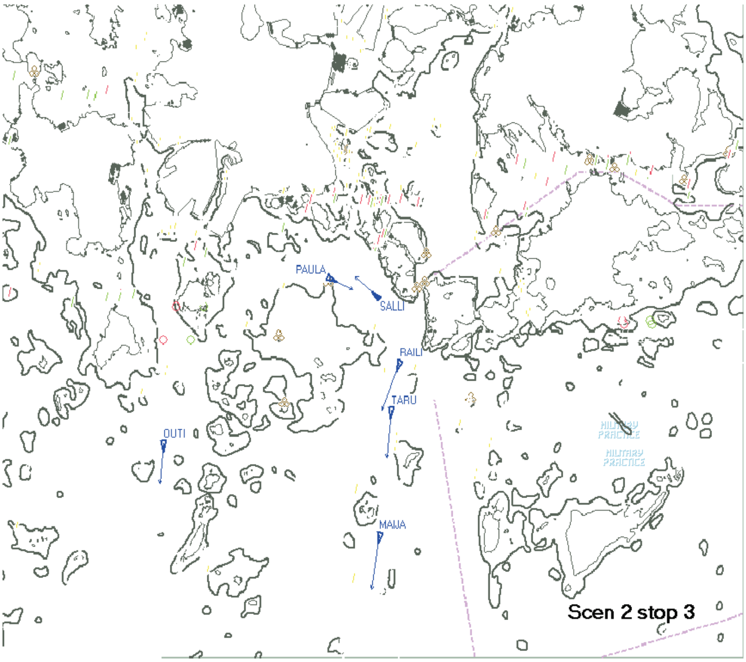
Scenario 2 break 1



Scenario 2 break 2



Scenario 2 break 3



Schedule at breaks

Operator	Test 1	Test 2
1, 5	Scenario 1 Path Prediction	Scenario 2 Vectors
2, 6	Scenario 2 Path Prediction	Scenario 1 Vectors
3, 7	Scenario 1 Vectors	Scenario 2 Path Prediction
4, 8	Scenario 2 Vectors	Scenario 1 Path Prediction

ANNEX 3

SATEST operator scores

Rotterdam Tables of SATEST Results

Actual number of ships scored in different trials

	Minute									
Operator	**5**	**8**	**10**	**14**	**17**	**20**	**35**	**37**	**40**	**45**
N1	10					14		16		15
N2	9					13		15		14
N3	8					13		15		15
E1	10					13		18		13
E2	13					12		17		13
N4		11		12					17	12
E3		11		12					18	13
E4		10		11					18	14
E5		10		11					15	16
E6		8		12					11	12
E7			8		12		17			12
E8			11		14		18			13
E9			10		12		14			12
E10			11		14		17			14
E11			9		13		18			15
Average	10	10	9,8	11,6	13	13	16,8	16,2	15,8	13,5
Actual	(10)	(11)	(11)	(12)	(14)	(16)	(18)	(18)	(17)	(14)

Helsinki Tables of SATEST results

Actual number of ships scored in different trials

Scenario 1 break 1	Max	Operators with predictor	Operators without predictor
Total number of ships in scenario	6	6/6/6	6/5/6
Ships that need extra attention in scenario	2	2/2/2	2/2/2
Ships that do not need extra attention in scenario	4	4/4/4	4/3/4
Total number of ships placed correct	6	2/3/1	3/0/1
Ships that need extra attention placed correct	2	1/2/0	1/0/0
Ships that do not need extra attention placed correct	4	1/1/1	2/0/1
Total number of ships placed on track	6	5/4/5	6/4/5
Ships that need extra attention placed on track	2	2/0/2	2/2/1
Ships that do not need extra attention placed on track	4	3/4/3	4/2/4
Total points of interest	1	0/1/1	1/0/0
Points of interest with advantage from prediction	1	0/1/1	1/0/0
Points of interest without advantage from prediction	0	-/-/-	-/-/-

Scenario 1 break 2	Max	Operators with predictor	Operators without predictor
Total number of ships in scenario	6	6/6/6	6/6/6
Ships that need extra attention in scenario	3	3/3/3	3/3/3
Ships that do not need extra attention in scenario	3	3/3/3	3/3/3
Total number of ships placed correct	6	0/0/2	3/0/2
Ships that need extra attention placed correct	3	0/0/1	3/0/1
Ships that do not need extra attention placed correct	3	0/0/1	0/0/1
Total number of ships placed on track	6	6/5/6	6/6/6
Ships that need extra attention placed on track	3	3/2/3	3/3/3
Ships that do not need extra attention placed on track	3	3/3/3	3/3/3
Total points of interest	1	0/0/0	0/0/0
Points of interest with advantage from prediction	0	-/-/-	-/-/-
Points of interest without advantage from prediction	1	0/0/0	0/0/0

Scenario 1 break 3	Max	Operators with predictor	Operators without predictor
Total number of ships in scenario	6	6/6/6	6/6/6
Ships that need extra attention in scenario	3	3/3/3	3/3/3
Ships that do not need extra attention in scenario	3	3/3/3	3/3/3
Total number of ships placed correct	6	2/1/3	2/1/3
Ships that need extra attention placed correct	3	1/1/1	2/0/1
Ships that do not need extra attention placed correct	3	1/0/2	0/1/2
Total number of ships placed on track	6	6/6/6	5/6/6
Ships that need extra attention placed on track	3	3/3/3	3/3/3
Ships that do not need extra attention placed on track	3	3/3/3	2/3/3
Total points of interest	3	2/1/3	2/0/1
Points of interest with advantage from prediction	2	1/1/1	1/0/0
Points of interest without advantage from prediction	1	1/0/1	1/0/1

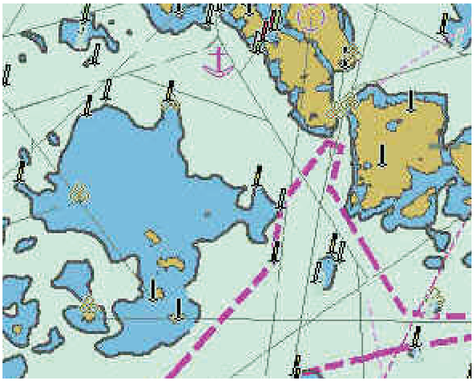
Scenario 2 break 1	Max	Operators with predictor	Operators without predictor
Total number of ships in scenario	7	7/6/7	7/7/7
Ships that need extra attention in scenario	4	4/4/4	4/4/4
Ships that do not need extra attention in scenario	3	3/2/3	3/3/3
Total number of ships placed correct	7	3/1/3	1/4/5
Ships that need extra attention placed correct	4	2/1/1	0/3/3
Ships that do not need extra attention placed correct	3	1/0/2	1/1/2
Total number of ships placed on track	7	7/6/7	7/7/7
Ships that need extra attention placed on track	4	4/4/4	4/4/4
Ships that do not need extra attention placed on track	3	3/2/3	3/3/3
Total points of interest	2	0/1/0	1/1/0
Points of interest with advantage from prediction	1	0/1/0	1/1/0
Points of interest without advantage from prediction	1	0/0/0	0/0/0

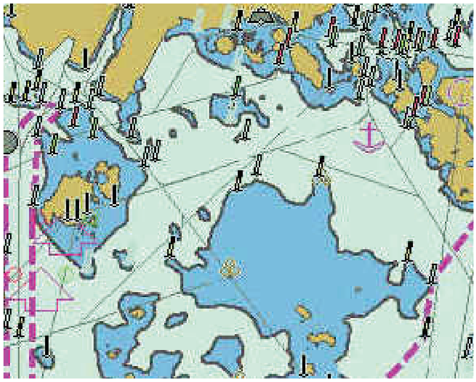
Scenario 2 break 2	Max	Operators with predictor	Operators without predictor
Total number of ships in scenario	6	6/5/6	6/6/6
Ships that need extra attention in scenario	4	4/4/4	4/4/4
Ships that do not need extra attention in scenario	2	2/1/2	2/2/2
Total number of ships placed correct	6	1/1/2	1/0/4
Ships that need extra attention placed correct	4	0/0/1	0/0/3
Ships that do not need extra attention placed correct	2	1/1/1	1/0/1
Total number of ships placed on track	6	5/4/4	5/4/5
Ships that need extra attention placed on track	4	3/3/3	4/3/4
Ships that do not need extra attention placed on track	2	2/1/1	1/1/1
Total points of interest	2	2/2/1	1/1/1
Points of interest with advantage from prediction	2	2/2/1	1/1/1
Points of interest without advantage from prediction	0	-/-/-	-/-/-

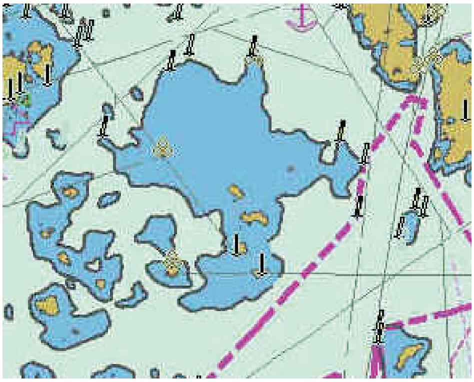
Scenario 2 break 3	Max	Operators with predictor	Operators without predictor
Total number of ships in scenario	6	6/5/6	6/6/6
Ships that need extra attention in scenario	3	3/3/3	3/3/3
Ships that do not need extra attention in scenario	3	3/2/3	3/3/3
Total number of ships placed correct	6	0/0/2	4/2/4
Ships that need extra attention placed correct	3	0/0/1	2/1/3
Ships that do not need extra attention placed correct	3	0/0/1	2/1/1
Total number of ships placed on track	6	4/4/5	6/4/6
Ships that need extra attention placed on track	3	2/2/2	3/2/3
Ships that do not need extra attention placed on track	3	2/2/3	3/2/3
Total points of interest	2	1/0/0	1/0/1
Points of interest with advantage from prediction	1	0/0/0	0/0/0
Points of interest without advantage from prediction	1	1/0/0	1/0/1

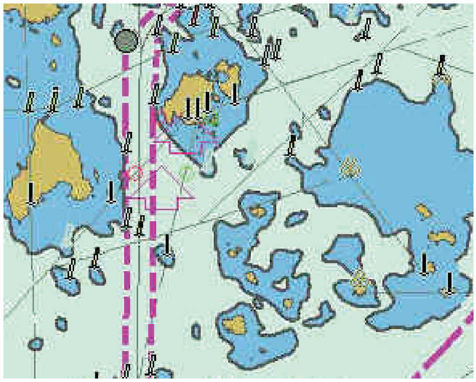
Total	**Max correct**	**Operators correct percentage**		
		with predictor	**without predictor**	**All**
Total number of ships per scenario	6.17	97.24	99.03	98.22
Ships that need extra attention per scenario	3.17	100.00	100.00	100.00
Ships that do not need extra attention per scenario	3.00	94.33	98.0	96.33
Total number of ships placed correct	6.17	24.31	35.17	30.63
Ships that need extra attention placed correct	3.17	22.71	38.49	30.60
Ships that do not need extra attention placed correct	3.00	26.00	31.33	28.67
Total number of ships placed on track	6.17	85.58	90.92	88.17
Ships that need extra attention placed on track	3.17	84.23	94.64	89.27
Ships that do not need extra attention placed on track	3.00	87.00	87.00	87.00
Total points of interest	1.83	45.36	33.33	39.34
Points of interest with advantage from prediction	1.17	52.14	33.33	42.74
Points of interest without advantage from prediction	0.67	32.84	32.84	32.84

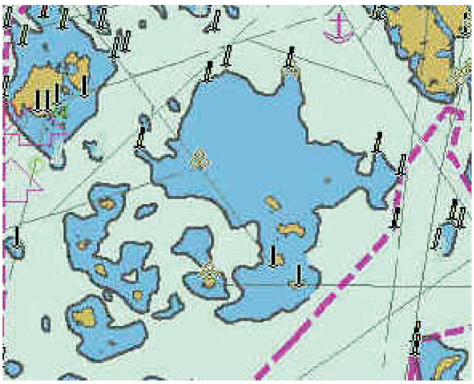
Helsinki Attention at breaks

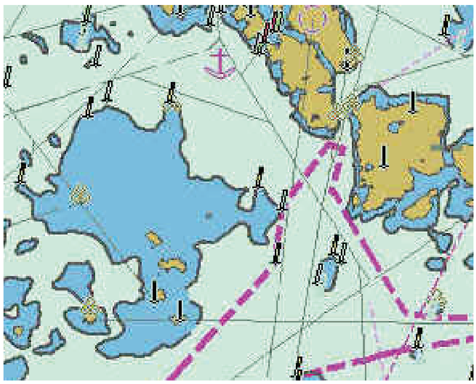

Operator 1 Break 1

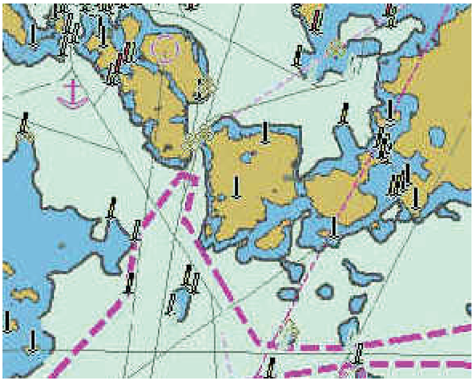

Operator 1 Break 2

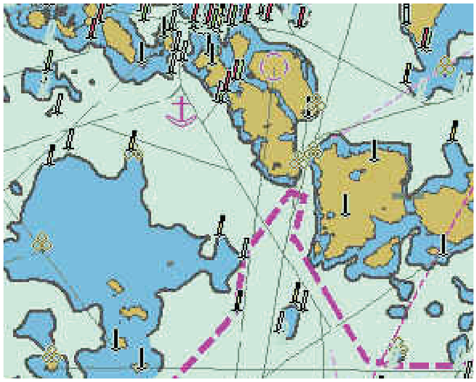

Operator 1 Break 3

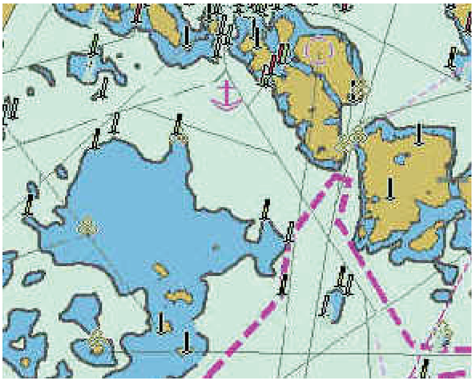

Operator 1 Break 4

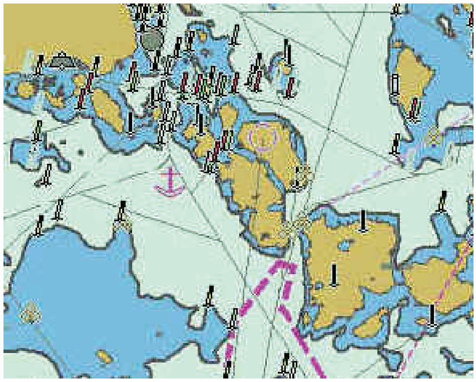

Operator 1 Break 5

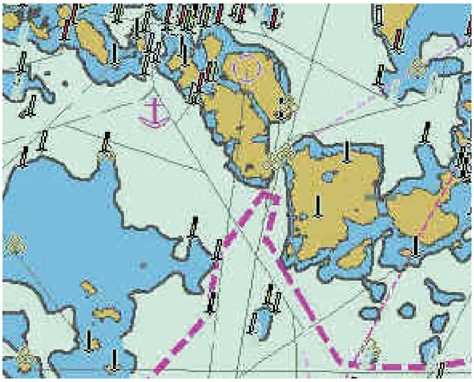

Operator 1 Break 6

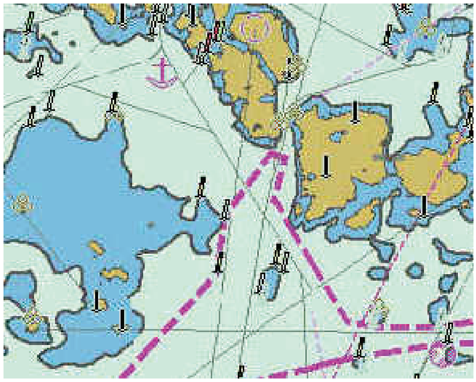

Operator 2 Break 1

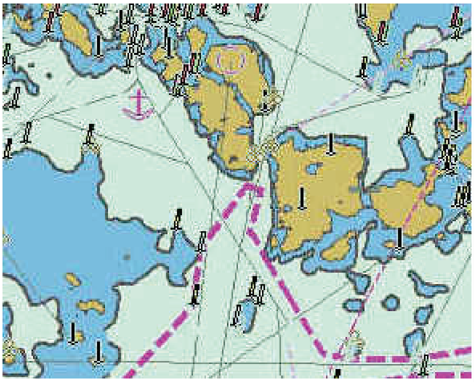

Operator 2 Break 2

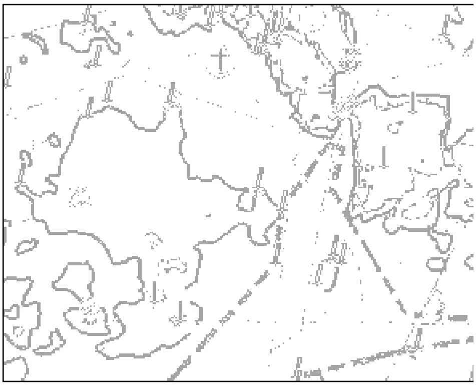

Operator 2 Break 3


Operator 2 Break 4

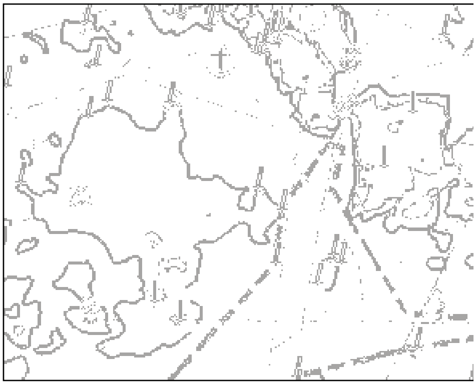

Operator 2 Break 5


Operator 2 Break 6



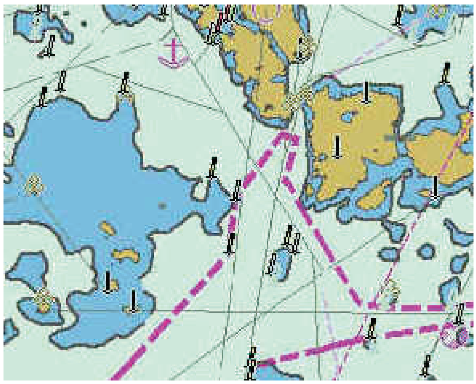


Operator 3 Break 1



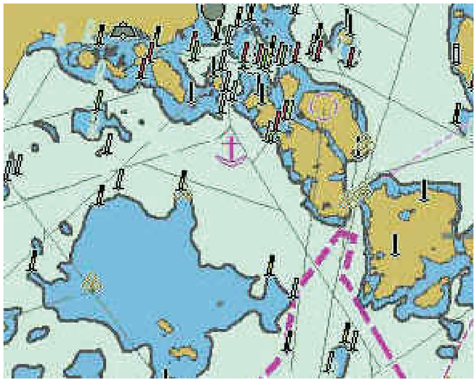
Operator 3 Break 4

Operator 3 Break 5

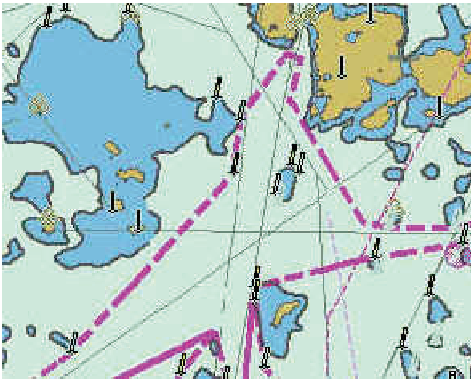


Operator 3 Break 6

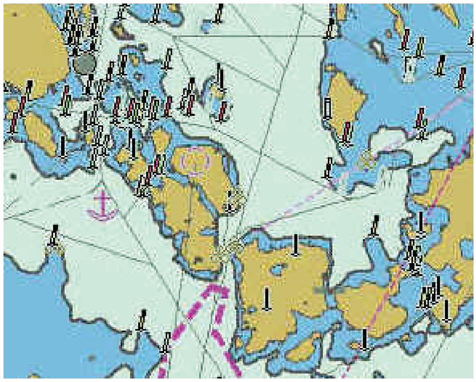
Operator 4 Break 1



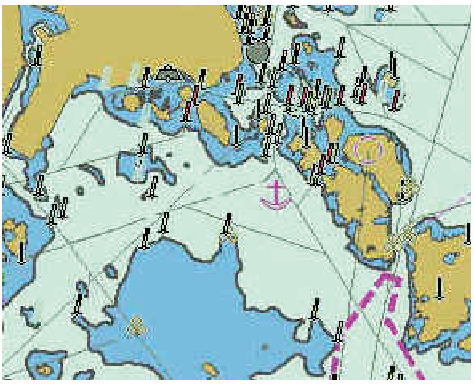
Operator 4 Break 2



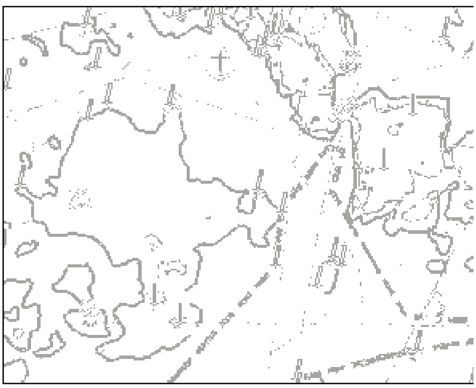
Operator 4 Break 3

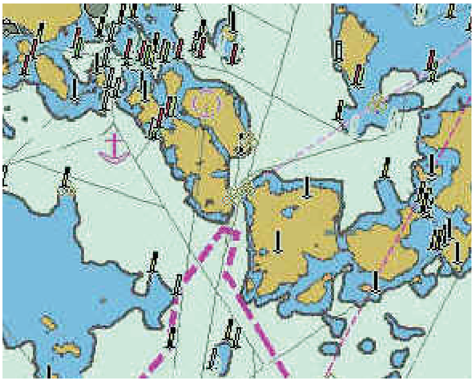


Operator 4 Break 4

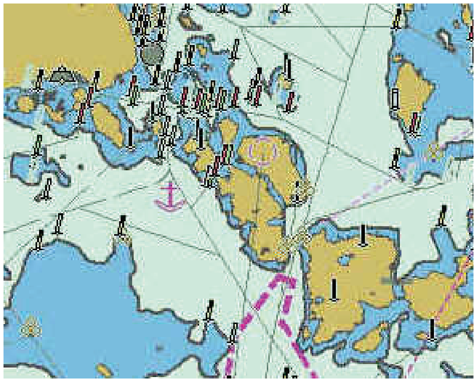


Operator 4 Break 5

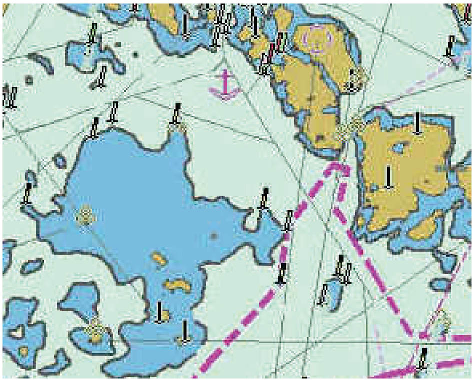




Operator 5 Break 1

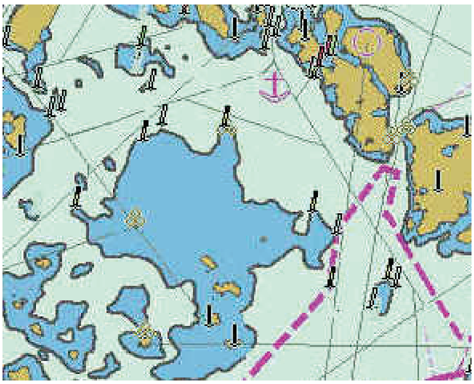


Operator 5 Break 2

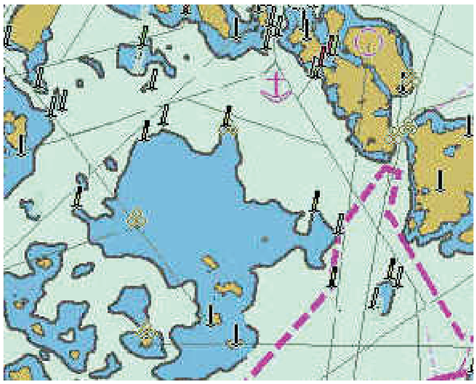


Operator 5 Break 3

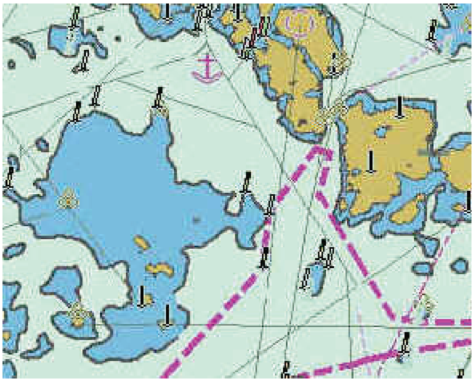
Operator 5 Break 4

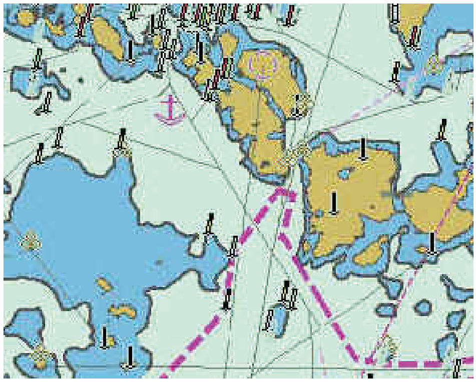


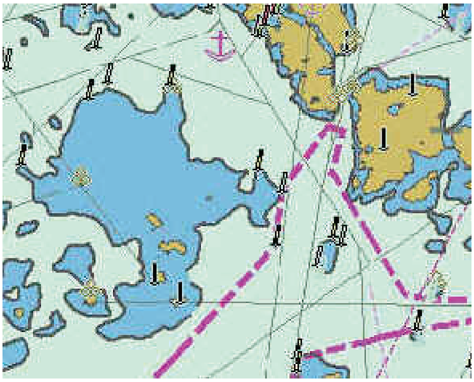
Operator 5 Break 5

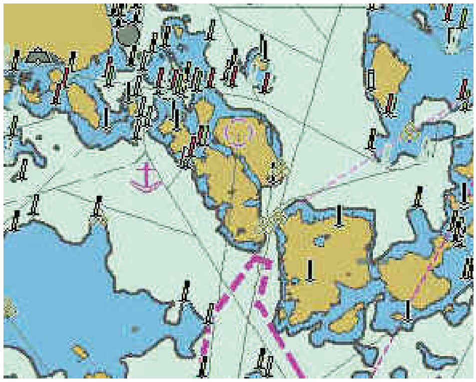


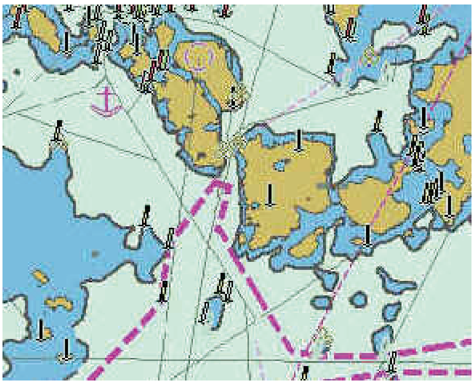
Operator 5 Break 6

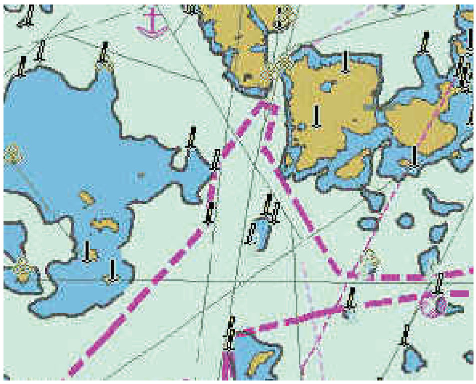

Operator 6 Break 1


Operator 6 Break 2


Operator 6 Break 3


Operator 6 Break 4


Operator 6 Break 5


Operator 6 Break 6

ANNEX 4

PMI protocols and score forms

SHIP		LOA	BRT	ROUTE	
BLUE SEA	Z	169	14515	GEMEERD ESSO	01
BONGO	Z	70	1133	INKOMEND NG-9 MOERMANSVAARW	02
CONDOR 1 + BAK	B	89	1800	SPUIHAVEN - PLEZIERHAVEN	03
CORNELIA	Z	112	5356	ZUIGEN NG/ZG BOEI - STORT ZEE ZG	04
DONGESTROOM	B	90	2105	SHELL-1 - ACHTERLAND (*spuihaven !*)	05
HAVLYS	Z	140	9367	INKOMEND ZUIDGEUL - TOTAL	06
H.D.17	P	34	50	GEMEERD WERKHAVEN	07
HOO PRIDE	Z	58	300	UITG. MMVW - ZEE/N-GEUL (ENGELS)	08
KOBULETI	Z	151	10937	CHEVRON 1 - ZEE	09
LILIANA	Z	96	3019	INKOMEND HANDELSHAVEN	10
PRINS MAURITS	B	95	1693	GEMEERD NRDZHVN ELF	11
SHELL-6	B	90	2154	VERH. SHELL-2 - WERKHAVEN	12
SMIT DENEMARKEN	S	29	193	ASS. KOBULETI (VOOR) - - SLBHVN	13
SMIT FINLAND	S	29	193	ASS. KOBULETI (ACHTER) - - SLBHVN	14
SZYMANOWSKI	Z	169	22130	INK. ZG - HANDELSHAVEN	15
TANGER	Z	120	5371	INK. NOORDGEUL - HDHVN. ZUIDPIER	16
VANUDEN 22	B	84	1599	GEMEERD NRDZHVN ELF	17
WATERBOOT 12	B	37	320	VERH. WERKHVN - ESSOSTEIGER	18
YEHYA	Z	70	1598	UITG. LEDIG, CENTRALE - ZEE / NG	19
FUTURO	Y	18	100	VERTREK PLEZHVN - OPVAART	20
MARIA	B	50	360	AFVAREND - MMVW	21
ALVRACHT 10	B	80	1252	VERHALEN KMR1009 - STEENSTEIGER	22
IMKE	B	50	434	VERHALEN ESSO (kade) - TOTAL	23
LAMMY	B	38	336	UITG. MMVW - HDHVN	24
DUTCH NAVIGATOR	Z	80	1410	GEMEERD NZH GULF	25

Wanneer zowel de tijd als de corresponderende regel **VET** gedrukt staan, geeft dit de **'trigger'** aan zoals die door het team tijdens de test wordt gegenereerd.
In de kolom **TIJD** staan behalve de minuten ook nog de volgende aanwijzingen:
w = **'waarneming'**, hetgeen wil zeggen dat er gebeurtenissen zijn waar te nemen, zoals kort omschreven in de corresponderende regel, waarop een actie/reactie van de kandidaat kan worden verwacht;
S = Standaard Procedure Melding
A1 = een **'actie/reactie'** die minimaal, maar ZEKER dient te worden genomen door de kandidaat;
A2 = een **'actie/reactie'** die als aanvulling KAN worden gegeven op A1;
tel = een **'actie/reactie'** afgehandeld per telefoon;
zicht = een zichtbericht verschijn per imperatieve melding op het scherm;
ook combinaties kunnen voorkomen.

TIJD	**FASE 0** **Bij aanvang van de run zijn er twee varende doelen binnen de sectorgrenzen; één vaartuig is bezig met werkzaamheden in de sector; één zeevaartuig op vertrek (met sleepbootassitentie); 7 binnenschepen, een jacht en een patrouilleboot afgemeerd en twee zeeschepen afgemeerd. Binnen de sectorgrenzen varen nog enkele onbekende binnenvaartschepen en een jacht, waarmee geen communicatie is geweest.Het IVS is ingevuld, met dien verstande dat de binnenvaart niet wordt vermeld, hydro en meteo gegevens en nautische bijzonderheden zijn**				
0	CORNELIA zuig-werkzaamheden ten noorden van de scheidingsboei NG/ZG				
0	KOBULETI op vertrek chevron 1 naar zee met 2 sleepboten				
0	LILIANA inkomend naar handelshaven t.h.v. HV-1 boei				
0	BONGO inkomend t.h.v. ZG-9 boei naar achterland via moermansvaarwater				
0	MARIA onbekende afvarende binnenvaart kmr. 1012				
zicht	2500 meter				

TIJD	**FASE 1** **Het op vertrek liggende zeeschip gaat met behulp van sleepbootassistentie aan het verkeer deelnemen; twee van de gemeerde binnenschepen en het jacht gaan aan het verkeer**				
1	**KOBULETI los en varend**				
S	Begrepen, melden binnen de monding				
01-w	***NIET MELDEN* FUTURO (jacht) los in plezierhaven**				
1	**YEHYA afvarend ledig naar zee (NG)**				
A1	binnenvaart (MARIA) stuurboord voor, geen verbinding mee				
A1	uitgaande zeevaart (KOBULETI), vertrokken van chevron-1 naar zee				
A2	eerste tegenliggende zeevaart (LILIANA) t.h.v. HV1-boei, door naar boven				
A2	zuigerwerkzaamheden (CORNELIA) t.h.v. scheidingsboei Ngeul/Zgeul				
02-w	***NIET MELDEN* FUTURO (jacht) verlaat plezierhaven naar boven, geen verbinding**				
02.30	**DONGESTROOM los Shell-1 naar achterland**				
A1	afvarende binnenvaart (MARIA) 2500 meter boven de monding NZH, geen verbinding mee				
A1	afvarende coaster (YEHYA) 3000m boven de monding NZH bestemming zee				
A2	informatie: vertrekkende KOBULETI los chevron naar zee				
A1	vertrek DONGESTROOM terugkoppelen met KOBULETI				

A2	opvarende coaster (LILIANA) recht vooruit bestemming HDH				
03.30	**WATERBOOT 12 vertrekt van de werkhaven, bestemming Esso steiger binnenkant**				
A2	zuiger (CORNELIA) werkzaam t.h.v. NG7/ZG14-boei				
A2	vertrekkende zeevaart NZH van de chevron (KOBULETI)				
w	eerste afvaart aan de spuihaven (YEHYA)				
03.30-w	***NIET MELDEN*** **afvarende MARIA blijft aan de Nzijde boeienlijn**				
430	**CORNELIA volle bak, storten, naar zee via ZG**				
A2	waterboot 12 draait de werkhaven uit naar boven				
TIJD	**FASE 2** **De twee vertrekkende binnenschepen komen vanuit nevenvaarwater/haven de rivier op; voormelding van vertrek van gemeerd liggend binnenschip; moet vaarwater oversteken wanneer vertrokken; één Engels sprekend zeeschip komt de sector binnen, uitgaand naar zee**				
5	**DONGESTROOM in de monding**				
A1	tweemaal afvaart 2200 meter boven de O-kop NZH 1x binnenvaart (MARIA) 1x zeevaart (YEHYA)				
A2	terugkoppelen met YEHYA				
w-zicht	2000 meter				

6	***VOORMELDING*: CONDOR 1 met bakje van de Spuihaven oversteken naar de plezierhaven**				
A1	opvarende zeevaart (LILIANA) 2500 meter beneden de spuihaven bestemming HDH				
A1	info terugkoppelen met LILIANA				
S	melden als u ontmeerd bent				
w	(FUTURO) onbekend jacht nevenvaarwater richting achterland				
06-w	***NIET MELDEN*** **WATERBOOT 12 is op de rivier, blijft aan noordzijde varen tot NG10**				
07-w	***NIET MELDEN*** **DONGESTROOM op rivier / CORNELIA draait Zgeul in / HOO PRIDE afv MMVW 500 mtr boven brug**				
8	**KOBULETI passeert Shell 1 op weg naar zee**				
A1	afvarende zeevaart (YEHYA) 1500 mtr. boven de O-kop NZH,				
A1	onbekende (binnenvaart MARIA) daar 300 mtr achter,				
A1	1x zeevaart opvarend 2000mtr beneden de Wkop NZH bestemming MMVW				
A1	KOBULETI terugkoppelen met YEHYA				
A1	KOBULETI terugkoppelen met BONGO				
w-zicht	**1500 meter**				

08.30-w	***NIET MELDEN*** **CONDOR 1(+duwbak) los en varend in de spuihaven, meldt zich niet en geeft geen antwoord !!**				
9	**HOO PRIDE below the bridge in moermanvaarwater, outward to sea**				
A1	1x zeevaart met bestemming MMVW, 2000m beneden Wkop MMVW				
A2	terugkoppelen met BONGO				
S	report with the entrance				
w-zicht	**1000 meter**				
TIJD	**FASE 3Het vertrekkende zeeschip met sleepbootassistentie komt vanuit de haven de rivier op; twee inkomende zeeschepen komen binnen de sectorgrenzen, één bestemming in de sector, één door naar volgende sector; het vertrekkende binnenschip steekt rivier over wordt koerskruiser voor overige vaart; één gemeerd liggend binnenschip gaat deelnemen aan het verkeer.**				
10.30	**HAVLYS inkomend zuidgeul naar Noorzeehaven Total**				
A1	tegenliggende zeevaart BB-voor (CORNELIA) op weg naar zee				
A2	vertrekkende zeevaart monding NZH bestemming zee				
A2	vanuit de Ngeul nadert inkomende zeevaart (TANGER) nog niet gehoord				
11-w	***NIET MELDEN!*** **CONDOR 1 in monding Spuihaven en geeft geen antwoord**				

Assessmentformulier Debriefing

kandidaat:
datum:
assessor:

Aandragen van oplossingen/improviseren

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Komt met adequate oplossingen voor voorgelegde problemen	Komt niet met adequate oplossingen voor voorgelegde problemen

Overzicht over het geheel

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Begeleidt afzonderlijke schepen expliciet rekening houdend met de interacties met het overige verkeer	Begeleidt afzonderlijke schepen zonder expliciet rekening houdend met het overige verkeer

Prioriteiten stellen

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Pakt complexe situaties aan door de belangrijkste problemen het eerst te behandelen	Maakt in complexe situaties geen onderscheid tussen het belang van problemen

Stemgebruik

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Stemgebruik duidelijk en overtuigend	Stemgebruik onduidelijk / niet overtuigend

Flexibiliteit

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Gaat adequaat om met plotselinge veranderingen in complexe situaties	Gaat niet adequaat om met plotselinge veranderingen in complexe situaties

Stressbestendigheid

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Blijft kalm en effectief handelen in complexe situaties	Handelt niet kalm en ineffectief in complexe situaties

Besluitvaardigheid

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Gaat in complexe situaties snel tot handelen over	Gaat in complexe situaties traag tot handelen over

Motivatie

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Geeft er blijk van een goede prestatie tijdens de simulatie te willen neerzetten	Lijkt er niet op gericht een goede prestatie tijdens de simulatie te willen neerzetten

Algehele effectiviteit

1 = goed 2 = voldoende 3 = matig 4 = onvoldoende 5 = ?

Weet tijdens de gehele simulatie de balans tussen veiligheid en effecientie te bewaren	Weet tijdens de gehele simulatie niet de balans tussen veiligheid en effecientie te bewaren

ANNEX 5

Improved PMI-P score form

EVENT	Time	Type	Text	Check	Early	Late
			FASE 1 **Het op vertrek liggende zeeschip gaat met behulp van sleepbootassistentie aan het verkeer deelnemen; twee van de gemeerde binnenschepen en het jacht gaan aan het verkeer deelnemen; één 'doorgaand' zeeschip komt vanuit naastliggende sector binnen de sectorgrenzen**			
1	**0:01:00**	**CALL**	KOBULETI los en varend			
		Standard	Begrepen, melden in de monding			
2	**0:01:00**	**Observe**	NIET MELDEN FUTURO (jacht) los in plezierhaven			
3	**0:01:00**	**CALL**	YEHYA afvarend ledig naar zee (NG)			
		A1	binnenvaart (MARIA) stuurboord voor, geen verbinding mee			
		A2	uitgaande zeevaart (KOBULETI), vertrokken van chevron-1 naar zee			
		B	eerste tegenliggende zeevaart (LILIANA) t.h.v. HV1-boei, door naar boven			
		B	zuigerwerkzaamheden (CORNELIA) t.h.v. scheidingsboei Ngeul/Zgeul			
4	**0:02:00**	**Observe**	NIET MELDEN FUTURO (jacht) verlaat plezierhaven naar boven, geen verbinding			
5	**0:02:30**	**CALL**	DONGESTROOM los Shell-1 naar achterland			
		A1	afvarende binnenvaart (MARIA) 2500 meter boven de monding NZH, geen verbinding mee			
		A2	afvarende coaster (YEHYA) 3000m boven de monding NZH bestemming zee			
		B	informatie: vertrekkende KOBULETI los chevron naar zee			
		A2	vertrek DONGESTROOM terugkoppelen met KOBULETI			
		B	opvarende coaster (LILIANA) recht vooruit bestemming HDH			
6	**0:03:30**	**CALL**	WATERBOOT 12 vertrekt van de werkhaven, bestemming Esso steiger binnenkant			
		B	zuiger (CORNELIA) werkzaam t.h.v. NG7/ZG14-boei			
		B	vertrekkende zeevaart NZH van de chevron (KOBULETI)			
		Observe	eerste afvaart aan de spuihaven (YEHYA)			
7	**0:03:30**	**Observe**	NIET MELDEN afvarende MARIA blijft aan de Nzijde boeienlijn			
8	**0:04:30**	**CALL**	CORNELIA volle bak, storten, naar zee via ZG			
		B	waterboot 12 draait de werkhaven uit naar boven			
			FASE 2 **De twee vertrekkende binnenschepen komen vanuit nevenvaarwater/haven de rivier op; voormelding van vertrek van gemeerd liggend binnenschip; moet vaarwater oversteken wanneer vertrokken; één Engels sprekend zeeschip komt de sector binnen, uitgaand naar zee**			
9	**0:05:00**	**CALL**	DONGSTROOM in de monding			
		A1	2x afvaart 2200 meter boven de O-kop NZH 1x binnenvaart (MARIA) 1x zeevaart (YEHYA)			
		B	terugkoppelen met YEHYA			

Summary

This thesis describes my study of situation awareness assessment of Vessel Traffic Service (VTS) operators. VTS operators are the traffic controllers on the water. They are responsible for a safe and efficient handling of vessel traffic. They monitor traffic, provide information on request and coordinate movement of ships in (emerging) conflict situations in harbours, rivers, and approach areas all around the world. Situation awareness is a term describing a person's internal mental model of the current situation. Situation awareness is a concept that has been applied with success in many domains where controllers need to maintain an overview of their work.

The concept of situation awareness presents new perspective for the study of VTS operator performance by investigating the mental picture of the operator. This perspective may lead to better understanding of the performance of VTS operators. The main objective of this thesis is to study how situation awareness of VTS operators can be assessed and to determine if the use and the assessment of VTS operator situation awareness leads to a better understanding of the performance of VTS operators.

The thesis presents a framework for the examination of the concept situation awareness and for methods for assessing it in the context of VTS operator performance. This evaluation is based upon a literature study on situation awareness and VTS research and upon observations on VTS posts in the Netherlands. Two methods to assess situation awareness are developed.

The first method, called SATEST, is derived from the SAGAT method developed by Mica Endsley and can be considered a traditional method for measuring situation awareness. This method presents a VTS scenario in an experimental setting. At unannounced, but previously specified times during these scenarios the system displays are blanked and the simulation is stopped. The operators are requested to answer a number of questions about

their perception of the situation at that time, designed to reveal their situation awareness. After completion of the query, the simulation continues.

SATEST was developed, tested and used in two experiments; in Rotterdam and Helsinki in the COMFORTABLE project. These experiments showed that SATEST was very useful in providing understanding of the VTS-work, the concept of situation awareness and its application in a VTS context. They also revealed the limitations of the method. It became clear in working with SATEST that the method covers situation awareness at any particular moment very well, but it is not very suitable to observe how situation awareness develops in a situation. The conclusion of the work with SATEST was that situation awareness is a useful concept in describing the work of the VTS operator, but that the method for assessing situation awareness needed more work.

In the discussions of SATEST the idea grew that communication might be used to analyse when operators become aware of conflicts and when they will solve them. The second method, called PMI-P (Performance Measuring Instrument – Performance module), uses the assessment of communication to assess performance and situation awareness. Communication is scored on a PMI-P score form. This score form was constructed using the concept of Required Situation Awareness (RSA), developed in this thesis. RSA refers to the situation awareness needed to correctly handle a situation. For every event in the scenario the PMI-P score form contains information that the operator needs to communicate, and additional information that an operator may want to communicate. Communication between VTS and traffic is regarded as a direct measure of VTS operator performance and an indirect measure of situation awareness. The PMI-P score form captures the development of VTS operator situation awareness by describing all possible communications that an operator can have during the scenario run.

This method was used in a project for the Rotterdam Port Authority. The results show that PMI-P is a powerful tool for assessing VTS operator performance. Understanding communication in PMI-P provides a solid handle on the assessment of VTS operator performance. This leads us to the question whether situation awareness is still needed as a concept in the assessment. The short answer to this question is yes: The required situation awareness of a situation that an operator requires has been used in the development of PMI-P to understand the scenario and to capture the required communication.

Together the two methods offer a wide range of opportunities to study situation awareness of VTS operators, to study impact of new technology on VTS operator situation awareness, and to training operators.

Samenvatting

Dit proefschrift beschrijft mijn onderzoek naar het meten van situation awareness van *Vessel Traffic Service (VTS) operators*. VTS operators zijn de verkeersbegeleiders op het water. Zij zijn verantwoordelijk voor een veilige en efficiënte afhandeling van het scheepvaartverkeer. Zij monitoren het verkeer, verstrekken op verzoek informatie en coördineren de bewegingen van de schepen in (opkomende) conflictsituaties in havens, op rivieren en naderingsgebieden over de hele wereld. Situation awareness is een term die het interne mentale model beschrijft dat een persoon heeft van de actuele situatie. Situation awareness is een concept dat met succes is toegepast in verscheidene domeinen waar controllers overzicht moeten houden over hun werk.

Het concept situation awareness biedt een nieuw perspectief voor de studie van de prestatie van VTS operators, door het mentale beeld van de operator te bestuderen. Dit perspectief kan leiden tot een beter begrip van de prestatie van operators. Het doel van dit proefschrift is om te onderzoeken hoe situation awareness van operators kan worden gemeten en om te bepalen of het gebruik en het meten van situation awareness leiden tot een beter begrip van de prestatie van operators.

De thesis presenteert een raamwerk voor het onderzoeken van het concept situation awareness en van methoden om het te meten in de context van prestaties van VTS operators. Deze evaluatie is gebaseerd op een literatuurstudie naar situation awareness en VTS onderzoek, en op observaties op VTS posten in Nederland. Twee methoden om situation awareness te meten worden ontwikkeld.

De eerste methode, SATEST geheten, is afgeleid van de SAGAT method, ontwikkeld door Mica Endsley, en kan worden beschouwd al seen traditionele methode voor het meten van situation awareness. Deze methode presenteert een VTS scenario in een experimentele zetting. Op onaangekondigde, maar vooraf bepaalde momenten tijdens het scenario worden de schermen zwart en wordt de simulator stil gezet. Operators wordt gevraagd om een aantal vragen te

beantwoorden over hun waarneming van de situatie op dat moment, ontworpen om hun situation awarenes te laten zien. Na het invullen van de vragen gaat de simulatie verder.

SATEST is ontwikkeld, getest en gebruikt in twee experimenten; in Rotterdam en in Helsinki in het COMFORTABLE project. Deze experimenten lieten zien dat SATEST zeer bruikbaar was voor het begrijpen van het VTS-werk, het concept situation awareness en de toepassing daarvan in een VTS context. Ze lieten ook de beperkingen van de method zien. Door het werken met SATEST werd het duidelijk dat de methode situation awareness goed dekt op elk specifiek moment, maar dat zij niet erg geschikt is om te observeren hoe situation awareness zich ontwikkelt in een situatie. De conclusie van het werken met SATEST is dat situation awareness een bruikbaar concept is voor het beschrijven van het werk van de VTS operator, maar dat de method voor het meten van situation awareness meer werk behoeft.

Tijdens de discussies over SATEST groeide het idee dat mogelijk communicatie kon worden gebruikt om te analyseren wanneer operators zich bewust worden van conflicten en wanneer zij ze willen oplossen. De tweede methode, PMI-P (Prestatie Meet Instrument – Prestatie module) genoemd, gebruikt de meting van communicatie voor het vaststellen van de prestatie en situation awareness. Communicatie wordt gescoord op een PMI-P scoreformulier. Bij het samenstellen van dit formulier is gebruik gemaakt van het concept Required Situation Awareness (RSA), ontwikkeld in dit proeschrift. RSA verwijst naar de situation awareness die nodig is om op de juiste manier met een situatie te kunnen omgaan. Voor elke gebeurtenis in het scenario bevat het PMI-P scoreformulier informatie die de operator moet communiceren en aanvullende informatie die een operator wellicht wil communicaeren. Communicatie tussen operator en het verkeer wordt beschouwd als een directe maat van prestatie van een VTS operator en een indirecte maat van situation awareness door het beschrijven van alle mogelijke communicatie die een oeprator kan hebben tijdens een scenario sessie.

Deze methode is gebruikt in een project voor het Havenbedrijf Rotterdam. De resultaten laten zien dat PMI-P een krachtig middel is om de prestatie van VTS operators te meten. Begrip van communicatie verschaft een solide handgreep voor de meting van prestaties van VTS operators. Dit brengt on tot de vraag of situation awareness dan nog steeds nodig is als concept in de meting. Het korte antwoord op deze vraag is ja: de situation awareness die een operator nodig heeft in een situatie is gebruikt in de ontwikkelilng van PMI-P om het scenario en de benodigde communicatie te begrijpen en vast te leggen.

Tezamen bieden de twee methoden een breed scala aan mogelijkheden om situation awareness van VTS operators te bestuderen, om de impact van nieuwe technologie op de situation awareness VTS operators te bestuderen, en om operators te trainen. v

Curriculum vitae

Erik Wiersma was born Zeist, the Netherlands, on December 26, 1960. In 1979 he completed the VWO at the Christelijk College Nassau Veluwe in Harderwijk in 1979. He went on to the "Technische Hogeschool Delft", currently known as Delft University of Technology, to study Industrial Design for two years. In 1981 he switched to the Leiden University to study psychology where he graduated in 1989. In 1990 he returned to the Delft University of Technology to work at the Safety Science Group, where he worked until 2009. Over the years he studied the role of Human Factors in Safety in many domains, and lectured and developed education programs on safety and security.