Assessing the Effectiveness of Naval Ship Automation: Opportunities and Challenges

Dr. ir. A.C. Habben Jansen[1]*, LTZ1 (TD) Ir. T. Stuivenberg[1], Ir. M. Viegers[1]

** Defence Materiel Organisation, Ministry of Defence, The Netherlands*
* Corresponding Author. Email: AC.HabbenJansen@mindef.nl

Synopsis

Automation is of significant interest to the field of naval ship design and operations, resulting in substantial research, development and implementation efforts being carried out in this field. Many of the expected benefits of automation, such as increased situational awareness and reduced operator workload, have a rather abstract, qualitative nature, and cannot easily be measured. This makes it difficult to predict or evaluate the effectiveness of innovation in automation. In order to address this challenge, this paper explores to which extent a need for such effectiveness assessment methods is experienced, and what type of information could be provided by such methods. A literature study into methods that are currently available for assessing the effectiveness of automation shows that most methods focus on either quantitative assessments for measuring specific effects, or qualitative assessments for measuring generic effects. Subsequently, a survey with Subject Matter Experts (SME's) of the Netherlands Ministry of Defence shows that qualitative assessments with user feedback are widely adopted to assess the effectiveness of automation. At the same time, many SME's recognise the subjective nature of this type of information as a challenge. Based on the literature review and the SME survey, the authors conclude that user feedback, though being subjective, remains an adequate method for assessing the effectiveness of automation, but that additional efforts are needed to apply user feedback effectively in automation programmes. More specifically, there should be an adequate mechanism for dealing with user feedback, as well as opportunity and commitment of all parties involved in the collaboration, in order to make innovation naval automation a success.

Keywords: naval ship design, automation, effectiveness metrics, KPI's, SME survey

1. Introduction

Automation and smart ship development is the number three research & development (R&D) topic within the Netherlands maritime industry, after emission-free ships and winning at sea (OECD, 2020). Furthermore, R&D expenses in shipping are on average higher than in many other industries (3,9% of net added value) (Ministerie van Infrastructuur en Milieu, 2015). From a naval perspective, human-machine interaction has been recognised by the U.S. Department of Defence as a critical technology to sustain military dominance (Lopez, 2022). Despite the effort spent on the realisation of automation programmes, limited in-depth research is publicly available into approaches for determining the effectiveness of automation programmes in specific and quantitative metrics or units, both prior to and after completion of the programme. Usually, similar arguments and expected benefits are provided as justification for automation programmes. These include cost savings, reduction of crew size, reduction of cognitive workload, increase of safety and improved system performance. Some of these benefits are difficult to express in specific or quantitative metrics or units. Therefore, more qualitative approaches, for example using Subject Matter Experts (SME's), are often used for assessing the (expected) effectiveness of naval automation. Although this type of input can provide valuable insights, it can be ambiguous or subjective. The input may be dependent on which expert is consulted, at what time, or under what circumstances.

The limited availability of quantitative and specific effectiveness assessments makes it challenging to reach substantiated consensus on how (further) developments of automation should be approached, and how the required level of automation in the design of new ships and systems should be determined. Furthermore, it remains challenging to measure the effectiveness of automation after implementation. Potential risks of the lack of suitable effectiveness assessment methods are limited realisation of the expected advantages, overruns in cost or time, and insufficient product quality. As such, a need arises for (re-)evaluating the usefulness, scope, and applicability of automation effectiveness assessments.

Author's Biography

Agnieta Habben Jansen works at the Netherlands Defence Materiel Organisation as a naval architect. In a previous role, she has worked on the development of platform management systems. She holds an MSc and PhD degree in Naval Architecture from the Delft University of Technology.

Tom Stuivenberg is a naval lieutenant, currently working within the Netherlands Defence Materiel Organisation with a focus on submarines. He served onboard submarines in various engineering roles, including Marine Engineering Officer, and holds an MSc degree in Aerospace Engineering from the Delft University of Technology.

Maria Viegers works at the Netherlands Defence Materiel Organisation as a naval systems integrator, focusing on platform management systems. She holds an MSc degree in Mechanical Engineering from the Delft University of Technology, which a specialisation in the design of control systems.

This paper explores to which extent the need for assessing the effectiveness of naval ship automation is experienced, and if so, what type of information could be provided by such a method. This analysis can serve as input for potential further method development or tailoring. First, a literature review on existing methods and approaches for measuring the impact of automation programmes within naval ship design and associated industries is conducted (Chapter 2). Subsequently, an SME survey with experts in the field of automation within the Netherlands Ministry of Defence (MoD) is described in Chapter 3. This SME survey evaluates the existence and the experienced satisfaction level of automation effectiveness assessments currently in use within the MoD. Chapter 4 provides several takeaways for effective assessments of automation for (future) development projects. Overall conclusions are drawn in Chapter 5.

2. Literature review

In this Chapter, various existing methods for assessing the effectiveness of automation – both in the maritime as well as in other industries – are explored. For each domain, it has been investigated what the most common automation benefits are, what methods are used to determine the level of automation before the automation programme is conducted, and how the effectiveness of automation is assessed. A distinction is made between automation benefits that focus on measurable values, such as costs and crew size, and more qualitative metrics, such workload, safety and quality. Section 2.1 focusses on automation in the maritime domain. Section 2.2 addresses automation in the military domain and Section 2.3 focusses on industrial automation in general. The combined results and conclusions are provided in Section 2.4.

2.1. Maritime

Within the maritime industry, most of the benefits related to automation have a focus on measurable and specific parameters such as cost reduction, fuel savings and crew size reduction, according to the consulted literature. Table 1 lists an overview of the most-mentioned benefits in this domain, as well as in the other domains that have been investigated. Recently, the Netherlands Ministry of Infrastructure and Water Management published an investigation on the economic effects of smart shipping, concluding that investments related to automation programmes within the maritime industry will result in emission-decrease and related fuel-savings, a reduction of waiting times, crew size reduction, decrease of operational expenses and increase of overall safety (Ministerie van Infrastructuur en Waterstaat, 2022). To measure these benefits, all parameters were expressed in terms of costs saving, applying a use case approach.

Most of the consulted literature describe models that aim to determine the feasibility of automation programs or the required level of automation prior to the start of the programme, rather than assessing the effectiveness of the programme afterwards. Lyridis et al. (2005) assess the added value of applying an automation suite for the retrofit of an icebreaker using a cost analysis, in which all benefits – including more qualitative parameters such as safety – were expressed in terms of costs. Kooij & Hekkenberg (2021) have developed an algorithm to investigate the effect of automation on the crew size of a cargo vessel, using specific and measurable input parameters. To determine the level of automation prior to the start of the programme, they applied this algorithm to a case study. Based on the input of crew capabilities and tasks (sorted by cost), the algorithm provides a tasks list, a crew composition (number and capabilities) and the occupation per crew member, expressed in percentages. Miller et al. (2021) have investigated control systems for navigating and manoeuvring in several circumstances and environments. Their research focusses on quality and performance as most important benefits of automation. They have developed a physical model including measurable parameters (including speed, heading, and ship dimensions) to assess the impact of the automation prior to the execution of the automation programme. Research from Hannaford & Van Hassel (2021) applies a qualitative method to assess the effectiveness of the automation programme by determining the increased level of ship autonomy using a literature review, surveys and SME-interviews. To a limited extent, these qualitative results have been quantified using statistics. This results in an overview of expected benefits and risks of autonomy and/or crew size reduction, which can be used for pre-assessments and evaluation after programme completion.

Linking these methods to the specific field of naval ship automation, it can be stated not all methods and metrics can be applied directly, because the importance of cost – though highly significant – is usually less prevalent in the naval domain as a single or main driver. The more qualitative approach of Hannaford & Van Hassel (2021) may be adequate for capturing a broader scope of benefits and risks, but does not yet solve the initial challenges with respect to ambiguity and subjectiveness.

2.2. Military

Within the military domain, limited (open) research is available with respect to measuring the effectiveness of automation. However, literature describing the most important benefits of military automation is available in abundance. Compared to the maritime domain described in Section 2.1, costs savings seem to be of limited importance when it comes to military automation. In the military domain, the most common benefits are the increase of situational awareness (Salmon et al., 2004) , (Endsley & Kaber, 1999), (Hoffman et al., 2017), (Ophir-Arbelle et al., 2012), improvements in the decision-making process (Cummings, 2003), (Bolia et al., 2004), (Parasuraman et al., 2009), (Geertsma et al., 2016), reduction of cognitive workload and enlarged capabilities (for example: overcoming terrain difficulties), and performance increase (Sarter & Schroeder, 2001). It is notable to mention that some work focusses on the disadvantages and misbeliefs related to military automation and its benefits (Hoffman et al., 2017), (Calhoun et al., 2009), (Parasuraman & Bahri, 1992). They describe that crew size and workload reduction within this domain is seldomly realised. Within the naval maritime domain specifically, benefits seem to have a focus on more operational benefits such as operator workload reduction, availability improvement, increase of survivability and enlargement of situational awareness (Janssen et al., 2016), as well as enabling rapid decision making and preventing casualties (Geertsma et al., 2016).

Most of the 'military benefits' are difficult to measure. It is challenging to capture subjective indicators such as 'performance increase' and 'comfort improvement' into measurable metrics. In the literature of this domain, several approaches are used to make these performance indicators more measurable. Di Flumeri et al. (2019) measured brain activity in a neuroscientific experiment to establish the required level of automation to achieve maximum cognitive workload reduction. Most of the literature describe models or experiments that assess the effectiveness of automation using experiments where participants have to conduct specific tasks for several levels of automation with pre-determined success criteria. For example, Ophir-Arbelle et al. (2012) have conducted experiments where the quality of a visual assessment of the environment (target recognition) is measured for different levels of integration of video feed from unmanned vehicles on an existing soldier information application. Although the soldiers experienced an increase in cognitive workload, they recognised targets faster and better compared to the situation where there was no automated system to include feed from other sources. More common evaluation methods such as interviews and SME surveys and literature reviews are also commonly used in this domain.

It can be stated that the areas of interest of the military domain align with the context of naval ship automation. In addition to cost, the examples of research in the military domain have a strong focus on aspects that are stronger related to operational effectiveness and operator performance. At the same time, the challenge of measuring these effects remains. Several efforts have been made to express these effects in a more quantitative way, but these methods mostly focus on single metrics of which a baseline has been defined at the beginning of the research. In practice, the effectiveness of naval ship automation may be a combination of multiple factors, while baseline data may often not be available.

2.3. Industrial

Compared to the other domains, most research on the effectiveness of automation has been carried out within the industrial sector. According to Mathur et al. (2011), measures of performance for automatic manufacturing systems were dominated by cost measures and partial productivity measures till recent past, lacking strategic and external focus. Modern performance measurement includes several perspectives of business at the same time, such as for example: strategic value, continuous improvement, technological flexibility, customer satisfaction and innovation. The design of a performance measurement system that serves the needs of the company, with a limited number of metrics, limited data-capturing effort and minimal information overload, has therefore become more complex. Tangen (2002) makes some suggestions on the selection of a metric. For example that: "the measures must provide timely, relevant and accurate feedback, from both a long-term and a short-term perspective; measurement should be undertaken in ways that are easily understood by those whose performance is being evaluated; and measurement should be accomplished by a limited number of performance measures that consist of both financial and non-financial measures". Mathur et al. (2011) conclude that the Overall Equipment Effectiveness (OEE) is an important and universally accepted metric for measuring the overall performance of equipment in the industrial domain that can be further adapted for use as a metric for automatic manufacturing systems. The OEE was introduced by (Nakajima, 1988) as the total productive maintenance (TPM) and is calculated as the product of availability index (breakdowns, set-ups and adjustments), performance index (reduced speed, idling and minor stoppages) and quality index (defects, rework and yield).

Another model in this domain is introduced by Russell (2010), who describes a metric to give a corporation a comprehensive assessment of the success of an automation project. The metric, which is a variation on the function point analysis (FPA) used for estimating software effort (Albrecht, 1979), includes five functional ratings (cost, operation, technology, productivity and integration) based on responses to a survey. The value of the metric is modified by 14 adjustment factors like communications, web enablement and productivity. In order to assure a balanced viewpoint, employees from different layers across the company and with different interactions with the system (executives, users, and technical staff) are selected for the survey.

Furthermore, Labi (2022) and Rashid et al. (2021) discuss several methods for measuring the benefits or effectiveness of automation applications, specifically related to civil engineering disciplines. Labi (2022) identifies measures of effectiveness (MOEs) expressed in system performance indicators that can help the system owner in assessing the degree to which an application has been, or is expected to become, successful in achieving its intended objectives. The paper emphasises the importance of the choice of MOEs because it can be enormously influential in determining the direction that a project will take. Several examples of social, economic and environmental benefits are given in the form of reduction of some disutility or cost (monetary or non-monetary) or increase of some utility, and related to different system stakeholders.

It can be concluded that performance measurements in the industrial domain show methods to translate multiple quantitative aspects into a quantitative metric, which could be relevant for the naval domain. Like in the maritime domain, the factor that has always been important in the industrial domain is the cost measure. However, in modern industrial performance metrics the focus has shifted to including multiple different and qualitative factors. The difficulty and challenge with these metrics, that is mentioned in different papers and should be taken into account, is the complexity which requires a careful selection of the design method and the parameters.

2.4. Conclusions

To estimate or measure the effectiveness of automation, key performance indicators (KPI's), i.e. benefits of automation, are usually established. These KPI's, which may vary per application, describe to what objectives automation shall optimize. The literature review has discussed several KPI's of automation. Table 1 lists an overview of all the benefits of automation that are mentioned in the literature that has been consulted. The table shows that the most common KPI for automation processes is performance increase (both quality as productivity), a KPI that is dominant in all examined sectors. Reduction of operator (peak) workload is dominant in both the military and industrial sector, but is not an important automation benefit within the maritime domain according to the literature. An important difference between the industrial (and maritime) sector compared to the military sector is that cost reduction is not an important automation effect within the military industry. The increase of tactical capabilities and expansion of operational reach, fault reduction and speeding up the decision making process seem to be more valuable in the military domain. In the industrial domain modern performance measurement includes the most different perspectives, making it more extensive, but also making the design of the measurement system and selection of included metrics more complex. Different papers emphasise the importance of the performance measurement design in relation to the project or business needs.

Literature review shows that several models exist to measure automation success or effectiveness when KPI's of a more quantitative nature are used (costs, crew size, production time, availability, etc.). Most of the models and assessments that are reviewed are capable of analyzing the viability of the automation programme, the establishment of the required level of automation and the afterwards evaluation of the results. They usually do this using physical models or experiments, complex algorithms, value functions or cost evaluations. However, as most of the automation programmes have multiple objectives – not all of them being quantitatively measurable – it is less straightforward to find an appropriate model. In the maritime domain, the KPI's for automation effectiveness seem to be more quantitative than in the military and industrial domain. Apart from cost savings and production time reduction, other dominant metrics for measuring automation success in industry are workload reduction, performance and/or quality increase and improvement of reliability and usability, while in the military sector the reduction of faults, workload and the increase of tactical capabilities and improvements in the decision making process are most common in literature. Evaluation models to assess the impact of automation relative to these more qualitative metrics mostly focus on a specific metric – rather than on the complete set of KPI's. When evaluating the entire set of KPI's in an integrated approach, SME surveys and interviews as input for value evaluations are used.

All in all, the literature review indicates that the effectiveness of automation is often measured either specific but fragmented, or integrated but subjective or abstract. Measuring the effectiveness of automation in a way that is both integrated and specific seems to remain a major challenge.

Table 1: Effectiveness criteteria described in literature

Category	**Effectiveness criteria**	**References**		
		Maritime	**Military**	**Industrial**
Performance	Increase product performance (productivity or quality)	[Mar1] [Mar2]	[Mil1] [Mil2] [Mil3] [Mil4] [Mil5] [Mil6] [Mil8] [Mil9]	[Ind1] [Ind2] [Ind3] [Ind4] [Ind7] [Ind8] [Ind9] [Ind11] [Ind12]
	Improve accuracy		[Mil4]	[Ind1] [Ind2]
	Increase system reliability		[Mil1]	[Ind1] [Ind4] [Ind7] [Ind8]
	Increase availability			[Ind1] [Ind2]
	Expand operational reach (weather, terrain, etc.)		[Mil1]	
	Reduce negative side effects (collateral damage)		[Mil13] [Mil14]	
	Increasing tactical capabilities, including situational awareness improvement		[Mil1] [Mil6] [Mil7] [Mil10]	[Ind8] [Ind10]
	Maximise remaining capacity after availability loss		[Mil14]	
	Improve data collection / data storage			[Ind2]
	Improve communication of performance results			[Ind2]
	Improve / increase use of performance measures			[Ind2]
Operator-related	Reducing operator workload / Increasing operator productivity		[Mil1] [Mil2] [Mil4] [Mil5] [Mil6] [Mil13]	[Ind3] [Ind4] [Ind8] [Ind10]
	Reducing number of operator faults	[Mar2]	[Mil2] [Mil3] [Mil6] [Mil13] [Mil14]	[Ind8]
	Increasing system owner productivity			[Ind3]
	Reduce time for decision-making		[Mil3] [Mil4] [Mil12] [Mil14]	
	Crew size reduction	[Mar3] [Mar4] [Mar5]	[Mil7]	[Ind4]
	Reducing training requirements		[Mil7]	
Economical	Reducing (production) time			[Ind2] [Ind4] [Ind5]
	Cost savings	[Mar1] [Mar5]	[Mil1]	[Ind1] [Ind3] [Ind4] [Ind5] [Ind11]
Quality	Improve customer satisfaction			[Ind2] [Ind4]
	Improve usability			[Ind1] [Ind4] [Ind7]
	Improve comfort level		[Mil1]	[Ind3] [Ind10]
	Increase technological flexibility			[Ind2] [Ind4] [Ind8]
	Improve maintainability			[Ind7]
	Improve innovation / learning			[Ind2]
	Improve long term / strategic value			[Ind2]
Social	Increase of safety	[Mar5]	[Mil1]	[Ind3] [Ind4]
	Increase passenger privacy			[Ind3]
	Improve employee satisfaction			[Ind4] [Ind8]
	Corruption mitigation			[Ind3]
	Increase job opportunities			[Ind3]
Environmental	Reducing ecological footprint / energy use	[Mar5]		[Ind2] [Ind3] [Ind4] [Ind6]

Definitions of references:
[Mar1] (Lyridis et al., 2005); [Mar2] (Hannaford & Van Hassel, 2021); [Mar3] (Miller et al., 2021); [Mar4] (Kooij & Hekkenberg, 2021); [Mar5] ((Ministerie van Infrastructuur en Waterstaat, 2022) [Mil1] (Boehm-Davis et al., 1983); [Mil2] (Vidulich & Bortolussi, 1991); [Mil3] (Cummings, 2003); [Mil4] (Parasuraman et al., 2009); [Mil5] (Taylor, 2015); [Mil6] (Ophir-Arbelle et al., 2012); [Mil7] (Hoffman et al., 2017); [Mil8] (Di Flumeri et al., 2019); [Mil9] (Calhoun et al., 2009); [Mil10] (Salmon et al., 2004); [Mil11] (Endsley & Kaber, 1999), [Mil12] (Bolia et al., 2004), [Mil13] (Janssen et al., 2016); [Mil14] (Geertsma et al., 2016); [Ind1] (Russell, 2010); [Ind2] (Mathur et al., 2011); [Ind3] (Labi, 2022); [Ind4] (Rashid et al., 2021); [Ind5] (Homem de Almeida Correia et al., 2019); [Ind6] (Zheng et al., 2018); [Ind7] (Karnouskos et al., 2018); [Ind8] (Parasuraman, 2000); [Ind9] (Boutraa et al., 2011); [Ind10] (Moray & Sheridan, 2004), [Ind11] (Parasurman & Bahri, 1992); [Ind12] (Parasuraman & Riley, 1997)

3. Subject Matter Experts survey

The literature review resulted in a list of most common KPI's per sector and the conclusion that limited methods are available for measuring the overall impact of automation that can be applied in the naval ship domain. To further assess the need and applicability of such methods, an SME survey has been carried out under 17 SME's, both military (5) and civilian (12), within in the field of naval ship automation within the Netherlands Ministry of Defence (MoD). The survey, that was distributed by email, included eight open questions and responding to the survey was voluntary. The SME's work in various areas of automation, covering platform, combat, and navigation systems. The authors believe that the combined responses give an informative impression of the way in which the effect of automation is assessed at the MoD, despite the small sample size. The results of the SME survey are presented in Sections 3.1 to 3.3.

3.1. Effectiveness of automation

First, the SME's were asked if the automation they are working on is (or has the potential to become) a success. 13 of the SME's agreed (almost) fully, or with minor reservations. 2 SME's agreed with more significant reservations, and 2 SME's did not agree, or only to a limited extent. Table 2 lists the reasons mentioned by the SME's for agreeing or disagreeing to the question whether the automation project they currently are working on is successful.

Table 2: SME responses on the succesfulness of automation projects

Reasons for agreeing that automation is successful	**Reasons for disagreeing, or having reservations on the success on automation**
• Data/information needed for on-board operations is available at a central position (e.g. navigation bridge or engine control room). • Data/information required for on-board operations is provided earlier to the crew, and/or in a better way. • The situational awareness of the crew has been increased. • Crew size reduction has been realised. • Operations can be executed with reduced operator load or fewer tasks. • On-board data/information is becoming increasingly available on shore for smart maintenance purposes. • Developments in automation are executed from the perspective of the user, rather than from the perspective of systems or technologies.	• Automation may result in an increase in number and complexity of system warnings, alarms or failures. • System warnings and alarms may be false, unclear, or unexpected. • The quality of the sensor data for automation is insufficient or unknown. • Crew members may have insufficient experience or training for working with the automation systems in an adequate way. • The time between development and deployment of new automation systems or applications is (too) long.

3.2. Key performance indicators and associated metrics

The second part of the survey elaborates on which KPI's are used by the SME's to determine the effectiveness of automation, and which metrics (if applicable) are used to express this. 6 SME's mention KPI's that are predominantly specific and measurable and are also dominant in Table 1 (see Section 2.4). These are the following:

- Time spent by the crew on interpreting and repairing failures
- Time saved in on-board or on-shore decision making
- Crew size reduction
- Reduction of crew-induced errors
- Availability and performance of (weapon) systems, based on the achieved sensor precision
- Costs (for production, maintenance, and modifications)

It must be noted that even though these KPI's could in essence be measured easily, the associated context may not always be specific. For example, time saved for decision making could be measured using a stopwatch, but the context in which these decisions are made, may vary with time, operational context or crew composition. As such, the specificity of KPI's has the nature of a continuous scale, rather than a binary concept (either specific or abstract). Nevertheless, the KPI's mentioned above have a different nature than the more abstract KPI's that are (also) mentioned by most SME's. These abstract KPI's are mentioned by 13 SME's and can be summarised into one category: user experience. Almost all of these SME's mention the abstract and/or subjective nature of user feedback as a factor that complicates the effectiveness assessment. 4 SME's mention explicitly that they do not (yet) use KPI's or associated metrics because they don't know how to define them. 3 SME's state that though some aspects on user feedback can be measured or made explicit with tests or experiments, this is usually not done in practice for the following reasons:

- The absence of baseline data on the initial situation makes it impossible to determine if the automation has the desired effect.
- Performing this type of assessments requires substantial time and effort.
- The uncertainty whether specific quantitative measurements have actual added value compared to the qualitative user feedback.

Several SME's mention methods to apply user feedback, despite it being subjective, for the effectiveness assessment of automation. 3 SME's mention that they assess the willingness of the crew to use the automation. Furthermore, 1 SME mentions using the feedback of the crew trainers to determine if the automation is effective, and 1 SME mentions the number of formal notifications in the ERP (Enterprise Resource Planning) system as an indication for assessing the effectiveness of automation.

2 SME's express reservations in using KPI's and associated metrics for assessing the effectiveness of automation. The reasons for this are the following:

- A more holistic approach is needed. The focus should lie on proof of value and benefit realisation. KPI's for automation are not (directly) suitable for expressing such higher level, overarching concepts.
- Using KPI's may induce the risk that they are used for judging crew members on their performance, rather than for providing transparency.

Linking these results to the literature review, it can be stated that the SME's experience the same challenge as the literature contributions: some metrics can be quantified relatively easily, but they reflect only a single part of automation efforts. User feedback can provide a broader and more integrated overview, but is more difficult to quantify, and may be subjective.

3.3. *Organising the development of automation*

The third part of the survey considers the way in which the SME's structure and organise the development of automation. First, they were asked which stakeholders are involved in determining the effectiveness of automation. A broad scope of representatives is mentioned, that aligns with the backgrounds of the SME's (requirement definers, (software) designers/developers, operational users, maintainers and in-service support experts). Most SME's mention that they determine the effectiveness of automation within a group in which several of these roles are combined, in order to get a complete overview. In some cases, the SME's seem to work on this topic predominantly with experts that are in the same project phase as themselves. For example, some SME's that are involved in defining requirements, which is an early project phase, mention product designers and scientists from research institutes as the group with which they assess the (estimated) effectiveness of automation. Similarly, some SME's that work on maintenance emphasise that operational users and system life cycle managers are consulted for this topic. 2 SME's mention that they do not assess the effectiveness of automation in a structured way. These are SME's that mentioned that they do not (yet) have metrics for this (see Section 3.2). 1 SME mentions not only persons ('actors') as being involved in assessing the effectiveness of automation, but also mentions doctrine, information and business operations.

The last question of the survey considers when the effectiveness of automation is assessed by the SME's: before a new development starts, when a solution or product is being developed, and/or when it is used on board (or on shore, if applicable). All SME's mention that the effectiveness of the automation is assessed during operational use. 9 SME's mention that only then the real impact of the automation can be assessed in a proper way. In addition to that, 13 SME's also assess the (estimated) effectiveness of the automation while it is still in development. This requires adequate methods, since the product is not yet finished at this stage. The SME's mention the following methods for dealing with this uncertainty:

- Defining a Minimum Viable Product (MVP), thus ensuring that the transition between development and deployment of the product can be executed in a smooth way, while delays or premature deployment are being avoided.
- Product assessments 'on paper', in which user stories are applied and to evaluate if the product meets the requirements.

- Product tests on a shore reference system, a training facility, or a factory test site.
- Initial product release to a group of key users, who act as accelerators for adopting the product in the organisation.

The start of a project is acknowledged as the most difficult project phase to assess, or estimate, the effect of automation. 6 SME's do not mention the initial product development phase for estimating the effectiveness at all. This may relate to the fact that these SME's work in later stages of the life cycle of a product, such as operational use or in-service support. The other 11 SME's mention that they perform upfront estimations, or that they believe they should be done. They mention the following challenges on performing these estimations:

- Acknowledging the need for effectiveness estimations, but omitting them due to the lack of suitable methods.
- Having to trust on the (not necessarily correct) input of only a few future operational users.
- Incompatibilities in the project timeline, such as having to procure physical equipment already while not all requirements have been defined.

4. Discussion

The combined results of the literature review and the SME survey provide insight in how the effectiveness of automation can be assessed. Especially, the SME survey results elaborate on this topic from a practical viewpoint, and are tailored to the case of naval ship automation. Since this viewpoint is considered as a key perspective of this paper, the discussion of this section will use the same viewpoint, while ensuring that takeaways and lessons learned from other related areas of the literature study remain kept in mind. It must be noted that the discussion is held from the viewpoint of the authors' experiences. These are subjective as well, and do not have formalised (literature) references. Nevertheless, keeping the practice oriented perspective in mind, the authors' experiences are regarded as a meaningful contribution to the discussion.

Several points of attention arise when the literature review, SME survey and authors' experiences are combined. First, it becomes clear that there are significant differences in the type of tasks that are automated. On the one hand there is automation of tasks and processes that are mostly based on physics, such as the developments for automated manoeuvring (Miller et al., 2021) (see Section 2.1). On the other hand, there is automation of tasks that are more human oriented, such as most examples in the military domain (see Section 2.2). These are not necessarily two separate categories, since automation efforts can consider both types of tasks at the same time. Yet, the physics based type of automation – though requiring significant in-depth R&D efforts – appears to be less complex in the sense of assessing its effectiveness. This is mainly because specific and measurable parameters are often available. As such, discussions on how to measure the effectiveness of automation are regarded more meaningful for human oriented automation. This aligns with most of the comments and concerns that were raised by the SME's.

Regarding the human oriented automation, the lack of specific and measurable KPI's is indeed recognised by many of the SME's. This aligns with the authors' motivation of initiating the research for this paper (see Chapter 1). Yet, defining such KPI's may not necessarily provide the added benefit that was originally assumed. Although the literature review provides examples of how unspecific KPI's such as 'increased situational awareness' can be measured, the SME survey shows that making such assessments is not necessarily achievable, realistic or useful. There seem to be significant practical constraints to carry out such assessments, mainly because they are time intensive and need baseline data, which is often not available. Focussing too strongly on assessing the effectiveness of automation with specific and measurable KPI's poses the risk that more effort is spent on the method than on the actual desired effect, namely that operators are provided with meaningful support in executing their tasks. Yet, the SME survey shows that the current approach, where user feedback remains subjective or unspecific, does also not yield the desired effect. Combining these arguments, the authors believe that not everything needs to be measured and that user feedback, even with its limitations, can serve as an adequate metric for assessing the effectiveness of automation. However, the authors also believe that essential steps and paradigm shifts are required to apply user feedback more effectively in the development of automation systems. The following statements, which are starting points rather than direct solutions, are proposed:

1. A method or process needs to be incorporated to 'smoothen' user feedback. At present, proposals for modifications or new features often emerge ad hoc, based on specific experiences from specific users with specific systems. There needs to be a mechanism for obtaining user feedback in a more continuous way, which enables gathering multiple perspectives over time, and priority setting of user inputs. Though in principle such procedures and mechanisms are available, the authors believe that they are currently not used to the full potential.

2. Once an adequate user feedback mechanism is available (or given renewed attention), the persons involved in applying the user feedback for automation developments need to stay involved and committed to this mechanism. This may require commitment to collaborative tasks that go beyond the initial responsibility of individual persons. For example: if an operational user is consulted for developing a new automation system, he or she needs to be given the time and opportunity to be involved in this collaboration, despite the fact that he or she may be an operational crew member rather than a development engineer.
3. Applying user feedback in automation developments needs to be aligned with the development stage of the product. A completely new product requires different user input than a product that has been there for years and is under continuous evolutionary development. In addition to the development stage, other aspects, such as the application field and scope of the project, also affect the type of user feedback and the extensiveness of the evaluation that is required. As such, the process should refrain from a one-size-fits-all system for gathering and applying user feedback. At the same time, separate processes for individual systems need to be avoided, because the overall effectiveness of automation is often the result of an overarching automation philosophy, rather than the result of applying separate systems. This is a delicate balance that needs more attention in the opinion of the authors.

5. Conclusion

Expressing the effectiveness of naval automation is a complex effort. There is general consensus in literature and among SME's that some form of effectiveness assessment is necessary during all stages of system development, but it often remains unclear how this should be done, or what type of metrics should be used. Especially developments in automation of human tasks (more than physics based tasks) often lack specific metrics to assess their effectiveness. It has been concluded that investing effort in defining such specific metrics does not necessarily provide the desired insight, or is sometimes simply not practical. Instead, user feedback remains to have the potential of being an adequate indicator for the (estimated) effectiveness of naval automation, even when it is subjective, unspecific or not measurable. However, in order to better benefit from user feedback in the development of (new) naval automation systems, the process for gathering and applying user feedback needs continuous attention. In practice, this means that there should be an adequate mechanism for dealing with user feedback, as well as opportunity and commitment of all parties involved in the collaboration, in order to make naval automation a success.

Disclaimer

The opinions presented in this paper are the personal opinions of the authors and the authors alone. Specifically, they do not represent any official policy of the Netherlands Ministry of Defence, the Defence Materiel Organisation, or the Royal Netherlands Navy.

References

Boehm-Davis, D., Curry, R., Wiener, E. & Harrison, L., 1983. Human Factors of Flight-Deck Automation: Report on a NASA Industry Workshop. *Ergonomics, Volume 26, Issue 10,* pp. 953-961.

Bolia, R., Todd Nelson, W., Vidulich, M. & Taylor, R., 2004. *From Chess to Changellorsville: Measuring Decision Quality in Military Commanders.* Daytona Beach, Human Pewrformance, Situational Awareness and Automation Technology Conference.

Boutraa, T., Akhkha, A, Alshuaibi, A & Atta, R, 2011. Evaluation of the Effectiveness of an Automated Irrigation System Using Wheat Crops. *Agriculture and Biology Journal of North America,* pp. 2151-7517.

Calhoun, G., Draper, M. & Ruff, H., 2009. Effect of Level of Automation on Unmanned Aerial Vehicle Routine Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting,* 1 October, pp. 197-201.

Cummings, M., 2003. *Designing Decision Support Systems for Revolutionary Command and Control Domains,* Charlottesville: Universiy of Virginia.

Di Flumeri, G. et al., 2019. Brain-Computer Interface-Based Adaptive Automation to Prevent Out-of-the-Loop Phenomenon in Air Traffic Controlles Dealing with Highly Automated Systems. *Frontiers in Human Neuroscience, Volume 13.*

Endsley, M. & Kaber, D., 1999. Level of Automation Effects on Performance, Situational Awareness and Workload in a Dynamic Control Task. *Ergonomics,* 42(3), pp. 462-492.

Geertsma, F., Badon Ghijben, N., Middeldorp, E. & Wouters, R., 2016. *Adaptive Automation to Maximise Warship Capabilities During Incidents.* Bristol, INEC 2016.

Hannaford, E. & Van Hassel, E., 2021. Risks and Benefits of Crew Reduction and/or Removal with Increased Automation on the Ship Operator: A Licensed Deck Officer's Perspective. *Applied Sciences,* pp. 11(8), 3569.

Hoffman, R., Sarter, N., Johnson, M. & Hawley, J., 2017. Myths of Automation and Their Implications for Military Procurement. *Bulletin of the Atomic Scientist, Volume 74, Issue 4.*

Homem de Almeida Correia, G. et al., 2019. On the Impact of Vehicle Automation on the Value of Travel Time While Performing Work and Leisure Activities in a Car: Theoretical Insights and Results from a Sated Preference Survey. *Transportation Rsearch Part A: Policy and Practice*, pp. 359-382.

Janssen, J. et al., 2016. *Autonomous, adaptive, aware: DINCS.* Bristol, INEC 2016.

Karnouskos, S., Leitao, P, Ribereiro, L.D & Sinha, R, 2018. *Assessing the Integration of Softwrae Agents and Industrial Automation Systems with ISO/IEC 25010.* Porto, 16th International Conference on Industrial Informatics

Kooij, C. & Hekkenberg, R., 2021. The effect of autonomous systems on the crew size of ships–a case study. *Maritime Policy and Management,* pp. Vol. 48, No. 6, pag. 860-876.

Labi, S., 2022. Measuring the benefits of civil systems connectivity and automation – a discussion in the context of highway transport. *Civil Engineering and Environmental Systems,* pp. 27-47.

Lopez, C. T., 2022. *Defense.gov.* [Online] Available at: www.defense.gov/News/News Stories/Article/Article/2904627/[Accessed 1 June 2022].

Lyridis, D., Psaraftis, H., Ventikos, N. & Zacharioudakis, P., 2005. Cost-Benefit Analysis for Ship Automation Retrofit: the Case of Icebreaker Frej. *Marine Technology,* pp. 113-124.

Mathur, A., Gangayach, G., Mittal, M. & Sharma, M., 2011. Performance measurement in automated manufacturing. *Measuring Business Excellence,* pp. 77-91.

Miller, A., Rybczak, M. & Rak, A., 2021. Towards the Autonomy: Control Systems for the Ship in Confined and Open Waters. *Sensors.*

Ministerie van Infrastructuur en Milieu, 2015. *De Nederlandse Maritieme Strategie 2015-2025,* Den Haag: s.n.

Ministerie van Infrastructuur en Waterstaat, 2022. *Economische effecten smart shipping,* Den Haag: Panteia & Ecorys.

Moray, N. & Sheridan, T., 2004. *Ou sont les neiges d'antan.* Daytona Beach, s.n., pp. 1-31.

Nakajima, S., 1988. *Introduction to TPM: Total Productive Maintenance.* Cambrdige, Mass.: Productivity Press.

OECD, Organisatie voor Economische Samenwerking en Ontwikkeling., 2020. *Peer review of the Dutch Shipbuilding Industry,* Paris

Ophir-Arbelle, R., Oron-Gilad, T. & Borowsky, A., 2012. Is More Information Better? How Dismounted Soldiers Use Video Feed From Unmanned Vehicles: Attention Allocation and Information Extraction Considerations. *Journal of Cognetive Engineering and Decision Making, Volume 7, Issue 1.*

Parasuraman, R., 2000. Designing Automation for Human Use: Empirical Studies and Quantitative Models. *Ergonomics,* pp. 931-951.

Parasuraman, R. & Bahri, T., 1992. *Theory and Design of Adaptive Automation in Aviation Systems,* Huachuca, Arizona: US naval Air Warfare Center.

Parasuraman, R., Cosenzo, K. & De Visser, E., 2009. Adaptive Automation for Human Supervision of Multiple Uninhabited Vehicles: Effects on Change Detection, Situation Awareness and Mental Workload. *Military Psychology,* 21(2).

Parasuraman, R. & Riley, V., 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors,* pp. 230-253.

Rashid, M. N. A., Abdullah, M. R., Ismail, D. & Saberi, M. H., 2021. Success Criteria for Automation and Robotics in Industrialised Building System (IBS). *International Journal of Academic Research in Business & social Science, Volume 11, Issue 2.*

Russell, D., 2010. A Metric for Rating the Effectiveness of Industrial Automation Systems Using a Derivative of Function Point Analysis. *Robotics and Computer-Integrated Manufacturing,* pp. 551-557.

Salmon, P., Stanton, N., Walker, G. & Green, D., 2004. *Situational Awareness in Military Command and Control (C4i) Systems: the Development of a Tool to Measure SA in Military Command and Control Systems and Battlefield Environments.* Daytona Beach, Human Performance, Situational Awareness and Automation Conference (HPSAA).

Sarter, N. & Schroeder, B., 2001. Supporting Decision Making and Actions Selection Under Time Pressure and Uncertainty: the Case of In-flight Icing. *Human Factors, Volume 43,* pp. 573-583.

Taylor, G., 2015. *Comparing Types of Adaptive Automation Within a Multi-Tasking Environment,* Orlando: University of Central Florida.

Vidulich, M. & Bortolussi, M., 1991. *The Effects of Speech Controls on Performance in Advanced Helicopters in a Double Stimulation Paradigm.* NASA Ames Research Centre, Behavioral Sciences

Zheng, J., Huang, B. & Zhou, X., 2018. A Low Carbon Process Design Method of Sand Casting Based on Process Design Parameters. *Journal of Cleaner Production,* pp. 1408-1422.