Ready for Takeoff?

The Application of Human Readiness Levels (HRL)

By Laura L. Moens

Thesis submitted to the Faculty of Behavioural Management and Social Sciences in partial fulfilment of the requirements for the degree of Master of Science in Human Factors and Engineering Psychology

Supervised by

Prof. Dr J. M. C. Schraagen (University of Twente)

External supervisor: Dr S. Bonelli (Deep Blue)

Second supervisor: Dr S. Borsci (University of Twente)

August 2021

Abstract

In order to mitigate HF-related risk and prevent costly late-stage design changes, it is crucial to include Human Factors (HF) throughout the development and validation process of new socio-technical (sub)systems. The Technology Readiness Levels (TRL) scale proved to be helpful as a systematic method for measuring the maturity of technology but incapable of capturing the human-related aspects of technology development. Previous research developed an initial Human Readiness Levels (HRL) framework to complement the existing TRL. Although first introduced in 2010, the HRL still has not been adopted by system development experts and project managers. The present work compared the most recent HRL framework to five systems development projects, four established validation frameworks and the experience of a seasoned project manager. Additionally, it compared the HRL to the Cognitive Systems Engineering (CSE) approach. This comparative analysis aimed to identify the gap and overlap between the HRL and currently-accepted practices. Consequently, 23 points of improvement for the HRL were identified and subsequently used to redesign the HRL framework. The adapted framework was then subjected to a user test. In total, five HF experts from different industries used the HRL in a think-aloud interview. The resulting conclusions with regards to the HRL's utility and usability are discussed. The outputs of the present work can be used to create an HRL that is useful and user-friendly and thereby increase its chance of receiving widespread adoption, like the TRL.

Keywords: Aviation, Human Factors (HF), Human Performance (HP), Human Readiness Levels (HRL), Human Systems Integration (HSI), Research & Development (R&D), Safety, Socio-Technical Systems, Systems Development, Technological Innovation, Technology Readiness Levels (TRL)

Contents

Ready for Takeoff? The Application of Human Readiness Levels

The present work argues for the importance of integrating the Human Factors (HF) perspective and methods into the development and validation process of new socio-technical systems to decrease the HF-related risk. The case of the Boeing 737 MAX provides a practical and illustrative example of the consequences of inadequate inclusion of HF.

In 2011, Boeing Commercial Airplanes announced the launch of a new aircraft: the 737 MAX (Titan Grey, n.d.). About six years later, the plane was deemed ready for commercial service. Among the first commercial airlines to fly with the 737 MAX was Lion Air.

> At 6.32 am local time on October 29 [2018], LionAir 610 crashes at extraordinary speed into the Java Sea at only 13 minutes after takeoff, killing all 180 men, women and children on board. (60 Minutes Australia, 2019)

Just after taking off, the stick shaker had been activated, multiple error messages had been presented on the display, and the aircraft had started to relentlessly trim its nose down (Campbell, 2019; Davies, 2019). What followed was a fight between Man versus Machine: Under the cascade of problems and warnings, every five seconds of the pilots' effort to lift the nose of the aircraft back up was alternated with ten seconds in which the aircraft would dive again towards the ground (60 Minutes Australia, 2019). After a 'rollercoaster' ride of about ten minutes, the machine had won.

Preliminary research pointed at a malfunctioning flight-control system that the pilots should have disabled for the aircraft to function normally again. In response to the accident, Boeing released guiding material to supplement their operational manual in which they advised pilots regarding procedures for dealing with "erroneous cockpit readings", as they called it (Titan Grey, n.d.).

On 10 March 2019, the Boeing 737 MAX aircraft operated by Ethiopian Airlines crashed six minutes after the takeoff, killing all the 157 people aboard (Titan Grey, n.d.) According to news sources, the accident was similar to the one of Lion Air. The investigations pointed to the Maneuvering Characteristics Augmentation System (MCAS) as the presumable source of the aeroplane's repeated dives when receiving erroneous systems data. Boeing designed MCAS to compensate for the changed aerodynamics of the aircraft caused by the more fuel-efficient but heavier motors (Gates & Baker, 2019). This system was mentioned neither in the 737 MAX pilot manual nor in its supplement introduced after the Lion Air crash (Titan Grey, n.d.). Since Boeing did not communicate it to the pilots and provide specific training, this MCAS was not part of the pilots' mental model (Mauro, n.d.).

This second crash led aviation regulators across the world to issue grounding orders for the 737 MAX. On April 4th, Boeing publicly announced the acknowledgement that MCAS had played a role in both crashes (Titan Grey, n.d.). Next to the severe loss of so many people's lives, Boeing experienced catastrophic financial and reputation loss. By August, its market capitalisation had fallen by $62 billion, accompanied by a 25% plunge in share price, and the halt of sales and impending cancellation of orders cost Boeing roughly $600 billion (Titan Grey, n.d.).

In December 2019, Dr Mica Endsley, Government Relations Committee Chair of the Human Factors and Ergonomics Society, testified before the U.S. House of Representatives in a hearing that examined the Federal Aviation Administration (FAA) 's certification process of the 737 MAX (HFES, 2020; Human Factors & Aviation Safety, 2019). In her testimony, Dr Endsley underscored that HF standards were not respected in the design and certification of the 737 MAX, with inherent brittleness of automation, lack of automation transparency and inadequate automated systems training as a result. Accidents like that of Lion Air Flight 610 and Ethiopian Airlines Flight 302, and their costly repercussions, could potentially have been prevented by allowing HF a more prominent role throughout the development of aircraft systems.

Often human elements are considered only after a system failure occurred, and negative consequences took place (Newton et al., 2017), such as in the case of the Boeing 737 MAX (Anderson, n.d.; Dekker, 2009; Nunes & Laursen, 2004; Titan Grey, n.d.). Taking a proactive approach instead is still a challenge as some have pointed out that the professional field of HF science struggles with unfamiliarity (or even ignorance perhaps), despite possessing the methodologies capable of contributing to the prevention of accidents (Mauro, n.d.). Smit (2020) studied the Dutch branch of Thales and found support for the need to increase HF awareness. She interviewed experts from across the company and concluded that knowledge of HF was minimal and dispersed. One interviewed Test & Integration manager said: "Often, we are not aware where HF begins and where it ends". A System Engineering manager further emphasised this when saying: "There is a need for some sort of HF instruction in order to raise awareness about the domain. For which aspects do we need to involve the HFE?" (Smit, 2020, p. 27).

The present work focuses on a framework that can ensure the structural consideration of the human in the system throughout the systems development life cycle: the Human Readiness Levels (HRL). First of all, several human readiness concepts will be discussed in Section 1.2. The thesis will then focus on the most recent HRL concept, namely the one developed by the HRL working group (see Appendix A). Sections 2.1 and 2.2 will describe the analysis of the HRL through careful comparison with current practices and relevant literature, respectively. In Chapter 3, a list of potential points of improvement for the HRL will be presented as well as an adjusted version of the HRL framework. Subsequently, this adjusted HRL will undergo a user test. More specifically, HF experts will be asked to use the HRL framework to assign a maturity level to a project in which they were, or are, involved.

The method and results of this user test will be described in Chapter 4. Finally, open issues and further recommendations will be presented in Chapter 5. This thesis aims to stimulate the successful application of the HRL framework within systems development. Therefore, it is of interest to systems developers, HF or operations specialists and validation practitioners who have to design and validate complex, safety-critical, socio-technical systems as well as the managers of these large innovation projects or programmes.

1.1 Advantages of HRL application

By assessing the maturity in terms of human-related aspects of technologies while they evolve from concepts into stand-alone prototypes and eventually into components or subsystems integrated with the operational system, the HRL ensures that the integration of the human and the system is considered throughout a system's development cycle. In this way, it can not only identify human-related issues but also track and resolve them, thereby decreasing operational risks associated with inadequate Human Systems Integration (HSI), similarly to how the TRL is used for managing technology risk (Phillips, 2010). As stated by Phillips, the lowest level of readiness is equal to the highest level of risk.

Apart from identifying and mitigating HF risk, the HRL can also improve project planning and decision-making. Like the TLR, the HRL is meant to smoothen the research and development (R&D) process and make more informed project decisions (Hale et al., 2011; Krois & Rehmann, 2005). For example, HRLs could be used to compare the maturity of different technologies. A project manager can utilise this information to adjust investments in order to stimulate the development of lagging technologies (Miller et al., 2016). Furthermore, the application of HRL helps to set program milestone criteria (Phillips, 2010). The HRL exists of levels that can be seen as milestones or steps in the process of new systems development and validation from concept to deployed system. Each level offers a description of the actions required to advance the system towards socio-technical maturity. Structured milestones and corresponding criteria could also support accountability and communication. Clear and reliable communication is essential in a multidisciplinary field like system design, where expertise in psychology, engineering, design, safety, and even law meets and where governmental, non-governmental, and commercial stakeholders collaborate (Scaravetti & Montagnier, 2009). The HRL could, just like the TRL, provide a common reference point for all parties involved and support the elimination of misunderstandings and ambiguities in the technology development process (Nolte, 2008). The HRL could also be an excellent overview and potential training material for novice HF specialists.

Finally, the application of the HRL can help to save costs. Identifying and changing a sub-optimal human-system interface during the developmental phase, rather than post-release, saves as much as ten times the costs (Johnston et al., 2013). Newton and colleagues (2017) demonstrated the economic impact of HF issues and the need for an HRL scale through the use of a retrospective case study. They let a team of HF experts look at a case study concerning the development of a weapons system within a national laboratory and concluded that 67% of the costs and 36% of the manufacturing time could have been saved by considering HF proactively. Many examples illustrate how costs significantly increase, up to 1500 times, when (HF) design flaws are detected after completion of the development process rather than early on in the development (Del Giudice et al., 2015; Haskins et al., 2004). Moreover, Phillips (2010) mentioned budget overruns and schedule delays as some of the main challenges that result from the difficulty of managing technology maturity. Apart from preventing expensive late-stage design changes, HRL could also support strategic financial planning and decision-making similar to the TRL. The TRL made it possible to estimate the relative cost of each development phase (Héder, 2017). It is now known that the development costs generally grow exponentially, peaking around TRL8 and that most Boeing projects reach TRL 6 using less than 10% of the total budgeted amount. The application of the HRL could help organisations collect valuable data that will enable them to better estimate project costs in the future.

One might wonder whether the benefits mentioned above are specific to the HRL or more general advantages of HF. The uniqueness of the HRL lies in the emphasis on continuity and scope. HF is often seen as the analysis and optimisation of an interaction between a human and a machine at a specific point (in time and space). For example, established policies often see HSI as a one-time act instead of a continuous process (Phillips, 2010). Moreover, HSI evaluation tends to take place too late into the process (Jander et al., 2012). Instead, the timeline and breadth of HF analysis are much more extensive, spanning across the complete system life cycle and including the organisation around the system. When HRL is applied as it is supposed to be, it ensures the inclusion of HF throughout the development process and across the system's scope. Thus, the HRL is a practical application of HF as a scientific discipline in the form of a process tool to develop systems. Moreover, because HRL is a science-based and consistent method, it supports a more quantifiable and less subjective judgment of a system's maturity (GAO, 1999). Because there is a lack of knowledge concerning human-centred measurements and requirements in systems development, HF is often overshadowed by more readily defined technology requirements (O'Neil et al., 2015). An encompassing but simplified framework like HRL can help fill this gap and institutionalise and 'advertise' HF within systems development.

1.2 Overview HRL

This section will briefly present the evolution of the HRL from its origin by discussing the various frameworks developed over the years. Furthermore, the open issues concerning the HRL frameworks will be highlighted, and the research objectives will be presented.

1.2.1 Origin of the HRL

The idea of HRL is based upon the more well-known Technology Readiness Levels (TRL) (Phillips, 2010). The TRL is a framework that was initially developed by NASA and later embraced and adjusted by the Department of Defense (DoD) and the European Union (Graettinger et al., 2002; Héder, 2017) to indicate the maturity of a technology. Phillips (2010) offers a good summary of the TRL concept. In a nutshell: "by the time the technology has reached a TRL 9, the technology has progressed through the formulation of an initial concept for application, proof of concept, demonstration in a laboratory environment and realistic environment, and integration into a system" (Phillips, 2010, p. 12). Essentially, TRLs describe the general Research and Development process in more detail, starting with basic research and concluding with demonstration. The TRL scale applied by the European Commission can be found in Appendix B (European Commission, 2014).

Many alternative and complementary readiness scales have spun off from the original TRL. Examples are the Manufacturing Readiness (Del Giudice et al., 2015), the Integration Readiness, and the System Readiness (Miller et al., 2016) level scales. The HRL is one of the newest derivatives and goes by different names. All of these concepts are meant to complement the existing TRL with the idea that the capability of technology is not equal to its performance but rather depends on a set of interacting components. The TRL proved to help assure that the technical components of the system will work as intended, but it does not secure operational success (See et al., 2017). Therefore, Phillips (2010) decided to publish a human-focused counterpart of the TRL. Phillips' HRL aims to integrate HSI strategies throughout the system's life cycle. HSI can be broken down into nine functional areas, called 'HSI domains', that is Manpower, Personnel, Training, Human Factors Engineering (HFE), Environment, Safety, Occupational Health, Survivability and Habitability (Johnston et al., 2013). A description of each of the domains can be found in Appendix C. To develop the

HRL framework, Phillips and Acosta established four principles to guide the development of the HRL framework, that is (1) the framework should be in line with existing processes and complement these, (2) it should define a process that supports the development and deployment of cost-efficient and human-centred systems, (3) it should facilitate the integration of HSI early on, and (4) it should help to identify the costs associated with the implementation of HSI. Next, they developed the definitions of the framework levels. They decided to develop an HRL with nine levels, to stay in line with the structure of the TRL and, thereby, stimulate the acceptance of the HRL. For each level, Phillips described a general activity as well as more specific sub-activities. The first activity, affiliated with HRL1 – "Base-lining and commitment." - is to activate the HSI process, comprising sub-activities such as conducting preliminary functional allocation, initial HSI evaluation, and front-end analysis (Phillips, 2010, p. 34-36). The scale continues with "Human Systems Integration analysis in support of component technology development" (HRL2), "Refining requirements thresholds" (HRL3), "Component human touchpoint engineering parameters and human performance indicators" (HRL4), "Limited system human performance parameters demonstration" (HRL5), "Field validation of human performance prototypes" (HRL6), "Final developmental test & evaluation human performance parameters" (HRL7), "Operational test & evaluation human performance parameters" (HRL 8), and concludes with "Initiation of capability gap feedback cycle" (HRL9). An overview of the framework, including short descriptions of the levels, can be found in Appendix D.

1.2.2 Different HRL concepts

Other authors followed in Phillips' footsteps and created their HRL frameworks. This section will provide an overview of all the HRL concepts in chronological order.

Hale and colleagues (2011) created a framework similar to that of Phillips (2010) and called it the HF readiness levels (HFRL). Instead of the HSI domains, they used the 24 HF anchors. These anchors were taken from the FAA's Human Factors Job Aid (Krois & Rehmann, 2005, p. 420-421). The complete list of anchors, including short descriptions, can be found in Appendix E. The HFRL consists of a scale with seven levels, ranging from level 0 ("Insufficient evaluation of HF anchor") to level 6 ("HF anchor completely addressed"). Hale and colleagues proposed two steps for assessing readiness. First of all, they proposed that the relevance of each anchor needs to be evaluated. An HF anchor is relevant if it is expected that there will be human-systems issues related to that particular anchor. Secondly, for each one of the 'relevant' anchors, evaluators have to indicate the HFRL. The more has been done to identify and mitigate the impact of the HF anchor in the system, the higher the HFRL. Additionally, a risk assessment approach can be used to prioritise the identified HF issues. Next to assessing risk, Hale and colleagues (2011) suggest that the HFRL of the system as a whole can be determined by taking the average of the individual anchors' scores. When the system's HFRL is lower than its TRL, Hale and colleagues state that the HF readiness is considered neglected. In other words, the levels of the TRL and HRL should more or less be followed in parallel. This principle makes sense because if one would first 'complete' all TRLs and subsequently all HRLs, the human component would not be a continuous part of the development process. A potential disadvantage of the HFRL framework created by Hale and colleagues is that it is issue-focused rather than system-focused. As explained before, the first step of assessing the HFRL is to gauge the relevance of each of the HF anchors. If it is decided that a particular anchor is irrelevant, this anchor is not considered any further. Therefore, it is not clear whether the application of this framework enables the identification of HF issues that might appear only at later validation stages (e.g. during simulation-based testing). As a consequence, these issues could be overlooked. For example, imagine the following scenario: At the start of the systems development process, several issues related to the HF anchor 'Workload' are identified, and possible solutions are subsequently proposed and tested to reach a higher HFRL. However, the HF anchor 'Situational Awareness' was not deemed relevant at the start of the development process. So, a situational awareness related issue is now overlooked because the validation focuses on the solutions and scenarios related to the workload issues. The situational awareness related issue might lead to costly incidents after the system has been fielded. Moreover, it could be challenging to identify the relevance of HF anchors at an early stage in which there is only a technological concept. This example reveals a possible weakness of Hale and colleagues' issue/solution-focused approach rather than the technology/system-focused approach of the TRL and the HRL proposed by Phillips (2010). Another potential disadvantage of the HFRL is the use of HF anchors. On the one hand, the anchors are more precise and offer more clarity. On the other hand, the fact that there are many anchors (compare 24 anchors with nine domains) implies more cluttering, especially as each (relevant) HF anchor receives an HFRL score.

Jander and colleagues (2012) stated that the approach by Hale and colleagues (2011) to determine HFRL is rather process-oriented and decided instead to take a more methodology-oriented approach inspired by User-Centred Design (UCD). Based on a literature review, they distilled 29 relevant Human-Machine Interaction (HMI) criteria. They then conducted seven interviews in which they investigated the needs and lessons learned from the interviewees' experience with the development and evaluation of systems and let the interviewees refine the identified HMI criteria. Subsequently, the refined criteria were then reviewed during a workshop with HMI SMEs. The overall result is a methodological framework consisting of questions for assessing the importance and fulfilment of the HMI criteria for the system under development, a table with methods to evaluate the criteria and a matrix that can assign relative values to the criteria based on their importance and fulfilment. The importance and fulfilment criteria are to be rated on a scale ranging from 1 to 6, but it is not specified further how one should assign these ratings. Because the HMI readiness is a matrix that assigns momentary values to criteria, in contrast to the process-oriented concept which provides a scale that reflects at which point of the development the system currently is, it is conceptually more similar to a system usability scale than to the HRL framework introduced by Phillips (2010). Moreover, there are only three possible outcomes in the HMI readiness matrix, that is (1) no redesign needed, (2) further investigation needed and redesign recommended, and (3) redesign absolutely necessary. Thus, the readiness levels granularity is much lower compared to other concepts.

Johnston and colleagues (2013) tried to provide an alternative to the process-oriented approach and promote a focus on the state of current issues. They proposed to use an HF Readiness scale from 1 to 10. Like Hale and colleagues (2011), Johnston and colleagues use the 24 anchors to define HF and assign readiness levels to system issues rather than the system itself. Most importantly, they developed an outcome-focused HR readiness evaluation tool because one of the insights derived from interviews conducted with HF practitioners was that only communicating risks related to HF issues is not sufficient for supporting system development. Communicating risks is usually followed by devising a plan on how to resolve the issue. So the authors integrated the HFRL scale into a software tool called SHARE. In this system, a user can insert the level of risk and the resolution for each issue, and the SHARE tool subsequently calculates a readiness level for each HF issue. The level of risk is determined by its severity and the probability with which the issue is expected to occur. The resulting HFRL score matrix is very similar to the matrix of Jander and colleagues (2012), replacing criteria importance with risk level and criteria fulfilment with issue resolution. Through empirical testing of solutions, the readiness score can be progressed towards level 10. It is unclear why the authors chose this matrix format and ten levels while stating that their scale is comparable to existing TRL scales.

Miller and colleagues (2016) state that existing human readiness concepts - they specifically refer to Phillips (2010) and Hale et al. (2011) - differ significantly from the TRL concept. As an alternative, they created the Human Capability Level (HCL), composed of Training Readiness Level, Personnel Readiness Level, Manpower Readiness Level, and Habitability Readiness Level. The scales have nine levels, in accordance with the TRL. The authors proposed to assign an HCL to the different classes of individuals that interact with the system (e.g. vehicle controller, maintenance personnel) and TRLs to the separate technological subsystems (e.g. health monitoring system, automated driving agent). In contrast to the 'readiness' concepts introduced so far, Miller and colleagues are the only ones to mention that the readiness level can decrease. For example, a budget cut might mean less personnel which impacts the Personnel Readiness Level resulting in a decreased HCL. Next to the HCL, Miller and colleagues (2016) proposed to use Human Integration Readiness Levels (HIRL), which specify and communicate the readiness of the human-technology interactions within the system. The HIRL scale is the human equivalent of the Integration Readiness Levels (IRL) and reflects the following HSI domains: HFE, Safety, Survivability, Occupational Health and Environment). Miller and colleagues do not further specify what falls under HFE. They also do not explain why they divided the HSI domains among the HCL and the HIRL in the way they did. One could argue, for instance, that it is difficult to consider training readiness separately from the integration between the human and the machine.

In 2019, an HRL working group was established to mature the HRL scale and validate its utility in various scenarios (M. Endsley, personal communication, February 24, 2020). The working group developed the factsheet that can be found in Appendix F. The scale presented in the factsheet consists of 9 levels. In HRL 1, relevant human capabilities, limitations, performance issues and related risks must be identified. In order to progress to HRL 2, a human-focused concept of operations needs to be defined, and Human Performance (HP) design principles should be established. To achieve HRL 3, it is necessary to analyse the human operational, environmental, functional, cognitive, and physical needs based on a proof of concept. Before attaining HRL 4, modelling, part-task testing, and trade studies of user interface design concepts must be completed. Then, for HRL 5, prototypes undergo user testing in a relevant simulation. To satisfy the criteria of HRL 6, the human-system interfaces need to be fully matured, based on HP (HP) analyses, metrics, prototyping, and high-fidelity simulations. Subsequently, human-system interfaces should be tested and verified in an operational environment with system hardware and software and representative users, yielding an HRL 7. For an HRL 8, the total human-system performance must be tested, validated, and approved in an operational context, using system hardware and software and representative users. To advance to HRL 9, the system should be successfully used across operations, and the human-system performance should be monitored consistently. From this scale, it becomes apparent that there is an increasing stringency of validation at each level; higher levels of readiness require the validation of more detailed concepts in more realistic environments. An elaborated version of the framework, including case studies on the application of the HRL, will appear in a special issue of Ergonomics in Design (P. Savage-Knepshield, personal communication, November 3, 2020). The framework presented there will be the most recent and elaborated version so far. A preliminary version of this framework can be found in Appendix A. Apart from some minor changes, the new HRL framework mainly distinguishes itself from the factsheet by providing more information; core questions, sub-questions, considerations, exit criteria, supporting evidence and links to the TRL. The HSI domains are integrated into the sub-questions. Take, for example, sub-question 17: "Are manpower, personnel, and training analyses being completed?". The framework is very similar to the HRL of Phillips (2010). They not only contain the same number of levels and the same domains, but both frameworks are process-oriented and system-focused.

As we have now seen, there are various HRL concepts, and they differ markedly in their approach. Table 1 provides an overview of the reviewed concepts and their main characteristics.

Table 1

Overview of different HRL concepts and their main characteristics

Authors	Name	Functional areas	Number of levels	Process-oriented or focused on current state	Issue-focused or system-focused	Individual sub-scales included?	Risk assessment included?	Methods included?
Phillips (2010)	Human Readiness Levels	HSI domains	9	Process	System	No	No	Yes
Hale et al. (2011)	Human Factor Readiness Levels	HF anchors	7	Process	Issue	Yes	Yes	No
Jander et al. (2012)	Human-Machine Interaction readiness	HMI criteria	3	Current state	System	No	No	Yes
Johnston et al. (2013)	Human Factor Readiness Levels	HF anchors	10	Current state	Issue	Yes	Yes	No
Miller et al. (2016)	Human Capability Levels & Human Integration Readiness Levels	HSI domains	9	Process	System	Yes	No	No
HRL workgroup	Human Readiness Levels	HSI domains	9	Process	System	No	No	No

1.3 HRL issues

Although the HRL comes with considerable benefits and could potentially contribute to the prevention of human-factor related catastrophes like those of the Boeing 737 MAX, there are no examples yet of its application. The HRL concept only seems to exist in the academic world, and even there, we cannot speak of one well-supported framework. As discussed before, there are various HRL concepts out there, and authors seem to prefer creating their own HRL rather than building on the foundation of previous researchers as they rarely cite previous research efforts and when they do, it often does not go further than one or two sentences in which they merely mention, rather than critically consider, the existing concepts. Some frameworks focus on the HF anchors (Hale et al., 2011; Johnston et al., 2013), while others, including the one developed by the HRL workgroup, follow the HSI domains structure (Miller et al., 2016; Phillips, 2010). Additionally, there is even a framework that formed its own human HMI criteria categories (Jander et al., 2012). In addition to this issue, there seems to be disagreement about whether the HRL should contain a risk assessment, whether it should focus on the process or the current state, and whether it should propose activities and methods. The frameworks that suggested the inclusion of a risk assessment only offer the possibility for someone to subjectively decide the relative risk level based on the estimated likelihood and severity of an issue (Hale et al., 2011; Johnston et al., 2013). A couple of frameworks try to offer an alternative to the more process-oriented approach by Hale et al. (2011) by specifying methods (Jander et al., 2012) or focusing on the current issues (Johnston et al., 2013) but consequently lack the process guidance and risk to become 'just' a usability scale for one-off measurements. Finally, most of the frameworks ended up losing their resemblance to the TRL (Hale et al., 2011; Jander et al., 2012, Johnston et al., 2013).

Héder (2017) critically reflected on the use of the TRL, and his arguments could be a valuable *a priori* lesson for the HRL. One of his main points is that the usefulness of the TRL's application cannot be assumed in every field of application but should be proved. He is afraid that, without discipline-specific guidelines, the TRL will sow confusion and get abused by those interested in obtaining EU funding. He further warns that various disciplines might already have ways of managing maturity. In these cases, the TRL could pose an additional external constraint rather than result from an 'internal evolution'. This could be the case for HRL, too, especially as there are no use cases yet.

1.4 Description of research objectives

The present work aims to contribute to the development of the HRL, taking the HRL created by the HRL workgroup as a starting point and improve it through two evaluation rounds. This HRL framework was chosen as it is the most recent and mature one. It is also the most exhaustive framework, describing activities, core questions, sub-question, considerations, exit criteria, supporting evidence, links to the TRL and an application example. From now on, I will refer to this HRL framework with 'HRL v1'. An overview of the structure of the present work can be found in Figure 1. First, the process prescribed by HRL v1 will be compared to *current practices*, that is, the work as done up until now without using the HRL. More specifically, several past and ongoing EU-funded Aviation projects and established validation frameworks will be studied to understand how systems are currently developed and validated. The comparative analysis of current practices and the HRL v1 will focus on identifying the gaps and the overlaps between current practices and HRL v1 and critically reflecting on these to distil points of improvement for the HRL v1. Additionally, a safety and HF specialist will be asked to give feedback on the HRL based on his extensive experience managing R&D projects. It is important to stress that it is not assumed here that current practices are perfect and that the HRL should be a mere reflection of these. Instead, the HRL v1 may add essential elements to systems development that are currently not commonplace. Thus, the present work will look at what important elements the HRL v1 might be missing but also at its utility or added value for current practices in European aviation research. The HRL v1 was explicitly compared with current practices from the field of aviation because Héder (2017) expressed the need to prove the usefulness of the TRL's application in every field of application. The present work considered this to be a potentially relevant consideration also for the HRL framework. Next to current practices, I will also analyse the HRL by comparing it to other approaches in the area of socio-technical systems development as described in the literature. The literature review will focus specifically on Cognitive Systems Engineering (CSE). CSE is an approach for designing cognitive systems, that is, systems in which joint cognitive activity of human and machine take place (Hollnagel, n.d.). In these systems, both the human and the machine have an (implicit) 'image' of the other's cognitive characteristics and these images need to be matched. CSE take a genuine systemic perspective rather than looking at system components and the interaction between them. This approach is especially useful for designing new complex socio-technical systems, which are also the target application of the HRL. CSE has been influencing the field of human-machine systems since 1982 and combines concepts from various disciplines such as engineering, psychology, cognitive science, information science, and computer science. This makes it a very interesting and rich resource for comparison with the HRL. The outputs of this comparison will be used to identify additional points of improvement for the HRL.

Next, the present work will design an adapted HRL based on the potential points of improvement that arose during the comparison analysis. The adapted HRL will be called HRL v2. The HRL v2 will be subjected to a user test, in which several HF experts are asked to use and evaluate the usability and utility of the process tool. The experts will be selected to represent different industries in order to explore the usefulness and utility of the HRL outside the field of aviation. Finally, open issues and further recommendations will be presented.

The present work was completed in collaboration with Deep Blue, a research and consultancy agency focusing on the role of the human in safety-critical and high-tech systems. Deep Blue provided the insights into their current practices that were necessary for the present work to succeed. Therefore, this thesis also has the goal to provide the company with valuable insights. First of all, I intend to explore to what extent Deep Blue's current way of working is already in line with the HRL framework. Secondly, an improved version of the HRL, that is, the HRL v2 will be provided to Deep Blue to apply in future projects and reap its fruits.

The aims of contributing to the development of the HRL as well as to the practices at Deep Blue are compatible: The application of HRL by Deep Blue in future projects would, in turn, also be a triumph for the HRL.

Taking all activities together, this study sets out to achieve the following five objectives:

Objective 1. Propose potential points of improvement for the HRL based on a comparison with current practices and relevant literature.

Objective 2. Validate the utility of the HRL within European Aviation projects.

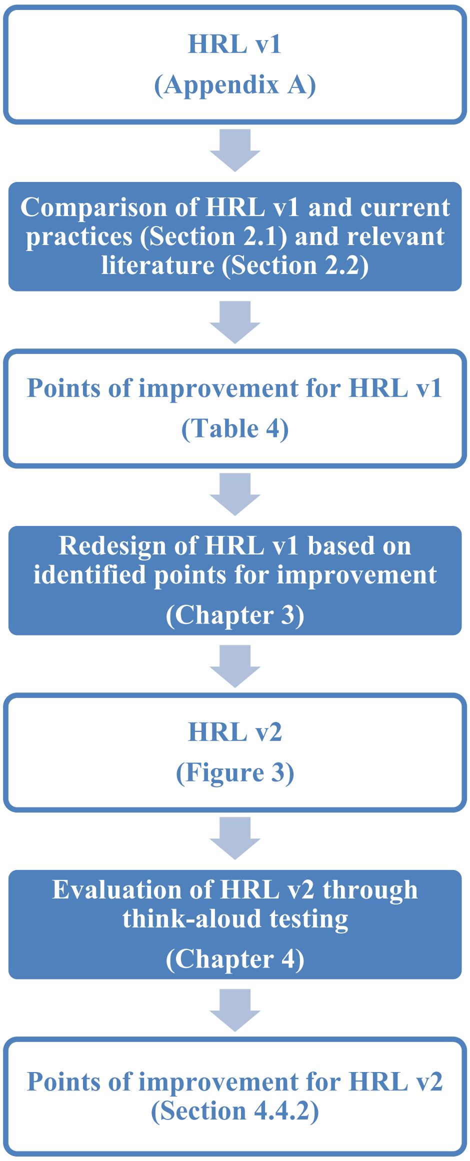
Objective 3. Evaluate the HRL v2 on its perceived usability and utility.

Objective 4. Explore to what extent current practices at Deep Blue are already in line with the HRL framework.

Objective 5. Provide Deep Blue with an improved version of the HRL.

Figure 1

Flowchart of the research structure. The blue boxes represent the activities performed, whereas the white boxes represent the inputs and produced outputs. In brackets is the *location of the products and descriptions of the activities.*



2. HRL v1 analysis

This chapter presents an evaluation of the HRL v1 through careful consideration of current practices at Deep Blue and CSE approaches as described in the literature.

2.1 Current practices

To understand how big the gap between current practices and the process prescribed by the HRLs is and to identify potential points of improvement for the HRL framework, a comparison was made between the work as it is imagined by the HRL and the current practices in Deep Blue projects. Next to identifying what the HRL v1 might be missing or lacking, this analysis also uncovers how HRL captures and complements current practices.

In total, five projects were studied. Additionally, special attention was paid to four existing HF maturity frameworks within the area of European Aviation projects. Finally, a Safety and HF expert provided his impression of the HRL based on his experience managing R&D projects. The projects and the expert were arranged by Deep Blue. Deep Blue is a research & consultancy agency addressing socio-technical challenges (https://dblue.it). They are located in Rome, Italy, and operate on a European scale, focusing on the role of the human in safety-critical and high-tech systems. Deep Blue participated in and led research programmes funded by international frameworks and agencies such as Horizon2020, Single European Sky ATM Research (SESAR), and Clean Sky. They have extensive expertise in Transportation, Healthcare and Environment, but their core business is in Aviation. The five projects selected for this analysis and the four validation frameworks all fall within the area of aviation.

2.1.1 Innovation projects

The projects were selected based on the richness of the available information. The collaboration with Deep Blue allowed me to access validation reports and other relevant unpublished or confidential documents of projects. The five projects that were selected were the best-documented ones. In case any relevant information was missing in the documents, this was resolved by directly contacting a Deep Blue employee that was or is involved in the project. This section will briefly introduce the projects and present the main findings that resulted from comparing the validation approach in the projects and the HRL v1 process. The full description of the analysis and comparison can be found in Appendix G.

ALICIA. Together with 40 European partners, Deep Blue participated in ALICIA (CORDIS, 2019a). ALICIA is a project that ran from 2009 to 2014 and aimed to increase time efficiency within the future Air Traffic Management (ATM) system. In order to contribute to this objective, the research project focused on developing innovative cockpit applications that improve operations of an aircraft under degraded (weather) conditions, or as they called it: 'All Condition Operations'.

ACROSS. The project Advanced Cockpit for Reduction Of StreSs and workload (ACROSS) ran from 2013 until 2016 (CORDIS, 2019c). In total, it had three objectives (Billiere et al., 2013, p. 5). It aimed at supporting the management of peak workload situations with the ultimate goal of contributing to the successful implementation of reduced crew operations. In order to do this, ACROSS integrated different technologies into an operational concept, called the 'Global Operational Concept' (P. Lanzi, personal communication, January 8, 2021).

NINA. NINA stands for Neurometric INdicators for ATM and is composed of two strands of work (Terenzi et al., 2014, p. 5). The first strand of work explored how the cognitive state of an air traffic controller (ATCO) can be described based on the monitoring of specific physiological parameters. The second strand of work explored how adaptive HMI solutions could be triggered by the cognitive states of the ATCO. In total, the project lasted slightly more than one year, starting in 2014.

STRESS. From 2016 to 2018, project STRESS took place, coordinated by Deep Blue (CORDIS, 2017). STRESS is about HP neurometricS Toolbox foR highly automatEd Systems deSign. As the title already partly gives away, STRESS concerns a project that is very similar to NINA. There were also two validation activities (Bonelli et al., 2017, p. 10). First of all, the project developed and verified indexes to assess attention, stress, workload and cognitive control. Secondly, the project had the objective of studying the impact of automated tools on HP.

SAFELAND. The last project that will be discussed in the context of this study is the project SAFELAND. SAFELAND is also coordinated by Deep Blue and started in 2020. The project is planned to be completed at the end of 2022. Apart from reviewing available documents, I spent six months observing the activities in project SAFELAND. The observed activities included meetings with consortium partners, informal discussions of Deep Blue colleagues and stakeholder workshops. SAFELAND aims to create a solution that can provide the SAFE LANDing of incapacitated single-piloted aircraft through enhanced ground support (CORDIS, 2021).

Main findings. Based on the comparison between the HRL v1 and the development and validation practices in the projects mentioned earlier, several elements that the HRL v1 is currently missing were identified. In contrast to the validation process in ALICIA, the HRL v1 does not include acceptance criteria and requirements for the validation exercises. Additionally, several validation assumptions were identified as part of project ACROSS and SAFELAND, like the assumptions of having a latency-free data link between the ground station and aeroplane in SAFELAND. These assumptions need to be fed back into the technology development (TRL), and it needs to be ensured that the assumptions are met at some point. Instead, the identification of assumptions is not mentioned in the HRL v1. Furthermore, the HRL v1 does not consider the impact on team factors, communication, social organisation, cooperation, commercial objectives and organisational culture, whereas these were deemed important in ALICIA, ACROSS and SAFELAND. In general, the HRL v1 focuses much more on the tasks that need to be performed using the system and the interaction between the operator and the system, and not so much on the context or environment in which the tasks have to be conducted and the interdependencies between different humans in the system. For example, the HRL v1 starts with identifying cognitive constraints (or 'user capabilities and limitations') while SAFELAND started by analysing the broader work domain constraints. Other aspects that the HRL v1 lacks compared to ACROSS and SAFELAND were the system's acceptability for the operator and other stakeholders and information presentation or flow. Furthermore, whereas the HRL v1 talks about 'human-system performance' only in a general way, HP was specified in ACROSS, NINA, STRESS and SAFELAND, and the HSI domains were not used. Finally, some projects conducted a risk assessment as part of the development and validation process. In addition to identifying elements that the HRL v1 lacks, a potential issue had been identified. In ALICIA and ACROSS, the technology had already partly been developed before HF were considered. Although it is clear that this is precisely the problem that the HRL tries to tackle, it is not clear what the strategy of the HRL would be or how it should be used in these cases.

The comparison between HRL v1 and the various projects also led to the identification of areas of overlap. Both the HRL sub-questions and project activities cover topics such as basic human research, performance requirements, metrics for successful HP, strategies for supporting human use, conformance of preliminary design to HP guidelines, key HP design principles, usage scenarios, user roles, procedures, tasks, human-machine function allocations, human-system issues, manpower, personnel, training, and the drafting of test plans and validation reports. The flow of the levels of the HRL v1, from identifying relevant HF guidelines, establishing needs and requirements for supporting HP, defining prototypes, to conducting a preliminary user test, also overlapped with the process that took place in NINA. Furthermore, there was a need to express maturity from an HF or operational point of view in all projects. Most of the projects fulfilled this need by using the Concept Life cycle Model (CLM), which is part of a validation framework that will be discussed in more detail in Section 2.1.2. As mentioned before, SAFELAND is an ongoing project. At the end of the project, it will need to identify the next steps towards higher maturity, and the HRL v1 could be useful for this.

2.1.2 Established frameworks for managing maturity

When analysing the project cases, it became apparent that there are already some structures to support the validation process with particular attention to HF. Héder (2017) already warned that various disciplines might have ways of managing maturity and that the TRL could pose an additional, external constraint rather than being the result of an 'internal evolution'. The same might be valid for the HRL. The current section will provide a closer look at relevant frameworks within the field of aviation. More specifically, the HF Case, the European Operational Concept Validation Methodology (E-OCVM,) the Human Assurance Level (HAL) and the SESAR Human Performance Assessment Process (HPAP) will be considered. The complete discussion per framework can be found in Appendix H. Only the main outcomes will be presented here.

Several additional gaps between the HRL v1 and established practices were identified based on the comparison to established validation frameworks. The HF Case, E-OCVM and the SESAR HPAP are more structured compared to HRL v1. For example, in the HRL v1, managerial and system-related sub-questions are lumped together, as are sub-questions on domain-level and sub-questions that pertain to more specific elements of an HSI domain. Also, the HF Case and SESAR HPAP both provide a specification of 'HP'. Moreover, the E-OCVM, the HAL and the SESAR HPAP offer examples of tools that can be used at different maturity levels. Furthermore, in line with the analysis of the projects, the validation frameworks captured some essential factors that the HRL v1 is lacking. For example, the HF Cases uses the HF categories, which capture HF elements the HSI domains do not capture, such as team and communication. See Table H1 for the definition of each of the HF categories (Mellett & Nendick, 2007). Similarly, the HAL covers people factors that relate to the social organisation of people (HAL) that go beyond manpower and personnel and in the SESAR HPAP communication is considered. In contrast to the HRL v1, the HF Case and E-OCVM specify the engagement of stakeholders other than system users. Additionally, the E-OCVM captures the validation assumptions that are made as part of system validation exercises. The HAL framework includes a risk assessment and suggests validation activities based on the potential risk in the system. Finally, the HAL and the SESAR HPAP both designed a strategy for dealing with technologies that have already been partly developed and therefore have a technological maturity level that is much higher than the HRL. Another potential point of improvement of the HRL v1 extracted from the analysis of the validation frameworks is that the HRL v1 should clarify how it relates to other maturity scales, such as the E-OCVM's CLM. As shown in Figure H2, an attempt at this was made as part of the analysis. This mapping is based on the mapping between the TRL and the E-COVM CLM V-phases found in various documents (Billiere et al., 2013; Morgan et al., 2018; SJU, 2020). Nevertheless, the information in these documents is not aligned, so it remains to be seen whether the mapping is appropriate.

Next to the gaps and areas for improvement, the analysis of the validation framework also resulted in the identification of overlaps and potential added value of the HRL v1. Like the E-OCVM, the HRL v1 assesses both whether the *right* system is being built (i.e. system validation) as well as whether the system is built *in the right way* (i.e. system verification). Both frameworks make it possible to link a validation activity to a certain sub-question or exit criteria, thereby keeping an overview of the effort that has already been put into the development and validation of a system across different programmes, projects and work packages. Additionally, both frameworks can assess maturity as influenced by HF and prescribe a process in which HF are taken into account from the beginning of technology development. However, the HRL v1 offers more maturity levels and thus more granularity compared to both the E-OCVM CLM and the HAL and is more in accordance with the TRL. Lastly, the following topics are covered both by the sub-questions of the HRL v1 and the various validation frameworks: human capabilities and limitations, user roles and procedures, human-machine function allocations, support for human use, manpower, personnel and training.

2.1.3 Expert viewpoint

Next to the review of various projects and frameworks, a project manager within Deep Blue was interviewed to gain preliminary feedback on the HRL from a practitioner's point of view. The interview was conducted with Luca Save, an Safety and HF expert at Deep Blue. He holds a PhD in Cognitive Ergonomics from the University of Florence and has been working within the field of HF since 1999. He has been at Deep Blue since 2004, where he takes care of R&D project management. Two weeks before the interview took place, the HRL framework was sent to Mr Save, from now on referred to as the *DB project manager*, so that he could have a closer look at it and form a first impression. At the beginning of the interview, the DB project manager was asked to describe an aviation project he had been involved in and reflect on the HRL v1's relevance to the project in hindsight. He was then invited to share his comments on the comprehensibility of the framework. After the interview, the DB project manager was asked to review this section to ensure that it reflects his impressions and opinions well.

Selected case. As an example, the DB project manager mentioned SPIN (Safety Nets Performance Improvement Network). SPIN is a project that aimed to develop specifications and guidance material for the design of Safety Nets, which is a set of functionalities integrated into the ATC's working positions to alert them in the imminence of a risk. The risks addressed by the Safety Nets included those deriving from conflicting trajectories of aircrafts, aircrafts flying too close to the terrain, and flights infringing restricted areas, such as those used for military exercises.

HRL v1 utility. The DB project manager perceived most of the sub-questions as relevant and said usage scenarios and task descriptions (see Appendix A; sub-questions 2, 8, 12, 13) played a prominent role in the project. Since the operational environment could vary from an airport to an open space high up in the air or even a military area, the system needed to adapt to several scenarios. Similarly, the HRL sub-questions about human-machine function allocation (sub-questions 14, 29, 40) would have been useful in the Safety Nets project as the level of automation was a vital part of the project. Thus, the first impression is that the HRL nicely captures the process followed naturally in the Safety Nets project.

However, the DB project manager noted that the HRL framework seems to be most suitable for projects that are at the beginning of the development life cycle. From his experience, he concluded that HF specialists are often engaged when the system is already in place, and the project will consist of a simulation although all the steps before might have been skipped. In that way, the system could be assigned HRL 5 while users were not involved in the preceding steps. The DB project manager wondered how the HRL deals with this risk.

During the interview, the DB project manager also noted that most of the projects he is involved in concern broader concepts rather than specific technologies, such as the head-mounted display offered as an example by the HRL framework. Procedures, or working methods, should have a more prominent role in the HRL v1 to accommodate the development and validation process of more complex systems. Procedures were indeed frequently mentioned in the Deep Blue projects that are reported in Section 2.1.1. Moreover, procedures and activities are often influenced by regulatory constraints. To illustrate this, the DB project manager referred to the 2002 Überlingen mid-air collision. The leading cause of this collision was that the Traffic Collision and Avoidance System (TCAS) and the air traffic controller provided the pilots with conflicting orders (SKYbrary, 2018). Since the accident, the International Civil Aviation Organization changed the international requirements so that pilots are required to follow TCAS advice regardless of air traffic control instruction. This regulated procedure now influences an operator's behaviour, explained the DB project manager, and should therefore be considered when developing new ATM systems. So, to be beneficial for complex systems development, the HRL framework should consider systems in their context.

Furthermore, the DB project manager felt that the framework is slightly biased towards a particular industry. First of all, because of the terminology that is used (e.g. 'mission'). Secondly, the DB project manager stated that in many of the projects he supervises, it is not necessary to conduct such a sophisticated ergonomics analysis as prescribed by the HRL. Most of the information related to human capabilities and limitations is taken for granted at the beginning of systems development. For example, engineers know that they cannot 'overload' a person, but it will only become clear how much workload exactly there will or can be later on in the development. Similarly, the potential risks and issues are generally analysed after the task analysis. Moreover, the HRL only treats human errors on a very high level of abstraction, which would not be precise enough for safety-critical projects Deep Blue deals with. The DB project manager remarked that they often conduct a hazard analysis such as the HAZOP. This analysis is currently missing in the HRL. On the other hand, the HRL contains sub-questions about Manpower, Personnel and Training (sub-questions 17, 25, 36, 49, 60, 71), about which the DB project manager was very positive. In conclusion, the HRL framework should acknowledge that human capabilities, limitations and errors take shape and start to make sense later in the development when tasks have been developed, and that training can subsequently serve to eliminate limitations and errors.

HRL v1 comprehensibility. When seeing the HRL, the DB project manager said he had experienced confusion about its relationship with other works such as the Standard of Excellence in HP, the HAL, and the SESAR HPAP. Moreover, most of these frameworks are based on a rationale. For example, he explained that the HPAP categories were based on a model which groups issues in different areas: Human-Technology, Human-Human, Human-Process, Transitions. This underlying model makes the HPAP better understandable. Although the HRL is also an analogy, namely to the TRL, it was not clear to the DB project manager based on what theoretical model the selection of sub-questions had been made. Furthermore, he recognised that there seems to be a recursive pattern of criteria across the levels. He noted that the sub-questions repeat themselves, constantly on a more advanced level: In HRL 3, you analyse; in HRL 4, you need to develop a strategy; in HRL5, you test this strategy, and so on. However, this is not very (visually) apparent and could be made more explicit. There was also a small side note on confusing terminology. At HRL 3, the word 'completed' is used (e.g., "Are safety analyses for human users being completed?", see Appendix A; sub-question 16). More correct would be to say 'performed' or 'conducted'. In addition, the DB project manager noted that the sub-questions seem to be a mix of goals and methods. Moreover, there seems to be some kind of hierarchy among the sub-questions in the HRL. For example, the sub-questions about requirements (sub-question 15) cover the ones about usage scenarios, function/task descriptions and human-machine function allocations (sub-question 12, 13, 14). Concurrently, the answers to sub-questions 12, 13, 14 are needed for the safety analysis (sub-question 16). Additionally, sub-question 22 (evaluating HP data to determine the feasibility of metrics) seems to be more at a process level in the sense of "Have you done this?" while other sub-questions seem to represent substantive understanding: "Have you understood this?" In conclusion, the HRL could support its potential users by clarifying how it relates to other existing frameworks, and some adaptations could be applied to increase consistency within the HRL framework.

Lastly, the DB project manager noted that the questions seem to be too open to different interpretations. He warned that the questions will probably be interpreted differently based on the person and organisation by which the HRL is applied. A Toolkit could help to tame these interpretations. The DB project manager was working on a Toolkit together with colleagues as part of project SAFEMODE. This Toolkit gives examples of three different types of methods: processes, models and techniques. He thinks this classification is a good idea because a user should be able to distinguish among different uses of methods. For example, there are methods to organise an HF intervention (Process), to understand better a given problem (Models), and to conduct a concrete analysis with step-by-step indications (Techniques). The HRL can be considered a process that includes models (e.g. the HSI domains) and some techniques (e.g. Task Analysis) but does not differentiate them. Adding techniques and separating them from the rest of the HRL by grouping collecting them in a Toolkit could improve the HRL v1's usability.

The identified points of improvement for the HRL v1 that were identified based on the interview with the DB project manager are summarised in Table 2.

Table 2

Points of improvement for the HRL, based on the interview with an R&D project manager from Deep Blue

The current HRL v1 framework does not provide specific examples of tools, such as the HAZOP, to identify human errors.
It is unclear whether and how to apply the HRL v1 framework when technologies have already been partly developed; the system could be assigned HRL 5 because a simulation took place while users were not involved in the preceding steps.
The head-mounted display example offered by the HRL v1 framework is a very straightforward technology rather than an extensive complex system.
Procedures, or working methods, do not have a sufficiently prominent role in the current HRL framework.
The current HRL v1 framework does not consider the context of the system.
In the current HRL v1 framework, it is not very (visually) apparent that the sub-questions repeat themselves, constantly on a more advanced level.
The current HRL v1 framework contains jargon (e.g. 'mission') and confusing terminology (safety analyses 'completed' at HRL3).
The current HRL v1 framework focuses too early in the development process on human capabilities, limitations and errors (even before the definition of tasks).
It is unclear how the HRL v1 framework relates to established maturity and validation frameworks.
It is unclear on which theoretical model the HRL v1 sub-questions are based.
In the current HRL v1 framework, process-level/methodological sub-questions are mixed with (both method- and goal-related) validation sub-questions.

2.1.4 Lessons learned from the comparison between the HRL v1 and current practices

The analysis of five ATM innovation projects, four validation frameworks, and the interview with an R&D project manager has provided insights into the differences and similarities between the HRL v1 framework and established validation strategies in practice. One of the critical learnings was that technology is often already developed to a certain extent by the time the operational or HF effort starts. It is not clear whether and how the HRL could be used in those situations. There is a risk that new operational concepts are tested in a simulation and achieve HRL 5, although the previous levels were skipped. Some of the established validation frameworks that we have seen take this scenario into account and have developed strategies. The HRL framework could adopt these strategies. In order to increase the uptake of the HRL by project managers and validation practitioners, it is also essential that the relation between the HRL and these other validation frameworks is clarified. Not only in terms of what benefits one has compared to the other but also in terms of mapping the levels on other maturity scale phases and understanding to what extent sub-questions and other objectives are complementary. Another important realisation is that the HRL v1 framework does not consider the broader context of the system that constrains the system. It advises conducting an environmental analysis, but it is unclear whether 'environment' actually stands for the 'work domain' or just the 'physical environment'. Moreover, this environmental analysis takes place after human capabilities, limitations, and errors have been identified. By looking at current practices, including the opinion of an experienced project manager, it can be argued that it is too early to say something useful about human limitations and errors at the beginning of the development and before the work domain and the tasks have been defined. Perhaps the environmental constraints that are missing the most in the HRL v1 are the social constraints. The HRL v1 focuses on the Liveware-Hardware and Liveware-Software components and misses the Liveware-Liveware ones, such as team structure, communication, information flow, and other factors relating to the system's social organisation and its objectives, culture and business model. These critical elements are not captured by the HSI domains or the other HRL v1 sub-questions (see Appendix A). In general, the HSI domains were not mentioned in any of the current practices. Perhaps the use of HSI domains is specific to certain industries or countries. It was also found that how HRL sub-questions were formulated based on the HSI domains is somewhat confusing. Some sub-questions are directly tied to one of the domains, whether HFE is not explicitly mentioned but covered by most of the other sub-questions that do not specifically address one of the other HSI domains. There is also a sub-question for the 'left-over' HSI domains that are not mentioned in one of the other sub-questions but could be relevant (i.e. Survivability, Habitability, and Occupational Health). This leads to a mix of sub-questions on domain-level with sub-questions about certain HSI domain elements. For someone using the HRL, it is unclear based on which model or reasoning the decision for the specific sub-questions was made. Not only could the HRL v1 framework be improved by using a relevant model as the backbone on which sub-questions can be based in a consequent manner, but it would also benefit from an HP model. The HRL framework aims to prevent HF from being overlooked. However, it does not define the HF by which HP could be influenced. Significant HF, such as acceptability, workload, situation awareness, stress, cognitive control, trust, and satisfaction, are not mentioned. If the HRL is to make a difference next to the established TRL, it should treat these factors and not only task performance.

The potential confusion concerning the HRL sub-questions goes beyond the issues related to the HSI domains. The HRL v1 framework mixes managerial activities and questions on programme and project level (see Appendix A; e.g. sub-question 5), relating to the validation process, with activities and questions on exercise level (e.g. sub-question 6), relating to the content of the validation. The sub-questions are also a mix of goals (e.g. Have strategies to mitigate safety implications for human users been identified and recommended?) and methods (e.g. sub-question 16: "Are safety analyses for human users being completed?"). Apart from these rather abstract indications of methods, the HRL does not provide examples of tools that could be used at each level. This is a pity, as it could help to more clearly convey the pattern throughout the HRL, namely that of increasing precision as the maturity increases. In the HRL v1, it is not very visually apparent that the sub-questions repeat themselves constantly on a more advanced level. The relationship between the sub-questions could be made more evident.

Since some of the projects and frameworks included a risk assessment, it would be easy to conclude that the HRL should also incorporate this. Alternatively, specialised existing risks assessments could be used complementary to the HRL. Furthermore, there is no strong need for the HRL to base its validation process on the potential risk of the system that is being developed, like the HAL does, because the HRL framework is specifically meant for high-risk safety-critical systems. This target application could be expressed more clearly to prevent confusion and faulty assumptions. Furthermore, it was observed that the definition of validation assumptions are a common practice in the projects that have been analysed. During development and validation, certain assumptions about the technology were made. These assumptions are an essential link between the TRL and HRL and should be visualised in the HRL framework.

Another gap between the analysis of current practices and the HRL framework is that stakeholders, including users but also other stakeholder communities, do not seem to play an active role in systems development according to the HRL framework, whereas Deep Blue uses a UCD approach. In the Deep Blue projects, stakeholders are actively involved in the needs identification, scenario generation, concept development, and function allocation.

As said at the beginning of the analysis, this thesis does not only look at points of improvement for the HRL but also at how the HRL framework captures and complements current practices. The presence of ATM or even SESAR specific process frameworks to structure the validation process and assess maturity from an HF point of view, and the eagerness of projects to adopt these, shows that there is a need that the HRL could fulfil. At the beginning of a project, current and target maturity are often identified, and relevant validation activities are selected. The TRL is often used but insufficient as the success of the operational concept depends on more than just technical readiness. The HRL is an appropriate solution that can help take HF into account from the beginning of systems development. The HRL's assets are its resemblance to the TRL and its fine granularity. Many HRL sub-questions have proven to be helpful when analysing the practice in hindsight. The HRL framework also takes into account both validation as verification concerns. Finally, as the practice shows, systems development and validation are often divided over numerous programmes, projects and work packages. The HRL framework makes it possible to link a specific validation activity to a certain sub-question, thereby keeping an overview of all HF efforts put into developing the system. Moreover, these efforts will be based on the shared principles and practices of the HRL, making it easier to build upon them for future projects.

2.2 Cognitive systems engineering

In the previous section, the HRL v1 has been compared to current practices in order to collect valuable insights on how to improve the framework. This section will review relevant literature on the engineering of cognitive systems and compare suggested approaches to the HRL v1 in order to collect additional points of improvement. The approaches and methods described in this section are particularly suitable for designing innovative future practices.

Vicente (1999, p. 101) argued that when we base the design of a technology on workers' current tasks, the new technology will “introduce constraints that often radically redefine the task for which it was originally developed”. In this way, the designer will always be one step behind. A *formative approach* helps to circumvent this so-called task-artefact cycle because it does not focus on specific tasks. Instead of focusing on the way work *should be* done (i.e., the 'normative approach'), an ideal that is rarely achieved, or on the way work *is* done (i.e., the 'descriptive approach'), the formative approach focuses on the way work *could be* done. More specifically, a formative approach identifies the (technological and organisational) requirements that need to be satisfied for a system to support work effectively. Rather than prescribing what should be done or how work is currently done, these constraints specify what *should not* be done - the boundaries of good work practice - thereby offering more flexibility and less dependency on current practices.

2.2.1 Degrees of freedom

Building forth on the work of Jens Rasmussen, Kim Vicente advocated a formative approach to the design of systems with the aim of leaving space for expert users to 'finish the design', that is, to respond to local circumstances in real-time (Vicente, 1999). It might seem counter-intuitive, but this increased freedom is expected to lead to a higher level of safety. That is because designers can never predict all the possible failures that could occur, so, in the face of unexpected events, they will have to capitalise on experts' adaptations. Onboard Lion Air's Boeing 737 MAX, neither the memorized checklists nor the Quick Reference Handbook applied to the critical situation at hand (Campbell, 2019). In the field of safety, we might often focus so much on the mistakes humans make that we tend to forget that humans can also positively impact the risk in a system as they can mitigate risk. Vicente wrote: "In complex socio-technical systems, the primary value of having people in the system is precisely to play this adaptive role" (Vicente, 1999, p. 121). Similarly, as captured by one of the general laws of adaptive systems: "Only [human] variability can destroy variability [in the physical world]" (Woods & Hollnagel, 2006, p. 19). However, as Vicente warns, we should deliberately design systems that help workers be reliable and effective adaptive actors rather than merely expecting them to be adaptive actors (Vicente, 1999). Experts should be given appropriate degrees of freedom to enable adaptive behaviour. To illustrate this, Vicente uses the example of a navigational task in which one has to find a destination in a city. This task can be supported either by a map of the city or a set of directions. Whereas the map shows the constraints of the navigational space, the set of directions provide an instruction. The directions might require less thinking, making it more efficient, but both instruments will lead the user to the destination. However, now imagine that an alternative route has to be taken to avoid an unexpected obstacle. The user with the map will still find the destination, whereas the set of directions became utterly futile. Thus, maps are more flexible and general. Moreover, maps are event-independent, meaning they can be used regardless of variables such as the starting and destination point. They also contain redundant information, making it easier for users to realise when there is a flaw in their mental representation. Checklists and quick reference handbooks are examples of ‘a set of directions’. On the other hand, a system with degrees of freedom is like a map. It shows the constraints and supports experts’ flexibility and adaptation, making it more resilient in the face of unanticipated events. This benefit should be enough of an argument for applying a formative approach to system design, but it is good to mention that a system with more degrees of freedom also supports experts’ sense of control, one of the main determinants of worker health, as well as their learning process, since boundary explorations are necessary experiments of learning. Moreover, supporting worker flexibility and adaptability will only become more critical in the future. With increased levels of automation, the primary role of an expert will shift from executing tasks to observing automation and dealing with unanticipated situations in which the automation breaks down. That means that the human in the system will deal increasingly with situations that the systems designers were unable to predict.

The need for designing flexible systems that support adaptive behaviour exposes a significant disadvantage of using task analyses because these generally assume a highly controlled environment and a sequence of prescribed actions with few degrees of freedom for the operator. In essence, these task analyses are like the set of directions provided to find a destination. In other words, they are event-dependent. However, just like with any general systems law, there is a tradeoff. The resilience of adaptive capacity should be balanced with optimality (Hoffman & Woods, 2011). As said before, using the set of directions requires less thinking than using the map. Thus, using a set of directions is optimal in nominal situations.

According to Vicente (1999), task analysis is proper as long as it is preceded by a focus on recurring patterns and generalisations about the macro-cognitive work domain. Work Domain Analysis (WDA) identifies the support required to deal with unanticipated events. Thus, Vicente suggested using both a work domain analysis and task analysis. However, there are two types of task analysis. The first one is instruction-based task analysis, which represents “a category of task analysis techniques that specifies what actions should *not* be performed if the goal is to be achieved” (Vicente, 1999, p. 4). The second type of task analysis is constraint-based, “a category of task analysis techniques that specify what actions should be performed (Vicente, 1999, p. 7). Vicente suggested using a constraint-based task analysis such as the control task analysis. A control task analysis is different from instruction-based task analyses as it identifies what needs to be done, independent of who needs to do it (p. 183). Constraint-based approaches only focus on the goal state and the constraints on action, giving workers more freedom to decide how to perform the tasks. Generally, constraint-based approaches are less dependent on the system being analysed than instruction-based approaches because constraint-based analyses are less detailed. Therefore, the more open a system is, the less suitable an instruction-based form of task analysis is.

In conclusion, rather than limiting future system users to instructions, we should use a constraint-based approach, consisting of a work domain analysis and constraint-based task analysis, to design safe and resilient systems. Currently, the HRL does not support defining the boundaries of safe work practices and the corresponding degrees of freedom. Instead, it is very task-focused (see Appendix A; see sub-question 8, 13, 42, 46, 57, 68, and 75), and it stresses the prediction of potential human-system issues (see sub-question 3, 10, and 21). Thus, the HRL could benefit from a constraint-based approach. The next question would be on what constraints system designers should base their design.

2.2.2 Environmental constraints

Following the philosophy of the CSE discipline, systems design should start by identifying environmental constraints (Vicente, 1999). 'Environmental constraints' might be a confusing phrase as it does not (only) stand for the constraints in the physical environment, nor does it concern the constraints laid upon a system by concerns about the environmental impact on nature. In this context, environmental constraints are "Work Demands associated with factors that are external to the Worker (e.g., physical or social reality)" (Vicente, 1999, p. 6). Vicente contrasts them with 'cognitive constraints': "Work Demands associated with Worker cognitive characteristics" (Vicente, 1999, p. 5). For an HF expert, it might seem counterintuitive to start designing a system focusing on environmental rather than cognitive constraints. However, it makes much sense when we consider that users' current mental models are based on current systems and, therefore, unsuitable for shaping the design of new and better systems. Moreover, mental models can be buggy because human rationality is bounded (Vicente, 1999; Woods & Hollnagel, 2006, p. 13). The current mental models might contain wrong assumptions about the system and the physical laws to which it is subjected. Furthermore, human thought and behaviour are not set in stone but are shaped by the constraints in the environment (Vicente, 1999).

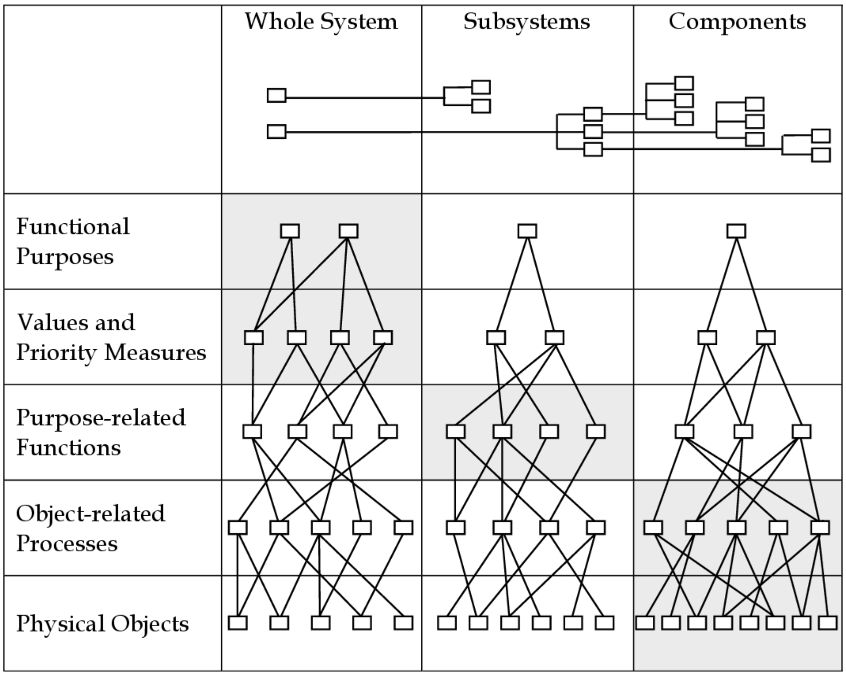
Thus, systems development should start with the identification of environmental constraints. The WDA is the perfect tool for this. The WDA is the first step of the Cognitive Work Analysis (CWA), a framework developed by Vicente to model complex socio-technical work systems. This thesis will not describe in detail what the CWA or WDA entail. For that, I can highly recommend reading Vicente's book 'Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work' for an introduction to CWA. Important to know, is that a decomposition hierarchy with part-whole links and an abstraction hierarchy with structural means-ends links can be used to conduct a WDA. The latter has several layers, depending on the constraints inherent to the work domain, essentially connecting the functional properties (i.e. purpose) and physical properties (i.e. material form). An example of a template for WDA can be found in Figure 2.

One sub-question in the HRL v1 is: "Are environmental constraints and impacts being analyzed?" However, it is unclear whether 'Environment' has the same connotation in the HRL as in the CWA. Moreover, even assuming that the sub-question refers to the same environmental constraints that the CWA talks about, the HRL places this sub-question after the sub-question related to cognitive constraints ("Have key user capabilities and limitations been identified?").

Once the WDA and control task analysis have taken place, it should be clear what external constraints shape the work domain and what tasks need to be accomplished. The next step would be to understand how these tasks could be accomplished or, in other words, what strategies could be applied. For this, Vicente proposes using a Strategies Analysis supported by an information flow map (Vicente, 1999, p. 215). Adding these elements to the HRL v1 framework would also be in line with the analysis of the projects, as it turned out that information flows played an essential role in several of the analysed projects.

Figure 2

Example of an Abstraction-Decomposition Space



Note. By Naikar and colleagues, 2005, p. 41

2.2.3 Social reality

Even though the CSE philosophy reasons 'from the outside in', it has everything to do with humans. Experts are embedded in a team and an organisation (and even in cultural history). These form the social reality, an essential part of the environment. Even legislators are part of the 'environment' as legislation constrains experts' behaviour (Rasmussen, 1997). This social aspect of the environment is also included in the CWA. After analysing the work domain, the tasks and the strategies for completing tasks, the CWA suggests conducting a Social Organisation and Cooperation Analysis (SOCA). Aside from several sub-questions about human-machine function allocations, the HRL does not explicitly support the consideration of cooperation between different system actors on the team and organisational levels. The integration of this element of the CWA into the HRL framework would thus be a potential improvement. However, even the SOCA does not provide much more than a function allocation. Johnson and colleagues (2014) created the Coactive Design method with which they go further than allocating tasks to different actors in the system. They argue that teammates engaging in a joint activity are interdependent and that this interdependence should be supported by enabling sufficient Observability, Predictability and Directability (OPD). Johnson and colleagues created the Coactive Design method to guide the designer in identifying the requirements for effective teamwork. The iterative process of the Coactive Design method starts with the identification of required capacities for each task and potential interdependence relationships and their corresponding OPD requirements. Subsequently follows the selection and implementation of mechanisms to support the OPD requirements for each relationship. Finally, there will be an evaluation process in which it is examined how the selected mechanism for each relationship affects the OPD of other interdependence relationships. For the identification phase, the Interdependence Analysis (IA) table can be used, which is an extension of traditional task analyses. Furthermore, team design patterns could be used to identify potential relationships, requirements and supporting mechanisms. Design patterns are abstract solutions to recurring challenges that allow designers to reuse concepts that worked before in similar situations (Van Diggelen & Johnson, 2019). More specifically, team design patterns capture the common invariant properties of a challenge and the relationships that are needed to meet the challenge. Johnson and colleagues (2014) applied the Coactive Design approach to a robot that can assist humans in recovering from disasters and found that their robot was more flexible and resilient than other designs. Thus, the Coactive Design method fits well with the CWA and the broader CSE philosophy of designing flexible and resilient systems. The approach of Coactive Design was specifically created to address the interdependence between people and robots but could be usefully applied to system actors (people and machines) in general. It would be a valuable addition to the traditional HF and performance evaluations already part of the HRL.

In summary, environmental characteristics are of primary importance in system development. However, it is important to emphasize that the CSE philosophy and the CWA process do not argue that we should only analyse environmental constraints instead of cognitive constraints. Instead, both are acknowledged to be crucial. A worker's behaviour is "constrained by the environment, but not completely determined by it" (Schraagen, 2018, p. 7). Theories specifying psychological mechanisms and representations 'inside of the head' - also called *process theories* - and theories specifying constraints that the environment imposes on humans - also called *product theories* – complement each other, "with product theories providing the constraints that process theories need to fulfil" (Schraagen, 2018, p. 7). Thus, the next step of the CWA concerns itself with cognitive constraints.

2.2.4 Skills, Rules, Knowledge

After analysing the work domain, the tasks, the strategies and the social organisation, the CWA continues with a Worker Competency Analysis (WCA). WCA considers basic capabilities and limitations and identifies the particular competencies workers should exhibit to function effectively in a specific job (Vicente, 1999). Vicente proposes to use the Skills, Rules, Knowledge (SRK) taxonomy, "a framework within which work domain requirements and existing knowledge of human cognition can be integrated" (Vicente, 1999, p. 275). Skill-based behaviour includes real-time sensory-motor acts that take place without conscious control and are directly coupled to *signals* in the environment (Rasmussen, 1983). Rule-based behaviour is goal-oriented and controlled by a rule that often has been derived from previous experience. This kind of behaviour takes place when familiar *signs* in the environment trigger a stored rule. In situations in which a human does not have stored rules, knowledge-based behaviour takes place. This type of behaviour is a very conscious effort based on a mental model that symbolises the environment in which alternative plans are (conceptually) tested. A system should ideally support all these three different ways of interpreting information in the environment. Although skill- and rule-based behaviour are generally what distinguishes experts from novices, the same skills and rules can lead to adverse effects in the face of abnormal circumstances. The design of a system needs to support the development of a correct mental model, supporting the sense-making process of experts during abnormalities. A practical example of such a design is the Ecological Interface Design (EID).

2.2.5 Ecological Interface Design

An EID displays to the user what is happening in the system rather than what should be done by the user. In this way, the EID helps users to be effective and adaptive. Vicente (1999, p. 295) presented three design principles for creating EIDs. First of all, information should be presented in the form of time-space *signals,* and users should be allowed to act directly on the interface itself. Secondly, there should be a consistent one-to-one mapping between the environmental constraints and the *signs* provided by the interface. Thirdly, the work domain should be represented in the form of an abstraction hierarchy that can serve as a faithful external model of the system that allows users to engage in knowledge-based problem-solving. The HRL could include some sub-questions linked to these principles in order to stimulate the design of EIDs.

2.2.6 Visual appeal

One of the major appeals of Vicente's successful models and methods is that they are highly visual, hence supporting the intuitive understanding of complex concepts (Waterson et al., 2017). This visual quality has contributed to the long-lasting success of Vicente's models and could be an inspiration for the HRL. First of all, the HRL should resemble as much as possible the recognizable thermometer-like shape that is used as a conceptual metaphor for the TRL. Furthermore, the visual representation of sub-questions could also help to highlight and clarify the inter-relations between the sub-questions

2.2.7 Lessons learned from the comparison between the HRL v1 and relevant literature

In Section 2.2, I have been exploring how a successful socio-technical system can be built from a formative point of view and translated this into requirements and sub-questions that can be incorporated in the HRL v2. Rather than describing how work should be done, a formative approach focuses on what may be done. These constraints form the boundaries of the space within which the expert is free to decide how to carry out the work. In other words, the system supports the expert's flexibility and adaptivity. In order to identify the behaviour-shaping constraints, several techniques can be applied, captured in the CWA. Here, it is vital to start with environmental constraints. The first step of the CWA is to conduct a WDA, followed by a control task analysis and strategies analysis. The environment includes all factors that are external to the expert, also the social organisation and cooperation. The CWA prescribes a SOCA to draw a picture of the social reality. In addition to this, the Coactive Design method can be applied to identify interdependent relationships and design mechanisms to support the requirements for these. After environmental constraints have been identified, the CWA proposes a Worker Competencies analysis to define cognitive constraints. The SRK taxonomy connects knowledge about the work domain requirements and human cognition. A system should be built to support all three types of behaviour that are described in the SRK. One way of doing this is by creating an EID. The first couple of HRLs could be adapted in order to be more in line with CSE's recommendation to start with environmental constraints. Sub-questions linked to the different steps of the CWA can be added to the framework, together with corresponding techniques and templates. However, the future users of the HRL can also be regarded as experts, and the HRL should thus be constraint-based rather than instruction-based. Continuing this analogy, the constraints of the work domain of developing and validating complex systems would be the sub-questions. The space within these boundaries, think of the decision to apply specific methods, forms the degrees of freedom that experts can explore based on their *internal* constraints such as available time, financial resources and expertise. So, in a way, the HRL will also be an EID. One of the main points of improvement for the HRL would be its visual shape. It should exhibit the constraints, or sub-questions, and the links between these as well as its connection with the more well-known TRL.

3. HRL redesign

This chapter describes the redesign of the HRL v2 framework.

3.1 Design suggestions

The HRL v2 is based on the analysis of current practices and relevant literature, as reported in Chapter 2. Table 3 was created to summarize all points of improvement, including recommendations.

Table 3

Overview points of improvement for the HRL framework, including the section on which each issue is based and suggestions for improvement

Issue	Reference	Suggestion
Technology (i.e. hardware and software) is often **already developed** when the operational or HF effort starts. Consequently, the technology is tested in a relatively high-validity environment, whereas previous HRLs have not been completed yet. It is not clear whether and how the HRL could be used in those situations.	ALICIA, ACROSS, HAL, HPAP, SME (refers to the interview with the Subject Matter Expert reported in Section 2.1.3)	HRL could create a strategy for this inspired by the approaches taken by other validation frameworks. HAL strategy: Those responsible for the project should make sure to address the objectives of the 'Definition phase'. (L. Save, personal communication, January 7, 2021) HPAP strategy: It is the responsibility of an HP specialist to verify that HP maturity mirrors the overall solution maturity and, in case there is a mismatch, he/she should refer to the HP argument structure for the previous phase to identify which HP activities should be carried out before the HP activities of the given project phase can be conducted (Morgan et al., 2018). Thus, the HRL could demand that at least the 'definition', or 'validation', phase be completed, regardless of the current maturity, to ensure that the correct system is *built*. In case the system already finds itself in the verification phase, in which the question is whether the *system is built in the right way*, it should be up to an HF specialist to refer to the HRL argument structure for the previous level(s) to identify which

		activities should be carried out before the activities of the given project level can be conducted.
Within aviation, there are already **other frameworks** to structure the validation process (e.g. E-OCVM). It is not clear how HRL relates to these.	E-OCVM, SME	The HPAP shows how parts of the outputs obtained during its four steps directly feed into the SESAR documentation. This is something that should also be done for the HRL. Additionally, the levels of the HRL scale should be mapped onto maturity levels of established frameworks (such as the CLM).
The HRL v1 framework does not consider the context of the system (incl. regulatory aspects) – the work domain. It is unclear whether **environmental** analysis pertains to the work domain or just the physical environment. Moreover, this environmental analysis takes place after identifying cognitive constraints (human capabilities and limitations).	SAFELAND, E-OCVM, SME, CSE (refers to Section 2.2)	The philosophy and methods of CWA could be used - starting with a work domain analysis at HRL 1. The identification of human capabilities and limitations can be moved to HRL 2, after the work domain and task analysis have been completed. The identification of potential human-system performance issues and risks will be moved to HRL 3 after task descriptions have been developed.
HRL prescribes a task analysis without preceding it with a **work domain analysis**. In this way, system development risks falling victim to the task-artefact cycle.	CSE	
Questions about risks and issues and human capabilities and limitations are asked very **early** in the process. These usually (in DBL projects) only get meaning when they are linked to a specific task.	SME	
The HRL v1 framework does not include **information flow/representation**.	ALICIA, ACROSS, SAFELAND, CSE	The CWA strategies analysis and organizational analysis, supported by an information flow map, should be added.

HRL uses the **HSI domains,** which are not common in European aviation research and do not capture all critical HF categories on the one hand. On the other hand, they include irrelevant domains (e.g. habitability).	NINA, HF Case	The HF pie can be used instead of the HSI domains to formulate sub-questions. In this categorisation, procedures receive a more prominent role. Additionally, procedures will also be mentioned at HRL 9, as after deployment, experience can cause operators or maintainers to develop their own procedures (Smit, 2020, p. 55).
It is unclear on which theoretical **model** the HRL sub-questions are based.	SME	
Procedures and working methods/operations do not receive much attention in the HRL (compared to how much attention they receive in DBL projects).	SME	
HRL focuses on Liveware-Hardware/Software components and misses the **Liveware-Liveware** components, such as team structure, communication, and other factors of the social reality such as the social organisation and its objectives, culture, and business model.	ALICIA, ACROSS, SAFELAND, HF Case, HAL, HPAP, CSE	This issue can be resolved by adding the Cognitive Work Analysis (including the social and organisational analysis) and using the HF pie categories (including Team & Communication).
Concerning **human-machine teamwork**, the HRL only considers function allocations.	CSE	Human-machine function allocations sub-questions (14, 29, 40) could be replaced with exit criteria based on the work by Johnson and colleagues (2014) about human-machine interdependencies.
HP is not specified in the HRL framework.	ACROSS, NINA, STRESS, SAFELAND, HF Case, HPAP	The HF Impact Wheel illustrates twelve critical HF impacts on human performance; *acceptance, cognitive processes, comfort, error, fatigue, job satisfaction, motivation, situational awareness, skill change, stress, trust and workload* (see Appendix I for a description). These could be used to identify cognitive constraints, metrics for successful human-system performance,

Acceptability of the concept for the user is missing in the HRL v1.	ALICIA, ACROSS	potential human-system issues and design principles.
Current HRL does not stimulate the development of an **EID**.	CSE	By applying the CWA, starting from work domain constraints and gradually working towards cognitive constraints, and adding Cognitive Control (SRK) to the impact wheel, the development of an EID can be indirectly stimulated. Additionally, one criterion can be added to the Environment category to follow up the CWA questions: Work domain constraints have been integrated into the interface.
The HRL does not include a **risk** assessment and is not adaptive to the potential risk.	NINA, HAL	It should be clarified that HRL is meant to support the development of safety-critical, highly complex and large-scale systems. For the developments of simpler, lower-risk, closed systems or technologies, other processes can be used. Existing risk assessments could be used complementary to the HRL.
HRL does not take into account 'validation **assumptions**'.	ACROSS, SAFELAND, E-OCVM	During the early stages of systems development, assumptions about technological affordances might be made. These need to be fed back into the technology development (TRL), and it needs to be ensured that the assumptions are met before deploying the systems (see Appendix A; sub-question 77).

HRL **mixes** activities and questions on programme and project level (see Appendix A; e.g. sub-question 5) and exercise level (e.g. sub-question 6). It also mixes goals (e.g. sub-questions 24) and methods (e.g. sub-question 16).	HF Case, E-OCVM, HPAP, SME	HRL could be divided into validation goals and project management goals or *content* (details) versus *process* factors (bigger picture). In the HAL, process factors are defined as activities and outputs associated with HF management within the context of different stages of the safety assessment. Process factors refer to how the assurance against people, equipment, and operational factors is ensured to follow an appropriate managerial and methodological process. (L. Save, personal communication, January 7, 2021) Therefore, process factors could capture the sub-questions related to planning and other managerial activities, such as determining the exercise requirements and criteria, reporting the activities and contacting stakeholders. Furthermore, specific validation activities, methods, techniques (the *means*) should be separated from the sub-questions (the *goals*).
HRL does not include a **toolbox** with methods/techniques.	E-OCVM, HAL, HPAP, SME	A toolbox could be added to the HRL framework. This toolbox will leave less room for the interpretation of individual experts without taking away too many degrees of freedom by leaving the *choice* of the activities and the exercise requirements up to the experts. The tools in the toolbox are thus merely suggestions. The additional advantage of this toolbox is that it emphasises the essence of the scale; that is, the increasing concreteness of the validation methods and representability of the validation environment as maturity increases.

In the HRL v1 framework, it is not very **visually** apparent that the sub-questions repeat themselves, constantly on a more advanced level.	SME, CSE	HRL should be made more visual and intuitive, preferably a design where interrelations become apparent more easily. The HF pie categories can be used to categorize the sub-questions so that it becomes apparent that the same topics are treated throughout the HRL, with increasing stringency as the system matures.
The HRL v1 framework contains **jargon** (e.g. 'mission') and **confusing terminology** (safety analyses 'completed' at HRL3).	SME	This is a minor point of improvement that can be addressed by adopting concepts familiar in the aviation industry (though, e.g. the HF pie) and by replacing words such as 'mission'. Instead of 'completed', the word 'conducted' will be used.
HRL follows the idea of an established concept (i.e. **TRL**) but does not visually resemble it.	HAL, CSE	The relation between the HRL and TRL could be strengthened further by using the conceptual metaphor that is recognizable for users (i.e. the shape of the thermometer).
The head-mounted display example offered by the HRL framework is a very straightforward technology rather than an extensive **complex system**.	SME	The HRL should be suitable for supporting the development and validation of complex systems. There are several characteristics of complex systems (Vicente, 1999, p. 14). They are often dynamic and can have long temporal delays. Moreover, they usually are composed of many highly interactive subsystems. Thus, it is vital to take a long-term perspective, not just a snapshot, when analysing, designing and evaluating a system. The following methods could be included in the HRL framework: Statistical Analysis of Recurrent Events, Relational Event Model (REM), and T-Pattern Analysis (TPA). The current HMD example is too simplified and should be replaced in the future with examples of the use of HRL for the validation of more open and complex systems.

Stakeholders do not seem to play an *active* role in systems development according to the HRL framework. In contrast, Deep Blue uses a more UCD where stakeholders are involved in the needs identification, scenario generation, concept development, and function allocation.	ACROSS, STRESS, SAFELAND, HF Case, E-OCVM	Identifying and actively involving stakeholder communities could be part of the 'managerial' subgoals in the HRL, and UCD methods could be part of the toolbox.

The integration of the CWA model to the HRL v2 is probably the most substantial change relative to the HRL v2. Other prominent changes are the replacement of the HSI domains by the HF categories and the introduction of the HF impacts. Also, the HRL v2 will look very different from the HRL v1 thanks to the transfer from an excel file to a visual. Finally, another significant change to the HRL v2 compared to the v1 framework is the toolbox that proposes HF methods experts could use at each level to develop and validate the system.

3.2 Toolbox

The tools in the toolbox are based on the methods and techniques mentioned in the various projects, the established frameworks, the interview and the CSE literature. An initial categorisation was made by matching the various tools to different maturity levels. Chapter 4 will explore whether the categorisation makes sense to the target users and whether additional categorisations are necessary.

In ALICIA, both quantitative and qualitative methods were proposed. Quantitative methods were interview, de-briefing, over-the-shoulder observation, Post-Simulation Questionnaire (PSQ), heuristic analysis, and scenario-based evaluation using the Supply Context Organisation Process Effect (SCOPE) framework. Qualitative methods were NASA Task Load Index (NASA-TLX), Situation Awareness Rating Technique (SART), and System Usability Scale (SUS). Bench testing and synthetic environments were used for the validation and are representative of HRL 4 and 5. Therefore, the methods will be added to HRL 4, which means they can be used from HRL 4 onwards.

In ACROSS, interviews, debriefings, cognitive walkthrough, observation, gaming techniques, Instantaneous Self Assessment of workload (ISA), NASA-TLX, SHAPE questionnaires (e.g. SASHA), heuristic evaluation, focus groups, hazard and operability study (HAZOP) and workshops were used to validate the operational concept in a relevant simulation with real users. The interviewed project manager had also recommended the HAZOP. ACROSS was an exploratory project, just like ALICIA, and it had, in fact, a very similar validation structure. However, the technologies were slightly more mature. The methods used in ACROSS will be added to the toolbox of HRL 5.

In NINA, validation took place in relevant ATM simulation facilities with real ATCs. This activity loosely corresponds to HRL 5, in which prototypes are evaluated in relevant simulations. Thus, the methods that were used in NINA can be added here. However, nothing will change as NINA only used methods that were also applied in ALICIA and ACROSS.

STRESS used physiological measures, task performance, and behavioural data in a simulation of en-route ATC activities on realistic training platforms (Bonelli et al., 2017). Both newly graduated and expert controllers participated in the simulations. The emphasis was on neurophysiological measurements: EEG, heart rate, eye blink, Galvanic Skin Response (GSR), eye tracking, and salivary cortisol. These methods can be added to HRL 5.

SAFELAND started with a WDA, ConTA and SOCA, partly executed in collaboration with SMEs in workshops (S. Bonelli, personal communication, May 7, 2021). These cognitive work analysis methods were also suggested by the literature and will be added to the first two HRLs. Workshops can be conducted from HRL 1 onwards. Furthermore, SAFELAND identified potential hazards using the Functional Hazard Assessment (FHA) method and the Safety Assessment Methodology (SAM). These methods will be added to the toolbox of HRL 3, at which the HRL v1 framework advises to start with the safety analysis.

The E-OCVM CBApp recommends the following techniques for V1: focus groups, HP issue analysis, task analysis, initial use cases, gaming techniques, review of low fidelity prototypes, and storytelling (Morgan et al., 2018). Since V1 corresponds to HRL 1 and 2 (see Figure H2), these methods can be added to these levels. More abstract methods such as focus groups will be added to HRL 1, whereas use-case analysis and storytelling can be used for HRL 2 when usage scenarios and task descriptions are developed. As discussed before, some of these activities take place later during development. Low-fidelity prototypes are central to HRL 4, and identifying potential issues starts at HRL 3 (in the HRL v2). Thus, HP issue analysis will be added to HRL 3. Gaming techniques will be added to HRL 4 to support the modelling, part-task testing, and trade studies of human systems design concepts. In V2, the operational concept is detailed enough for cognitive walkthroughs, using mock-ups, prototypes and simulators. The evaluation of prototypes in a simulated environment takes typically place at HRL 5. The cognitive walkthrough was already added to the toolbox of HRL 5 based on ACROSS, so this will remain unchanged. HF activities in V3 will be conducted in an even more realistic environment, corresponding to HRL 6, using operational trials and Real-Time Simulations.

In addition to proposing some methods, the E-OCVM referred to an online SESAR HP repository. In this repository, there are methods, techniques and tools that were not mentioned so far (https://ext.eurocontrol.int/ehp/). Firstly, two job analysis questionnaires can be used to identify the human abilities and skills needed to perform the job, namely the Fleishman Job Analysis Survey (FJAS) and the Position Analysis Questionnaire (PAQ). These can be used in HRL 2 to fulfil the exit criteria about human capabilities and limitations. In addition to the ISA and NASA-TLX, the repository contains various techniques for measuring mental workload. Air Traffic Workload Input Technique (ATWIT) is a technique to measure mental workload by presenting auditory and visual cues that prompt the controller to press a particular button. The Assessing the Impact on Mental Workload (AIM) questionnaire focuses on the effect of automation on mental workload. The Bedford Scale is a rating scale created to identify the spare mental capacity of an operator while he or she is completing a task. The Subjective Workload Assessment Technique (SWAT) assesses different kinds of workload: time load, mental effort load, and stress load. The Malvern Capacity Estimate (MACE) can be used to obtain a quick and direct measure of an operator's maximum mental capacity. Applications are typically in simulation environments and will thus, just like ISA, be added to HRL5. Furthermore, there are several techniques related to errors. For example, the GEMS is an error classification tool that is based on the SRK model. The Human Error Assessment and Reduction Technique (HEART) can be used to derive an overall metric of error probability. These methods could be used to identify potential issues and risks, starting from HRL 3 onwards. The Human Error Reduction in ATM (HERA) and the Technique for the Retrospective and predictive Analysis of Cognitive Error (TRACEr) are suitable for retrospective analysis of air traffic incidents in HRL 9. Additionally, Net-HARMS could be used to identify risks that emerge when several risks across different system levels interact with each other (Dallat et al., 2018). This system-based method fits well with the complex systems for which the HRL framework was designed. One of the suggestions for the HRL v2 is to add the factor Acceptance to the framework. The repository also includes several questionnaires for assessing operator satisfaction and acceptance, such as the Questionnaire for User Interaction Satisfaction (QUIS) and the Controller Acceptance Rating Scale (CARS). CARS is intended for simulations of the operational system before it is put into daily use. Thus, this scale can be used in HRL 6 with shadow mode trials and real-time simulations. The context of use for the QUIS has not been specified, but it might already be used in HRL 5 to evaluate prototypes. The repository also contains several methods for assessing situational awareness apart from the SART. The Situational Awareness Linked Indicators Adapted to Novel Tasks (SALIANT) consists of a behavioural checklist that can be used to assess situational awareness in teams. Dr Endsley developed the Situation Awareness Global Assessment Technique (SAGAT) for real-time simulations (SESAR Joint Undertaking, 2012). Similar to the SAGAT is the SALSA, which is also administered when the simulation is temporarily paused. In contrast to SAGAT, the Situation Present Assessment Method (SPAM) uses response latency as the primary dependent variable rather than the operator's memory of specific information in the environment. Situation Awareness Verification and Analysis Tool (SAVANT) is a combination of SAGAT and SPAM. These tools will be added to HRL 6 when real-time simulations of relatively high-fidelity take place. Finally, the repository contains a questionnaire that measures team effectiveness, namely the Team Effectiveness Audit Tool (TEAT). The questionnaire has proved to help provide a context for improving the effectiveness of working teams in organisations. The TEAT could be used at HRL 9, once an organisation and its teams deploy the system.

One of the suggestions in Table 3 was to include time-sensitive methods, such as the Statistical Analysis of Recurrent Events, Relational Event Model (REM), and T-Pattern Analysis (TPA). For example, the REM was initially developed in the social sciences by Butts (2008) to analyse the timing or order of communications. These models will be added to HRL 6 as this is the first level at which real-time simulations are conducted.

As suggested in Table 3, human-machine function allocations sub-questions (14, 29, 40) were replaced with exit criteria based on the paper by Johnson and colleagues (2014) about human-machine interdependencies. To support the accomplishment of these criteria, the Interdependence Analysis Table was also added to the toolbox. This tool can help describe interdependence relationships that should be supported in the implemented system and the associated Observability, Predictability, and Directability requirements. The Interdependence Analysis uses a traditional task analysis and knowledge of the team members' capabilities and the environment as inputs. Thus, this tool fits best with HRL 3, after the CWA has been completed.

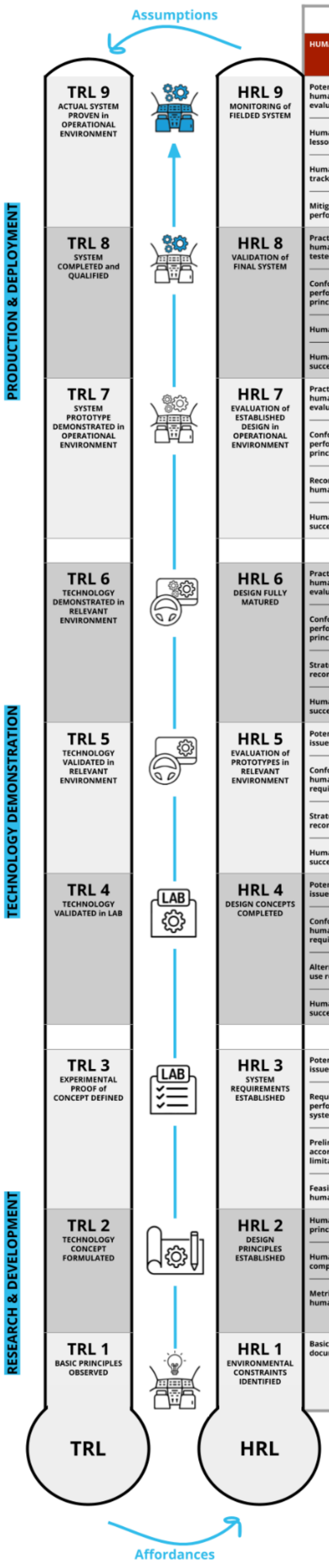
Based on SAFELAND, WDA, ConTA, and SOCA were already added to the toolbox. Based on the CSE literature, the other Cognitive Work Analysis methods were also added. For each of the CWA steps, an exit criterion was added to the HRL. In order to reach HRL 1, the functional structure of the work domain and the inputs, outputs and constraints of control tasks should be identified. For this, the WDA and ConTA can be used. For HRL 2, the process of how tasks can be performed, the organisation of actors into teams and the characteristics of users need to be defined. Here, the strategies analysis (including the information flow map), the SOCA and the worker competencies analysis can be used.

The suggestions in Table 3 were used to adapt the HRL v1. Some additional adaptations were made. First of all, the sub-questions about maintenance and sustainment (i.e., 20, 28, 52, 63 and 74; see Appendix A) were removed as it was unclear what they added to the HRL. Technical maintenance should be included in the TRL instead of the HRL. HRL sub-questions about a system to track and resolve human systems issues after fielding (i.e., 47, 58, 69) already cover preservation from an HF standpoint. Just like 'mission', Sustainment is again a concept that is probably industry-specific. The dictionary defines it as 'the support or maintenance of someone or something, especially military equipment or personnel' (Lexico.com, 2020). Thus, Sustainment might be relevant in a military context, but it was not mentioned in any of the aviation projects. In sub-questions 47, 58, and 69, 'system' will be replaced by 'practice' as it could otherwise be confused with *the system* under development. Moreover, the word 'system' might imply that it must be a technology, although it could also be a procedure. Throughout the HRL framework, 'user' will be replaced by ‘operator and maintainer’ to emphasize that both these users should be considered when developing a system. The sub-questions about the Safety & Health of the user were also removed. Instead, Safety & Health was added as a performance indicator to the impact wheel. Several other changes were made to the impact wheel. The impact factor 'Cognitive Processes’ was replaced by separate circles for Attention, Vigilance and level/type of Cognitive Control (SRK). Cognitive Processes is defined as ”how people think, make judgements and solve problems on the job, including information processing capability, memory, decision-making, vigilance and attention span and level and type of control (SRK)“ (Mellett & Nendick, 2007, p. 75). Cognitive Processes is not a clear performance indicator but rather an umbrella for several performance outcomes, so it was decided that it would be better to mention the specific outcomes instead. Attention, Vigilance and Cognitive Control have also been addressed in DBL projects, and Cognitive Control (SRK) is the basis of the EID. One HP impact that was mentioned in ACROSS and is also an essential aim of the EID is the ease of building a correct mental model. This factor has been added to the impact wheel. Skill Change was removed from the impact wheel as it fits better in the Training & Development category. Finally, icons were added to emphasize the evolution from HRL/TRL 1 to HRL/TRL 9, namely that of an increasingly detailed system (concept) in an increasingly more realistic environment.

The final design of the HRL v2 can be found in Figure 3.

Figure 3

The HRL v2, including the 'Impact Wheel' supplement



	HUMAN FACTORS CATEGORIES						TOOLBOX (aviation)
	HUMANS IN THE SYSTEM	WORKING ENVIRONMENT	ORGANISATION & STAFFING	TRAINING & DEVELOPMENT	PROCEDURES, ROLES & RESPONSIBILITIES	TEAMS & COMMUNICATION	
HRL 9	Potential upgrades to address human-system performance issues evaluated Human-system performance data and lessons learned documented Human-system performance issues are tracked and resolved Mitigations to improve human-system performance implemented			Training evaluated and modified if required	Procedures for operator and maintainer updated if required	Team effectiveness evaluated and improved if required	TRACEr, HERA, TEAT, NET-HARMS and all of the below
HRL 8	Practice to track and resolve human-system performance issues tested Conformance to human-system performance requirements and principles verified Human use issues satisfactorily resolved Human-system performance metrics successfully met	Effectiveness of strategies to address environmental constraints demonstrated	Effectiveness of strategies to accommodate manpower and personnel demonstrated	Effectiveness of strategies to accommodate training demonstrated	Procedures for operator and maintainer roles approved Operator and maintainer scenarios and tasks tested	Interdependence supported by system through careful attention to requirements for observability, predictability, and directability	All of the below
HRL 7	Practice to track and resolve human-system performance issues evaluated Conformance human-system performance requirements and principles verified Recommended strategies to support human use incorporated Human-system performance metrics successfully met	Effectiveness of strategies to address environmental constraints evaluated	Effectiveness of strategies to accommodate manpower and personnel evaluated	Effectiveness of strategies to accomodate training evaluated	Procedures for operator and maintainer roles tested Operator and maintainer scenarios and tasks tested	Mechanisms to support interdependence relationships and their observability, predictability, and directability requirements evaluated	All of the below
HRL 6	Practice to track and resolve human-system performance issues evaluated Conformance to human-system performance requirements and principles verified Strategies to support human use recommended Human-system performance metrics successfully met	Strategies to address environmental constraints updated	Strategies to accommodate manpower and personnel updated	Strategies to accommodate training updated	Procedures for operator and maintainer roles finalised Operator and maintainer scenarios and tasks tested	Mechanisms to support interdependence relationships and their observability, predictability, and directability requirements evaluated	Shadow Mode Trials, Real-Time Simulations, CARS, SALIANT, SAGAT, SALSA, SPAM, recurrent events analysis, REM, TPA, and all of the below
HRL 5	Potential human-system performance issues and risks identified Conformance of prototypes to human-system performance requirements and principles verified Strategies to support human use recommended Human-system performance metrics successfully met	Strategies to address environmental constraints updated	Strategies to accommodate manpower and personnel updated	Strategies to accommodate training updated	Procedures for operator and maintainer roles updated Task flow and sequencing optimised	Mechanisms to support interdependence relationships and their observability, predictability, and directability requirements evaluated	Cognitive walkthrough, ISA, ATWIT, AIM, Bedford Scale, SWAT, MACE, SHAPE questionnaires, focus groups, HAZOP, neurophysiological measures, QUIS, and all of the below
HRL 4	Potential human-system performance issues and risks identified Conformance of designs to human-system performance requirements and principles verified Alternative strategies to support human use recommended Human-system performance metrics successfully met	Strategies to address environmental constraints recommended	Strategies to accommodate manpower and personnel recommended	Strategies to accommodate training recommended	Procedures for operator and maintainer roles designed Task flow and sequencing optimised	For each interdependence relationship, mechanisms to support the observability, predictability, and directability requirements designed	Interview, debriefing, over-the-shoulder observation, PSQ, heuristic evaluation, SCOPE, NASA-TLX, SART, SUS, gaming, and all of the below
HRL 3	Potential human-system performance issues and risks identified Requirements to support human-system performance identified and included in system level requirements Preliminary design features to accommodate human capabilities, limitations, and needs investigated Feasibility of metrics for successful human-system performance evaluated	Work domain constraints integrated in the conceptual interface design Constraints of physical environment analysed	Manpower and personnel analyses conducted	Training analysis conducted	Usage scenarios, function and task descriptions updated	A set of interdependence relationships that should be supported in the implemented system - and the associated observability, predictability, and directability requirements - described	FHA, SAM, HP issues analysis, GEMS, HEART, Interdependence Analysis Table, and all of the below
HRL 2	Human-system performance design principles researched Human-system performance on comparable systems analysed Metrics and benchmark for successful human-system performance identified			Relevant human capabilities and limitations or characteristic identified	Usage scenarios and task description developed	Organisation of operators and maintainers into teams, communication and cooperation with each other, and authority relationships governing cooperation described	Strategies analysis (information flow map), SOCA, worker competencies analysis, use case analysis, storytelling, FJAS, PAQ, and all of the below
HRL 1	Basic human factors research documented and/or conducted	Functional structure of the work domain identified			Inputs, outputs and constraints of control tasks identified and process of how tasks can be done described		WDA, ConTA, workshop, focus group


Acceptance
Workload
Attention
Vigilance
Comfort
Trust
Cognitive control (SRK)
Impacts on Human in the System
Stress
Mental model completeness & correctness
Situational awareness
Error
Safety & health
Fatigue
Motivation
Satisfaction

4. HRL v2 evaluation

After adapting the HRL v1, the resulting HRL v2 was subjected to a user test in order to collect feedback from experts belonging to the target group. To my knowledge, the HRL has been evaluated by users only once before. It concerned a quantitative study in which the SUS was tailored to the HRL scale and distributed in several data-collection events to students and professionals. The study will be published in the special HRL issue of Ergonomics in Design (Handley, 2020). Rather than applying the same method to the HRL v2, this thesis aimed at qualitative data. In this way, I wanted to find out what obstacles still stand in the way of successful HRL adoption. This evaluation of HRL v2 also opened the floor to industries other than aviation in order to draw meaningful conclusions across industries. Furthermore, participants represented not only research-focused organisations but also industrial companies.

4.1 Participants

In total, five experts participated in the HRL evaluation. The participants had various educational backgrounds, most of them holding a PhD. They had a minimum of six years of experience with HF projects and 12 years on average. Two employees from Deep Blue were approached to participate in an HRL user test. One of them had a PhD in Cognitive Psychology and 12 years of experience in HF projects, mainly focusing on aviation. The other Deep Blue expert had a PhD in Human-Computer Interaction and over 13 years of experience with HF projects in aviation and manufacturing. Three other participants were recruited through the network of one of the thesis supervisors. The first participant worked at TNO, an independent research organisation in the Netherlands that focuses on applied science (https://www.tno.nl/en/). This expert graduated in Ergonomics and had almost 20 years of experience as an industrial designer with HF projects in the naval domain. Two other participants represented ProRail, a government organisation that takes care of maintenance and extensions of the Dutch national railway network infrastructure, rail capacity, and traffic control (https://www.prorail.nl/). One of them has a background in Informatics and has already worked at ProRail for 20 years, the past six years specifically on topics related to Innovation and Research & Development. This expert has been involved in the introduction of CSE methods to the work of rail traffic controllers. The other expert from ProRail obtained a PhD in HF and worked as an HF specialist at ProRail for five and a half years and five and a half years as a Guest Researcher in Human Factors & Gaming Simulations at the University of Delft.

4.2 Procedure

Participants were contacted via email and, if they were interested, a video meeting was scheduled. Before the interview took place, the participants received the document with the HRL and an informed consent form. They were told that there was no need to study the HRL framework but that they could look at it in advance of the interview if they wished to do so. The interviews lasted between 1 and 1,5 hours and were recorded with the permission of the participants. The participants were first asked to briefly introduce themselves and say something about their experience with HF projects. Subsequently, they were provided with a brief presentation of the HRL. After the introduction, participants were asked to think of a previous or ongoing project that would fit the target application of the HRL. More specifically, they were asked to come up with a project in which a complex and safety-critical socio-technical system was (partially) developed and validated. After defining a system, the participants were asked to apply the HRL framework to the example to determine the achieved HRL. They were asked to do this while sharing their screen and thinking out loud. In addition to the think-aloud exercise, participants were asked questions related to the HRL v2's utility and usability.

4.3 Material

The same scale used in the study presented by Handley (2020) was used to prepare questions for the interviews here. As mentioned before, Handley used the SUS, a 10-statement tool for measuring a system's usability (Usability.gov, 2020). For the evaluation of HRL v2, two of the statements were omitted. The other eight were adapted to fit better in the context of the evaluation and were formulated as open-ended questions. Apart from the adapted SUS questions, some additional questions were included. First of all, a question was added that asked participants to think of any industry-specific frameworks similar to the HRL they used. The analysis reported in Sections 2.1 and 2.2 pointed out that similar frameworks are used within European ATM research. So, this question was mainly meant for those participants working in fields other than aviation. Furthermore, a question was added about certain words used in the HRL. These words were identified by the supervisor of this thesis and myself as unclear or potentially confusing. The Coactive Design sub-questions were new, replacing the former human-machine function allocation sub-questions with exit criteria about human-machine interdependencies, and therefore addressed during the interviews to understand participants' interpretation of these. The toolbox was also a new addition compared to the HRL v1 framework shared by the HRL workgroup. Therefore, participants were asked about their opinion on the utility of this toolbox. They were also asked how they would like to use it (if at all) and whether some kind of classification would be helpful. All questions can be found in Table 4.

The questions were asked whenever they related to something the participants said while they were walking through the HRL to assess a maturity level to the example case they selected, such as in the following scenario:

Interviewee: So really the pilots and the people that will be using the technology. *functional structure of the work domain identified*[1]... Yeah, I mean, this goes with the concept probably, yes. So, yes. *input, output and constraints of control tasks identified and process of how tasks can be done or described* and this is procedures, roles and responsibilities. Yeah, I mean, at a high level, I would say yes. I mean, for me, this goes all in the, let's say, concept definition. Maybe what I would... *environmental constraints identified* uh yeah, not sure if this tells exactly what is here to me.

Interviewer: Yes, I'm actually curious; what do you see uh what is environmental? What does it mean here? According to you?

Interviewee: Yeah, exactly. I mean, for me, what it means is... let's say uh I can tell it in my words, that is what are the boundaries of my design space, basically, which includes physical space of work, but it includes also activities of people

In case the questions did come up during the think-aloud exercise, they were asked afterwards if there was still time left.

Table 4

Questions that were prepared for the evaluation interviews with HF experts. In brackets are the SUS statements on which the questions are based.

Introduction question	Please tell me something about yourself. What is your current role (your key tasks and responsibilities), how long have you been working in that role, and what is your previous experience?
	Can you think of an example of a project you were or are involved in that would fit the description of the HRL's target application (i.e. complex and safety-critical socio-technical systems)? Please describe shortly what the project was about.
Think aloud question	What HRL would you assign to the system that was developed and validated in the project you mentioned? Please think out loud while you determine the HRL so that I can follow your thinking process.
Usability questions	Would you like to use the HRL framework in future projects? Please explain why. If yes, when would you use it?

[1] Text between *asterisks* indicates that the interviewee was reading it out loud

	(SUS #1)
	According to you, who (else) would/could/should use this framework? How much time do you think most human factors experts (and other people you mentioned) need for learning to use this HRL framework? And how much time would they need to use it after they have learned to apply it? (SUS #7)
	According to you, what are the prerequisites for learning to use this framework? (i.e. the skills/knowledge you need to learn before you can get going with this HRL framework) (SUS #10)
	Do you think the HRL framework is easy to use or complex? Please explain why. In case it is perceived as rather complex: Is this complexity unnecessary? How could it be simplified? (SUS #2, #3, #8)
	How confident would you feel using the HRL framework? Please explain why. (SUS #9)
	How consistent or inconsistent do you think this HRL framework is? Both in itself as well as compared to current practices. (SUS #6)
Additional questions	Which process or frameworks do you normally use to ensure that Human Factors are systematically identified and considered in operational and technical developments (if any)? Do you know of a similar framework specific to the industry in which you are working? In what ways is the HRL better/worse compared to these?
	According to you, what is meant by: - Environment - Human-machine interdependencies - Human use - Strategies
	Would you like to use the toolbox? If yes, how would you like to use the toolbox? What kind of classification would be helpful?
Closing questions	Is there anything you would change about the framework? Is there something you would improve or add?
	Do you have any other comments about the framework that we have not discussed yet?

4.4 Results of the HRL v2 evaluation

The original transcripts of the interviews can be found in Appendix J. The data was analysed using a Thematic Analysis, as described by Braun and Clarke (2006). Braun and Clarke created a structured process for analysing qualitative data, consisting of six phases. The first phase is about familiarizing oneself with the collected data, which is best done by transcribing the data. In phase 2, initial codes are produced based on the data. A code identifies a piece of raw data that the analyst deems interesting. Subsequently, in phase 3, the analyst sorts the codes into potential (sub)themes. These can be visualised in an initial thematic map. In phase 4, the themes are refined. Some of the potential themes might not contain enough data or contain data that is too diverse. Several themes might form one or might need to be separated into different themes. In this phase, it is vital to ensure that the codes fit the themes and that no relevant data was missed in earlier coding stages. For this, the analyst might go through the complete dataset again. During the fifth phase, themes are defined. To do this, the analyst focuses on the data extracts of each theme and organises them into a coherent account. Once it has been determined what aspect of the data each theme captures, the analyst moves to phase 6. During this last phase, the analysis of the themes is written down in a concise, coherent, and compelling way.

4.4.1 Vignettes

Before focusing on the themes, this section describes the cases that the interviewees proposed to use as examples for applying the HRL.

Multiple remote tower concept. The first concept that was used as an example is part of a SESAR SJU project. It concerns a new setting in which two or more *modules*, representing air traffic controller positions, in a remote tower environment serve up to three airports. The concept includes a new role, *the supervisor*, who has to allocate the airport to the different modules according to specific criteria such as controller workload. In order to manage this task, the supervisor is provided with a specific tool. The project is aiming at bringing the concept to V3 of the E-OCVM CLM V-phases. The HF expert from Deep Blue involved in this project confirmed that each of the exit criteria belonging to HRL 1 and 2 was fulfilled. It was also confirmed that the design concept had been completed; in other words, HRL 3 had been achieved. The project is currently evaluating a prototype in a simulated environment but at quite a high level of abstraction. The design has not yet been fully matured. The expert assigned HRL 5 as the HRL v1. However, the exit criteria of HRL 5 were fulfilled when it comes to the Human in the System category, whereas the exit criteria belonging to the other categories, such as Training & Development and Organisation & Staffing, were not yet completed, as they are typically investigated starting from V3. Therefore, the multiple remote tower concept is currently being evaluated at HRL 5, as a prototype in a simulated environment but still needs to undergo another iteration, focusing on the lagging HF categories, before actually achieving HRL 5.

Head-up display. The second example, brought in by the other HF expert at Deep Blue, concerns one of the solutions developed within Alicia, a project focusing on enabling flight operations in all weather conditions that was also referred to in Section 2.1. Fog, snow and rain can impede the full visibility in the cockpit, which can be problematic when the aircraft needs to be landed. The head-up display projects additional information that is impossible to see due to the bad weather onto the reality and thereby supports safe landing in bad weather conditions. The HF expert confirmed that the exit criteria of HRL 1, 2, 3, 4 and 5 were covered. Reading the exit criteria belonging to HRL 6, the HF expert concluded that the system did not yet achieve HRL 6. Especially the part regarding the Organisation & Staffing, Training & Development and Teams & communication were not that much considered yet.

Control room layout. The HF expert from TNO has much experience designing control rooms for naval ships and chose one of these projects as a case for applying the HRL. They started the project by identifying the environmental constraints, so HRL 1 had been completed. Similarly, the exit criteria for HRL 2 were fulfilled. The HF expert estimated that the concept would probably be in HRL 4, 5 or 6, that is, the technology demonstration phase, since they built demonstrators of the space in 3D. They recently handed the concept over to the industrial partner who is going to develop it further. From this point onwards, that is HRL

6, their involvement decreases, and they simply support the industry and act as a reviewer to safeguard the HF knowledge that was developed.

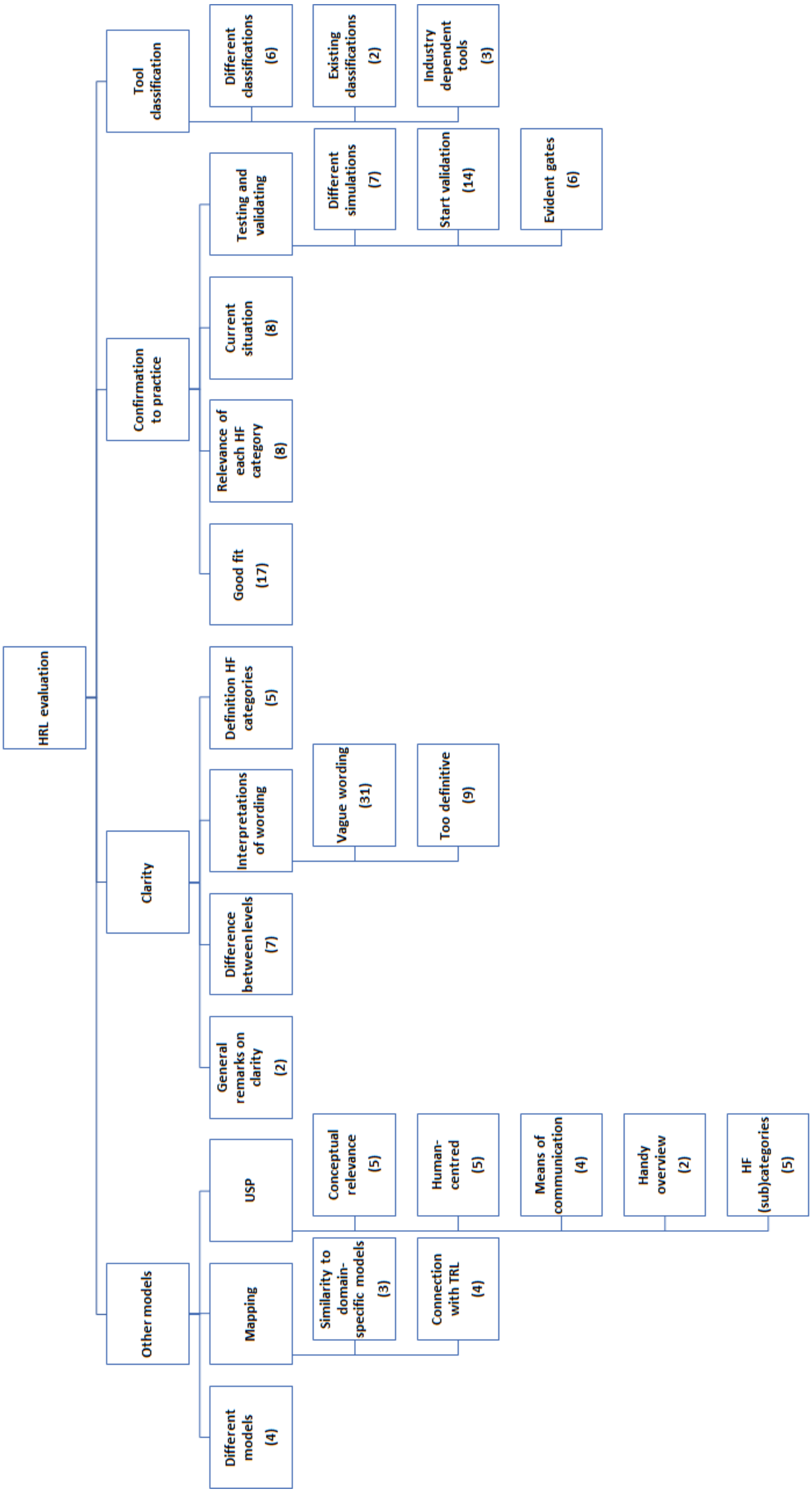
ERTMS safety system. The two HF experts from ProRail proposed to take ERTMS as an example to apply the HRL to. ERTMS is a safety system that ensures the train driver does not have to look out of the window anymore to see the signals but instead receives all this information inside the cabin. As a consequence of this 'smarter train', the train dispatcher can be equipped with more information. One idea is to provide the train dispatcher with information about the status of the train. This system change has a potential influence on the work of the train dispatcher, and the first step of the project was to compare the current and future scenarios. They were especially interested in the impact on workload and (shared) situational awareness. According to the HF experts, this fundamental research brings the concept to HRL 1. Additionally, they prepared collaboration principles, or 'user processes' as they call them, and built a first user interface. This new user interface concept has been tested in an (abstract) simulation environment. The HF experts were unsure which level of maturity would match the current test environment best; HRL 3, 4, 5 and 6 were all mentioned.

4.4.2 Utility and usability

As suggested by Braun and Clarke (2006)'s process, the interviews were first transcribed. After transcription, 157 codes were assigned to the raw data. These codes were subsequently grouped into 26 subthemes, which were then combined to form four main themes: Other Models, Clarity, Confirmation to Practice, and Tool Classification. The thematic map with the themes and sub-themes can be found in Figure 4.

Figure 4

Thematic map themes and subthemes. The number between brackets stands for the number of codes captured by the particular (sub)theme.



Other Models. During the interview, the HF experts regularly mentioned other models they use to structure and support the system's development and validation process. For example, the E-OCVM and the SESAR HPAP were mentioned repeatedly during interviews with the HF experts in aviation. During the interview with the HF experts within the railway industry, other concepts like the HF Fase and the Software Engineering V Model were mentioned. The reason for the number of various models out there boils down to the following:

> You often work with different suppliers. They also bring their own model every time. For example, we are now working on a project with NLR [Dutch Aerospace Center]. And they also have their own model. … You just see that there are quite a few different models. You have different suppliers, but you also have colleagues who then bring other, different ideas with them. I think there is probably no ultimate model. That is also the point.[2]

The same HF expert noted:

> Within innovation, we have all kinds of steps for how you progress through innovation. The V model from systems engineering is often referred to, and yet another model about concept development has also been developed by NASA. ... So you actually want this [HRL] model, and I think also it is true that it could be matched well, but you also want to be able to match it with other models.[2]

Thus, it looks like the HRL is entering a space that is already quite crowded and to be useful, it should be clear how the HRL maps onto currently used models. Some similarities were already identified by the HF experts themselves during the interviews. For example, one of the HF experts in aviation noticed that the exit criteria of the HRL are similar to the arguments in the HPAP. The HF expert from TNO mentioned that HRL 7, 8 and 9 are

[2] This quote has been translated from Dutch to English.

comparable to the harbour acceptance test, the ship acceptance test and the factory acceptance test that the Navy uses. Currently, the link between the TRL and the HRL is the only one that has been presented. Regarding this, interviewed HF experts noted that “it's very powerful that it's set against the TRL levels because they are generally accepted”[1] and “it's really like showing that there is a parallel path, that the two things are in connection.” Also, the assumptions and affordances connecting the two scales were said to make sense. Moreover, when discussing whether the HRL should be industry-specific or generic, one of the HF experts said:

> Well, I think it's just a general model, across all industries. I think that this is also the strength of the TRL. I would really keep it at the same level as the TRL because that is a worldwide known model, and I think that it is very powerful, that we can put this against it. So it may remain general.[2]

So the clear connection of the HRL to the TRL is seen as a strength and should be maintained as much as possible.

Apart from identifying overlap and similarities between the HRL and other models, it was also stressed that the HRL should have a clear, unique selling point. For example, one interviewee said:

> To be used in SESAR, for example, it should propose something more, that makes some difference, provides some added value compared to the others. This can be true in all fields but in SESAR particularly mainly because they have this consolidated use of the E-OCVM methodology.

According to the same expert, the HRL’s focus on the human in the system is its added value compared to the TRL and E-OCVM. This human-centeredness could be highlighted even more, she thinks. For example, one of the phases is currently called ‘technology demonstration’, which does not do justice to both TRL and HRL equally. Both the railway and the naval industry experts agreed that the HRL could be a means for communicating with engineers and suppliers. The interviewees underlined the conceptual and methodological relevance of the HRL and its importance as a counterpart of the technology-focused movement. As one of them expressed it:

> We often run into the fact that technology is considered more important than the HF aspects. And that's a bit of an age-old struggle I think. And then this could help in that, to also be able to give it a formal status. So I see it mainly in that, as a means of communication.[2]

The HRL could function as a handy checklist, for example, for a project manager, and the overview of the tools can be practical for HF experts. Similarly, the HF categories and Human in the System subcategories were mentioned as a strength of the HRL. For example, one of the interviewees mentioned the following about the Human in the System impacts overview: “I think it's a very nice schedule because I actually know all the topics, but it's nice to see them together for once.”[2]

Clarity. This theme captures all comments related to the (un)clarity of the HRL, a significant component of its usability. Generally, it was noted that the HRL has a clear philosophy and that the different levels contain well-detailed criteria and clear expectations. However, when assigning HRLs to the cases, the experts struggled with the difference between the levels. The difficulty seemed to be concentrated on levels 4 to 6. For at least two of the interviewees, this was partly caused by redundancy. For example, one expert said:

> Well, what I notice is that there is a lot of overlap in the text of the rows, so I think that makes it difficult to really distinguish those different levels. And yes, if there is so much redundant text in it, that also makes the model less readable. So I think that could be improved.[2]

Nevertheless, the experts understood that each level is a further refinement of the previous level, validating the concept in a different environment. As part of a solution to this ambiguity, one interviewee suggested explaining the differences in an accompanying document. Also, the wording needs to be selected carefully and explained if necessary.

Seven words and phrases were particularly confusing or open to multiple interpretations. Firstly, one of the exit criteria was said to be unclear, namely: ‘Alternative strategies to support human use recommended’. It needs to be specified what kind of strategies are expected at this level. Additionally, the word ‘recommended’ could be interpreted in different ways. After reading the particular exit criteria, one HF expert said: “yes, not only recommended but identified, tested, integrated into results and then define the requirement for that, for the users, and recommendation yes.” Another expert also took a guess at defining this exit criterion:

> Here, for me, it's something that is somehow generated by the fact of doing the validation in a lab. So something that I'm not searching for but it's something that pops up in a way. Because I’m putting the technologies, I'm somehow not stressing but I'm validating the proof of concept of my technology, let's say the prototype of my technology. ... I'm really thinking about something that you didn't consider before, not because you were not requested to consider it, because the consideration of these aspects is enabled by the fact of taking the technologists to the validation in the lab. … But again, I mean, this is probably a super personal interpretation of what they meant here by ‘alternative strategies to support human use recommend’. … I'm probably taking the wrong direction.

It is unclear whether this description matches the original meaning that the HRL workgroup intended to get across. The HF expert should not have to guess the meaning of the exit criteria. Unfortunately, this was not the only exit criterion that led to confusion. Likewise, the criteria saying ‘analysis conducted’ were experienced as vague by the interviewees. One interviewee compared them to the Coactive Design sub-questions and said:

> I think what I remark between this and that, for example, is that here you tell me three keywords like observability predictability and directability that, I mean, if I knew that, I would exactly know what you expect as an output here, which is good. Whereas here, when I say analysis conducted, I mean, what does it mean? What shall I look for in this analysis? Here, you told me what shall I look into at this level. So you tell me at level three you need, from teams and communication, you need to look at how relationships are observable and interdependencies are observable, predictable and directable. Or, I mean, you have to be able to define the requirements for that. Whereas here, for example, you don't tell me what analysis means.

The Coactive Design exit criteria that replaced the human-function allocation sub-questions thus received positive feedback. The HF expert also said she found it super interesting earlier on in the interview. Returning to the unclarities, ‘basic HF research’ is another phrase that elicited slightly different interpretations. Whereas one expert interpreted it as fundamental research in which you have a look at the current situation, that is, the current limitations such as workload, another expert focused on human needs:

> In this stage [HRL 1], I would say that basic HF research is actually the definition of the need from the human factor perspective. So the need of the pilots and definition of the needs of the different actors in terms of users. … So it can also be like, I speak to five pilots, two questions, three questions: What do you do when you are in bad weather conditions? What would you need to help you? And tell me a story of a near-miss or, you know, a situation where you felt uncomfortable. So I would really ask five pilots these three questions and then derive, define a sort of list of basic needs that comes out of this type of documented research.

Not only the exit criteria but also the HRL titles themselves created confusion. Environmental constraints, the focal point of HRL 1, was a much-discussed concept. It was said to be misleading as it could be confused with 'environment' in a sustainable sense and because it is not clear whether it offers full coverage for the exit criteria in HRL 1 or merely stands for the physical environment. When one of the interviewees was asked what environmental constraints stand for, she replied the following:

> I mean, for me, what it means is, let's say it in my words, that is what are the boundaries of my design space, basically, which includes physical space of work, but it also includes activities of people. So it's really, I mean, for me environmental, I don't know if it's, when I think to environmental here at this high, at this level number one, I think, really of the context and the work of people within the context, which is basically what I tried to say, at the first level, I imagine, you need really to make an appraisal of the needs that are there. ... So the definition of the situation, environmental is the situation, so not just what is out, let's say, in the environment, but also what is in the persons and in the work that people have to do.

Another expert described it as the sum of HF assumptions, operational demands - the situations and work domain - and the technical principles. Thus, the word 'environment' is unclear and, depending on the chosen definition, limited in the eyes of HF experts. One of the interviewed experts mentioned that the exit criteria of HRL 1 are all part of the 'concept definition'. The next level, HRL 2, also led to several interpretations. One HF expert described it as follows:

> It's about the basic principles, about how you want to function or work with the system, I think. Here too, we immediately make a, I have to think about it, what do we call that uh interaction pattern. In which you actually look at what generic elements you can use in that interaction that people are already familiar with. ... What we

> always take as a starting point with regard to control room layout is collaboration, are the collaborative relationships, so the people, knowing which people have to work in that space and it is very much about what the collaborative relationships between those people within that team are and that must be supported. ... 'Interaction design patterns' was, by the way, the word I was looking for, which we use for those design principles.[2]

One of the HF experts from ProRail also had her own idea of design principles. They established interface principles but also a new way of working which goes further than principles. A few levels up, the phrase ‘relevant environment’ was questioned two times. The interviewees were unsure what it meant precisely apart from being more advanced than a prototype and less than the operational environment. The last level which caused confusion was HRL 6, that is, ‘design fully matured’. This level was discussed mainly with the expert from TNO as the selected case concerns a relatively mature concept. Although the HF expert was unsure when a design can be called fully matured, she noted that HRL 6 symbolises the moment when the design, at a demarcated level of detail, is ready to be handed over to the industry. One of the experts from ProRail said that HRL 6 could be the level where it is decided to implement the design idea, which is when options are being excluded. The other expert from ProRail said:

> What I understand is that after HRL 6, you know this is going to work from a technical point of view and a human point of view, and then you make it for real and then you can throw away your prototype and then you go to 7.[2]

Related to the ambiguity surrounding the meaning of ‘design fully matured’ is the experts’ feeling that the choice of words, in general, is too definitive. This was not only said about HRL 6 but also with regards to HRL 4, ‘design concepts completed’, and one of its exit criteria ‘strategies to support human use recommended’. As a result, experts were reluctant to say that an HRL was completed. As becomes clear from the following data excerpt, these are issues of word choice rather than misunderstandings of the general HRL philosophy:

> Interviewee: 'Strategies to support human use recommended' uh yes, I find it hard to say this is finished actually.
>
> Interviewer: Yes? And why?
>
> Interviewee: Because the product is not fully developed yet. There are still all kinds of choices that you can make at a finer level of detail that also affect this. So uh.
>
> Interviewer: But as you said yourself, it also comes back at the next level.
>
> Interviewee: Yes, exactly.
>
> Interviewer: So it's more about whether you completed it at the level of a prototype in a simulated environment.
>
> Interviewee: At that level, you have completed it. Yes. Yes.[2]

It was emphasized that there is a continuous feedback loop to include things that they run into later on:

> As soon as you start the next phase, you will always come across things that you still need to refine. Otherwise, you would not work efficiently, if you have to complete everything 100% before you can proceed to the next phase. … And then it does indeed go with the specifications to the industry. They will then really build a system prototype. But there are always things that need to be changed and uh, so you occasionally go back to HRL 6 to make some adjustments of things that have not yet been fully worked out in detail, they have to be worked out in HRL 7, and you will evaluate that again.[2]

Thus, the way the levels are currently formulated seems to signal that no changes can be made anymore at later levels.

In addition to the unclarity of the difference between the levels and the words mentioned earlier, some experts commented on the HF categories. More specifically, the meaning of the categories Organisation & Staffing and Training & Development was not evident. Some solutions for this already came up during the interviews. One of the interviewees mentioned that in SESAR, leading questions are used to define each category. Another interviewee suggested making a similar breakdown for the other categories as for the Human in the System, referring to the 'Impact Wheel' supplement (see Figure 3).

Confirmation to practice. In general, the philosophy of the HRL is easily understood by HF experts as it represents a very well-known process. This was literally stated by one of the experts from Deep Blue:

> Interviewee: To me, the philosophy is quite easy to understand, because it's actually straightforward, because it's the work that we do for each time, it's a very well-known process.
>
> Interviewer: Okay. It fits well with the practice that you're already doing?
>
> Interviewee: Yeah.
>
> Interviewer: Okay.

It was also reflected in comments specifically expressing similarities between the HRL and the practices as-is. This was especially true for TNO and Deep Blue, for which 6 and 9 codes were identified, respectively. To give some examples, identifying environmental constraints is indeed something with which the expert from TNO and her team always start. Similarly, the analysis of comparable systems, one of the exit criteria of HRL 2, is foundational for them. One of the HF experts from Deep Blue stressed that starting from HRL 2 onwards, it is vital to broaden the temporal scope of the usage scenarios. Although the time-sensitive tools REM and TPA are placed only at HRL 6 and onwards, adding these tools had a similar goal: to broaden the analysis's temporal scope. Broadening the scope of usage scenarios is thus supported by current practices and could start even at a lower level. At HRL 3, the HF expert indeed expects to identify human system performance issues and risks or, as she called it, ‘bad cases’, which she defined as follows:

> I was thinking that at one point I [HRL 3] would probably, early in the design process, I will probably think also of, let's say, bad cases. So scenarios that I don't want that could happen or scenarios that I don't want to happen. In order to understand what could go wrong basically.

Additionally, the expert from TNO confirmed that the identifying issues and risks and the conformance of prototypes to human systems performance requirements and principles is indeed something that they are working on throughout the process and thus recur at several levels. The same expert also mentioned that after the hand-over to the industry, they get another receive moment which corresponds to HRL 7. In line with HRL 9, they monitor the system after it has been fielded to check how satisfied users are and to generate inputs for new developments. Furthermore, she noted that procedures, roles and responsibilities are a challenge and can indeed still change in HRL 9, after the system has been fielded.

It seems that the HRL and the current practice are pretty aligned. However, when it comes to the HF categories, some incongruencies came up during the interviews. In general, the HF categories do not all get the same amount of attention in the projects in which the HF experts were involved. More specifically, the Human in the System, Working Environment and Procedures, Roles & Responsibilities are often the focus of the HF effort. In contrast, the others - Organisation & Staffing, Training & Development, Teams & Communication - are often lagging. For example, the HF experts from ProRail said the following:

> Interviewee 1: So there is awareness that training has to be looked at and prepared for, but apart from that, it's not really uh that's not really discussed during development. It

is like, yes, we have to do that, and it is written on the roadmap that it should be developed, but nothing else is being done with training in those phases, under 7.

Interviewee 2: No, the moment you are really sure that you are going to do it, we often look at how we are going to implement it, how much time is needed for training, who is going to do it, and so on.[2]

This difference between the HRL and current practices was pointed out during three of the four interviews, only the HF expert from TNO did not mention it. To the question of whether it would be better to attend all HF categories throughout the development process as the HRL suggests, the reactions were divided. The experts from ProRail think that the relevance of the HF categories is project-dependent. An expert from Deep Blue said:

> For me, what will be really useful is to be able to demonstrate, actually, that having this combined approach, this integrated approach that makes all the categories advance together, is actually a good way of doing things because it's helpful to the development of technology. So it's not just because we ask to do it, but because it brings an added value, and benefits in a way the development of technology. ... That it's not just, you know, a box to tick, but it's something that actually gives me, I mean, the strategies to accommodate training are updated as an exit criterion here, it's super crucial for me to go to the next level because it produced benefits for me in the whole system design.

So the attention to all HF categories throughout the process could be an added value of the HRL compared to how systems development is currently done. However, it would need to be proven that this approach is beneficial.

Another potential conflict was identified with regards to the consideration of the current situation. Here, the current situation refers to the reference or baseline situation, that is, the cockpit, ATC tower, or control room as it is right now, without the envisioned system.

One of the experts from ProRail noted that before looking at the future, there should be an HRL 0 in which they conduct fundamental research of the current situation. The goal of this is to know how things are going at the moment, what works and what does not work. For example, at ProRail, they started the project with an impact analysis which entailed conducting a hierarchical task analysis of the current situation and the situation with the new system. Her colleagues agreed and added: "Yes, because usually, our projects are such that we don't make something completely new, but in a certain work domain, we introduce a new technology." Similarly, the expert from TNO delineated the importance of using the knowledge of the current process as an input for building the new concept. In both interviews, the idea emerged that, in a way, the analysis of the system after deployment, at HRL 9, brings the system back to HRL 0/1 and, specifically, to the analysis of human performance on legacy or comparable systems. In other words, systems development and validation can be seen as a circular process.

Testing and validating. This theme could be seen as a large sub-theme of confirmation to practice. It talks about different types of simulations used by the HF experts and when to start testing the design. This theme attempts to capture and explain the initial impression that experts use simulations at lower maturity levels. For example, the HF experts from the ERTMS case conducted a simulation and said the following:

> Because it already says 'validated in lab' then uh I always think to validate is a big word and the phase we are in is uh first just exploration in which, in my opinion, you are in level 3 for that reason, it is the first time that we have tested it, but with components of level 4 already but, yes, okay you can also argue it is a first simulation or a first investigation as part of level 4 and there will be several iterations. ... Yes, okay. So if you see it like this, you can say eventually there will be a lot of iterations in 4, I think.[2]

So it seems that the simulation was used to conduct the first exploration. When asked why they would not assign HRL 5 to the concept, given that they tested a prototype in a simulated environment, they answered:

> That is also something that plays a role within the company. We have different types of simulation systems, and that is actually related to the representativeness of the simulation in which you simulate. So at 5, you would, for example, uh look at what it says uh you would want an environment in which, for example, the system is a bit more representative for the operator, also in particular because you have neuropsychological measures in it for example, or I can imagine that something like workload that this also becomes more important and here it says 'potential human-system performance issues and risks identified' yes you want actually to have a simulation environment which comes closer to the real system and in which, for example, certain parts of your system, and in this case you are talking, for example, about the security system that is underneath, or the functionality, that is represented are a bit better in that simulation environment. Because you can also say, look, in 4, they could also just even be more advanced mock-ups, so to speak.[2]

Later on, one of the experts explained what type of simulations they use:

> We often think of 3 types of simulations, and I think there are also more resources that we should think about, but the last step is actually a training environment, and there you are actually almost with 1-on-1 systems, you can say. That's about 7, I think. You can use it for that. And if you look back then, before you build such a system, uh before you build your real operational system, you want a first, you also want to have some kind of proof of concept, and that is mainly in 6 uh because that is the phase you have to go through before you go to 7, before you start developing it, in an operational development uh so then you have that first in 6, and then you have an environment for

> that, that you use to quickly prototype and to formulate requirements that give feedback at level 6. And I think that's the environment we're in right now. That is a simulation environment that you can simply change quickly. So you do get a better feeling or understanding as an operator, but they are no longer, as it were, cardboard mockups.[2]

So, there are at least three different kinds of simulators they use, one that includes mock-ups, one that is more representative but still allows for quick prototyping and, thirdly, a training simulation that is practically a copy of the actual system. They related the first environment to HRL 4 and the last one to HRL 7. The second type of environment has been linked to both HRL 5 and 6. In this fragment, it can also be seen that it is not apparent what type of simulation exercises belong to what HRL. Whereas the current simulation was first assigned an HRL 4 or even 3, HRL 6 was mentioned later during the interview. From the interview with the other expert from TNO, it became apparent that demonstrators of the product, or at least the general functioning, can be built quite early in the process to align expectations. So the simulation is not only used to get feedback from users but also clients.

So far, it seems that there are different types of simulation and, although they might serve a broader purpose than validating the design with users, they seem to be pretty in line with the HRL. More discussion arose once it was argued when the validation should start and how many iterations there should be. This topic was intensely explored by the expert from Deep Blue that is very familiar with the E-OCVM and the HPAP. She noted that it is not clear how at HRL 3, for example, human capabilities and limitations and human performance metrics are investigated. She believes that potential issues and risks and the 'goodness' of the identified principles and requirements should be tested, based on a proof of concept and involving the users. She used the E-OCVM to clarify her point:

Interviewee: I'm more familiar, I have to say, with the E-OCVM scale. But I know that the TRL is quite overlapping for some parts, and this is supposed to be for E-OCVM the core part. V1, 2 and 3 are considered the core part, the more relevant aspects. In that phase, in E-OCVM, you have several iterations at different levels. You know, it's a continuous iteration made with the concept which is actually already, in the worst case, it's at V1. But at V1, you already start with something that can be validated, with qualitative results, with workshop, focus group, and so on, because it's not mature enough, of course, but then at V2 and V3, you start with interacting with the concept to validate and test if it's feasible, if it's workable uh if it's supportive for the controller and so on. It's something that, in this description at least, it seems to not be reported. So my doubt is that it's not reported here, but maybe it's hidden below.

Interviewer: So it's not clear whether it's iterative or?

Interviewee: Yeah, it's not clear if it's iterative, it's not clear how they measure that it's mentioned here in the description of human in the system for example, how these metrics, these capabilities, needs, limitations are investigated or, maybe it's my fault, at that phase, there is no need according to the HRL to investigate in concrete, it's just a description.

Thus, it is not evident that the HRL is iterative and that (qualitative) validation starts as soon as HRL 1. Compared to the E-OCVM, the HRL seems to start later with validating the concept and involving users, also suggesting fewer iterations. The HF expert indicates that she would need another testing and feedback moment at least at HRL 3 and 4 before evaluating the prototype in a simulated environment in HRL 5. She suggests the following:

Interviewee: Introduce some gates that clearly identify the potential end of one stage, so one cluster, the ones that we mentioned at the beginning: research and development and then this demonstration phase, technology demonstration phase, and then the deployment phase. But that also clearly then allows us to identify the work that has been done correctly and in a good way, in the right way. Because it's, yeah, it's a, you know, there is a two-fold benefit. Also for the researcher, that needs to have feedback on that, on the activity performed.

In order to make sure there was no misunderstanding, the idea behind the HRL was quickly repeated:

> Interviewer: In fact, the criteria for HRL three are the gate criteria to go to HRL four. So.. uh because I don't know if you noticed that at every level, actually, the topics are the same, the questions are the same. But instead of testing it in a workshop, you will test it in a simulation, then you will test it in operational environments, etc. So they should follow each other up. And, actually, HRL three, if you finish that, it also includes HRL 2 and HRL 1, so I'm trying to understand what uh.
>
> Interviewee: Yeah yeah yeah, I am reading now. Yes, yes. Okay. I got it. Yeah, so maybe you just need to make more evident the gate and maybe to make it concrete, associating the gate to some activities, some toolbox, as you call it here, which is something that needs to be assessed, like experimental concept test, or initial validation, early validation of the concept or something like that.

In conclusion, it should be more evident how it should be assessed whether a level has been achieved. Although this sub-theme (‘start validation’) contains 14 codes, it should be noted that they all originate from the same interview.

Tool classification. The last theme summarises all comments made regarding the toolbox or, more specifically, the classification of the tools. The experts from ProRail stated that they do not distinguish between tools used in HRL 4, 5 or 6. Instead of classifying the tools in HRL 4 to 6 by maturity level, they suggested the tools could be classified by type, in the sense of whether they are, for example, psychophysiological, questionnaires, and so on.

Additionally, they suggested a classification based on which Human in the System impact the tool assesses. The expert familiar with SESAR projects noted that Eurocontrol and SESAR already develop a tool repository and that it would be nice to integrate this instead of creating a new one. The selection of tools generally seemed adequate. However, some tools were not used by all experts because the tools are more related to Safety than to HF or because they are specific to the aviation domain. One of the experts suggested:

> That link, that step to the toolbox, is much more domain-specific, I think. Or there should be some, yes, at least a lot of things in here. Yeah, look, interviews, debriefing, over-the-shoulder observation, of course, I know that. But it also contains many tools that we do not use at all. So I can imagine that you have a very general variant and one that is domain-specific, that you can perhaps consider that as two versions.
>
> Interviewer: Yeah, okay, a sort of classification, so interviews and things like that, and then you can add domain-specific tools to it yourself.
>
> Interviewee: Yes. Yes.[2]

So, a classification based on the industry is also one of the suggestions.

4.5 Lessons learned from HRL v2 evaluation

Five HF experts applied the HRL v2 to four different cases of different maturity levels. Based on the interviews, it can be concluded that many systems development models are already being used and that these must be mapped to the HRL. The connection with other models should be clear. The link between the HRL and TRL strengthens the HRL and should be maintained as much as possible. On the other hand, the HRL should emphasise its unique selling points relative to existing models. Its focus on the human in the system, including specific HF categories, can make it a valuable overview for project managers and HF experts. Furthermore, the HRL generally seemed to fit the current practices well. However, the relevance of each HF category for each project and at every level of maturity can be questioned. It should be demonstrated that the integrated approach of including all HF categories throughout the systems development process leads to increased safety. In addition, there are some potential misalignments between the timing and number of testing iterations found in the HRL and the current practice. Like in the HRL, different types of simulations of different levels of representativeness are used in practice. However, these simulations might start earlier than the HRL suggests. The interviewed experts use simulations as soon as possible to test their concept with users and clients. The HRL should better specify how HRL 1, 2 and 3 are assessed. Moreover, whereas the HRL was built to support the design of *new* systems, the current practices always view the new scenarios in light of the current situation. Therefore, it is vital that systems development starts with the analysis of the current situation and tests the system’s impact compared to his baseline or reference scenario. The input for this comes from HRL 9, and in that sense, the HRL is a circular model. In general, the philosophy of the HRL came across as clear. Unclarity can be found with regard to the different levels. Redundancy makes it currently difficult to distinguish the levels. In addition, the choice of words regularly led to confusion and multiple possible interpretations. Therefore, they should be either rephrased or better explained, for example, with the help of examples. A notable comment concerning the phasing was that the words make the statements sound too definitive and thus undermine the idea of continuous feedback that is the backbone of the HRL. Furthermore, the HF categories should be specified, possibly with a leading question or statement and a framework of subcategories as was developed for the Human in the System category. Lastly, the toolbox could become more user-friendly if the right classifications are used. The classification based on maturity levels makes sense, although there is probably no particular difference between tools used in HRL 4 to 6. Additionally, tools could be classified by type, for example, methods, techniques and tools, or by which HF (sub)category they are related to. In this regard, other organisations have already developed frameworks that could be integrated into the HRL. Moreover, there could be a toolbox section with generic tools and a section with tools specific to a particular domain.

5. Discussion

In this chapter, the main findings and conclusions for each of the research objectives will be presented. In addition, the last section of the chapter, Section 5.5, provides a reflection on the strengths and weaknesses of the study.

5.1 Objective 1: Potential points of improvement for the HRL v1

First of all, current practices were explored and compared to the HRL v1 in order to generate potential points of improvement for the framework (that is, Research Objective 1). To generate a rich but manageable insight into current practices, five systems development projects, four validation frameworks and one expert were consulted. Secondly, the HRL v1 was compared to CSE approaches as described in the literature. In total, the comparison yielded 23 potential issues and 18 suggestions for improvement of the framework, which were presented in Table 4. The main conclusions are presented in this section, together with points for discussion and suggestions for potential follow-up research.

It was found that technology is often already developed to a certain extent by the time the operational or HF efforts are employed. Consequently, the system will be tested in an environment that corresponds to a higher HRL, even though it could be the first validation in terms of HF. Based on the strategies that other validation frameworks use to deal with this issue, it was suggested that the HRL should require that at least the ‘definition’, or ‘validation’, phase is completed regardless of the current maturity to ensure that *the right system is being built*. In case the system is already in the verification phase, in which the question is whether *the system is built in the right way*, it should be up to an HF specialist to refer to the HRL argument structure for the previous level(s) to identify which activities should be carried out before the activities of the given maturity level can be conducted. However, this can easily be abused, and the HF experts might be pressured to give the ‘good to go’ by the other parties involved in the system development. The HRL is meant to signal that it is essential to test systems from an HF point of view from the beginning of the systems development. Therefore, it would be a bad sign to give the impression that it is acceptable to assign a high level when HF have not been considered in previous development and validation steps. On the other hand, making the framework too idealistic will make it impractical for experts in the field and therefore decrease the chances that it will be adopted. The present work has not found a satisfying solution for this issue. So, it would be interesting food for thought for the HRL workgroup. Among the projects considered in this research, there were a couple of cases in which the technologies were developed before the system concept had been defined. In both ALICIA and ACROSS, several (relatively mature) technologies were combined to form a new system. In this case, the system would need to start from HRL 1 again. These cases also show that systems development is not a linear process. This aligns with Lo (2020, p. 194), who states that iterations between design process stages are possible as tested designs may need adaptation.

Another challenge the HRL has to deal with is the abundance of existing frameworks specific to the aviation industry, such as the E-OCVM. The HRL should show how the outputs obtained during its nine levels feed into the currently required documentation. Additionally, the levels of the HRL scale should be mapped onto maturity levels of established frameworks. A first step was made in this thesis by mapping the HRLs to the CLM V-phases, but the results of this are not definitive. Moreover, the proposed mapping has not been evaluated by experts. A follow-up study could let experts assign V-phases and HRLs to several systems under development to understand better how the two maturity scales are linked.

In other validation frameworks and projects, validation or maturity was linked to a form of risk assessment. The HRL does not include a risk assessment, so this could be a point of improvement. However, the risk assessments offered by other frameworks are only based on subjective ratings of probability and gravity of potential issues. In essence, this is also captured by the sub-question on identifying potential human systems issues and risks. Here, additional existing risk assessments could be used complementary to the HRL. So the HRL v1 sub-questions are not adapted to the system's level of risk but prompt one to consider potential risks and create requirements based on this. It could also be stated that HRL is meant to support the development of safety-critical systems only. For the developments of simpler, lower-risk, closed systems or technologies, other processes can be used.

Other suggestions included integrating the CWA, the Coactive Design method, the HF pie categories and the HP impacts. Additionally, several suggestions were made for making the HRL v1 more comprehensible. Namely, it was suggested to move the more managerial sub-questions to a separate framework, add a toolbox, and make the framework visually stronger so that the relation between the HRL and the TRL as well as the interrelations of the sub-questions within the HRL framework would be more evident.

It should be emphasised that the suggestions made based on the comparison of the HRL v1 to both current practices and relevant literature could be replaced or complemented with alternative strategies for dealing with the detected areas for improvement.

5.2 Objective 2: The utility of the HRL v1 within European Aviation projects

Apart from points of improvement, the analysis of current practices also provided insights into the potential utility of the HRL framework (that is, Research Objective 2). This section will briefly describe and discuss the main conclusions concerning the HRL v1's utility within the European Aviation projects.

In all projects, there was a need for a framework that could structure the validation process, more specifically, one that can express maturity from an HF perspective. A project starts typically by identifying the current and target maturity and, subsequently, selecting corresponding validation activities. One of the analysed projects is still ongoing at the time of writing and needs to recommend the next steps towards higher maturity. In order to do this, a validation framework could be of great use. In the studied projects, a couple of other frameworks, such as the E-OCVM and SESAR HPAP, were used to fulfil the need of structuring the (HF) validation process. This could be seen as proof against the utility of the HRL since there are already established processes or in favour of the HRL because there is a 'market' for frameworks like the HRL. Many similarities were found between the HRL and the other frameworks. For example, the E-OCVM and the HRL both treat validation as well as verification, provide a means of supervising and synchronising different exercises and workstreams throughout the development life cycle, and promote HF validation activities throughout the system development life cycle. The various frameworks also showed overlap in terms of topics they considered. For example, they all paid attention to manpower, personnel and training. At the same time, the HRL also distinguished itself from the existing validation frameworks. The HRL offers more levels and, in this way, more granularity. But most importantly, the HRL is more in accordance with the TRL than other maturity frameworks. The TRL was introduced into EU funded projects in 2014 as part of the Horizon 2020 program but is insufficient as the success of the operational concept depends on more than just technical readiness (Enspire Science, n.d.). If the HRL would be adopted in a similar way as the TRL has been, it could help prevent HF efforts from only starting once it is time for a simulation exercise.

5.3 Objective 3: The perceived usability and utility of the HRL v2

The HRL was adapted based on the points of improvement that resulted from the analysis of current practices and relevant literature. HF experts applied the resulting HRL v2 to four different cases of different maturity levels. This 'user test' led to additional insights into the HRL's usability and utility (that is, Research Objective 3). The main findings will be presented in this section, together with points for discussion and suggestions for potential follow-up research.

Link between the TRL and HRL. The HF experts believed that the link between the HRL and TRL especially strengthens the HRL and should be maintained as much as possible. Visualising the HRL next to the TRL, in the same thermometer-like shape, thus appears to have been a meaningful redesign suggestion. However, the question is to what extent the HRL can mirror the TRL. For example, the target application of the TRL might be broader than that of the HRL. The current practices, as represented in this thesis, were formed by projects and experience concerning systems that were developed for 'workers'. This made sense as the CWA is specifically suitable for very complex systems that require to be operated by workers with a high degree of competency (Vicente, 1999). It is uncertain for what degree of complexity the HRL can be used and whether this is similar to the range of applications for the TRL. Vicente mentioned several criteria for calling something a complex, or 'open system', that is, "(a) large problem spaces, (b) social, (c) heterogeneous perspectives, (d) distributed, (e) dynamic, (f) potentially high hazards, (g) many coupled subsystems, (h) automated, (i) uncertain data, (j) mediated interaction via computers, and (k) disturbances" (Vicente, 1999, p. 14). Nevertheless, the open-closed system dimension is continuous, so it is hard to draw a line. For now, the decision to apply the TRL and HRL to certain systems and not to others remains in the skilled hands of systems developers and HF specialists, at least until experience leads us to identify best practices.

The relevance of the HF categories. Although the HRL generally seemed to fit the current practices well, the relevance of each HF category for each project and at every level of maturity was questioned by some of the HF experts. That does not mean that they thought the categories were irrelevant. On the contrary, the HF categories received only compliments. This is an improvement in contrast to the HSI domains. Thus, the problem does not concern the categories' utility by itself but their utility across the various HRLs. According to the HF experts, Organisation & Staffing, Training & Development, and Teams & Communication are often not considered until later in the systems development process. Similarly, in a participatory design process, Training is usually a step in the design process rather than a factor that needs to be continuously considered (Lo, 2020). Right now, the HRL framework is structured so that each of the exit criteria needs to be fulfilled before it is possible to continue to the next level, just as a chain is only as strong as its weakest link. That would mean that not attending, for example, Training, leaves a system in HRL 1 no matter how advanced the validation on other HF categories has been. Nevertheless, the HRL is system-centred rather than project-centred. The advantage of this is that it can be used to keep track of system maturity when the system development is divided over various programmes, projects, partners, and work packages. This implies that an HRL can be obtained by combining the results of one effort with the results of other efforts. Thus, in the example above, HRL 2 could be completed by a project that focuses solely on the Training aspect. Alternatively, the HRL could be split into separate readiness level scales for each category, as Miller and colleagues (2016) did for the HSI domains. However, it can be argued that the complexity of this structure only discourages system developers from using the HRL. However, it is necessary to demonstrate that including all HF categories throughout the systems development process increases system safety. If this cannot be demonstrated, this demanding approach should not be suggested by the HRL. This open issue could remind one of the observation that the HF effort often starts once the system can be tested in a simulator. In both cases, it would be a bad sign to give the impression that it is acceptable to step in at a higher maturity level rather than following all the HRLs starting from HRL 1, either for certain HF categories or for HF in general. On the other hand, making the framework too idealistic will make it impractical for experts in the field and render it useless. Future efforts should concentrate on finding the right balance based on evidence of what improves current practices and what does not.

Modifying systems versus designing new ones. Whereas the HRL was built to support the design of *new* systems, the HF experts made it clear that they always view the new scenarios in light of the current situation. Therefore, it was concluded that it is essential that systems development starts with the analysis of the current situation and tests the system's impact compared to this baseline or reference scenario. This might already be captured by the HRL or, more specifically, by the exit criteria on the analysis of comparable systems, the documentation of basic HF research and the identification of a benchmark for successful performance. Whether this is the case or not should be clarified in a guiding document. If it is not the case, additional exit criteria could be added to support the development of a baseline based on the appraisal of the current situation. However, this is only relevant *if* the HRL is to be used for validating changes rather than new systems. Alternatively, the utility of the HRL could be specific to a particular type of innovation, that is, radical rather than incremental. In line with this, Vicente (1999, p. 352) pointed out that "the CWA is based on a formative approach that takes a clean slate perspective on design, whereas most corporate design projects are evolutionary in nature". He did, however, demonstrate that the CWA *can* be used for evolutionary system design projects, but he also warned that using it in this way might make the CWA lose its value to a certain extent. At the same time, there might not be a stark difference between the two. Even in the case of a new system, there is some sort of baseline. For instance, when introducing a head-up display in cockpits to support pilots when landing, this landing support system might be completely new, but its influence on the pilot's performance can be compared to the performance of pilots without the system. In that sense, we always speak of introducing a change rather than a new system. The question becomes:

When should the development and validation start from HRL 1, and when is it part of the upgrades and modifications of HRL 9.

Timing and number of validation moments. In terms of usability, there are some potential misalignments between the timing and number of testing iterations found in the HRL compared to the current practice, which complicates the assignment of an HRL to a project. Like in the HRL, different types of simulations, of different maturity levels, are used in practice. The different simulation environments in the HRL v2 were not all clear to the experts. Mainly the 'relevant environment' provoked questions. In Lo (2020), the following research environments are distinguished: laboratory, simulated, and field. The first one is in line with TRL/HRL 4, the second one could replace the 'relevant environment', and the last one would be the same as the operational environment. A more severe issue is that simulations might start earlier than the HRL suggests. The interviewed experts use simulations as soon as possible to test their concept with users and clients. At lower levels of maturity, these could include workshops. So, workshops should not only be used for concept definition but also concept testing. Focus groups and workshops are currently already suggested by the framework from HRL 1 onwards, so it is possible that there is no difference between current practices and what the HRL suggests, but that this is just a misunderstanding. The HRL should better specify that these tests take place at HRL 1, 2 and 3 or propose another strategy for assessing the achievement of the levels. The HF expert with much experience with SESAR projects was explicitly convinced that she would need to test the principles and requirements before going to the next phase. However, this is precisely what happens in the next phase. From HRL 4 onwards, a prototype will be developed based on the principles and requirements, and this will be tested in different simulations. Perhaps the HRL v2 is confusing because the shape of the levels is equally large. Therefore, it might seem as if each level takes equally long and contains a similar number of internal iterations. However, as the HF experts from ProRail noted, some levels might contain many iterations relative to the other levels. Perhaps it would have helped the HF expert in SESAR to show that HRL 1, 2 and maybe even 3 are all part of CLM V1. According to the CLM V-phases, analysing, modelling and running simulations is at the centre of attention in V2, corresponding to HRL 4 and 5.

Remaining unclarities. There was still unclear terminology present in the HRL even after the first round of redesign in which industry-specific words such as “mission” were omitted and the HSI domains were replaced by the HF categories. The choice of words regularly led to confusion and multiple possible interpretations. Therefore, it was concluded that these should be either rephrased or better explained. A notable comment concerning the phrasing was that the words make the HRL statements sound too definitive and thus undermine the idea of continuous feedback that is the backbone of the HRL. It should be clarified that completing, for example, HRL 3 does not mean that requirements cannot be updated anymore later on in the process based on results from a simulation exercise. However, in Section 5.1, it was mentioned that iterations between stages are possible, for example, when several new technologies are combined in a new technological system. To a certain extent, updating the concept, the requirements or building a new prototype as a result of higher-order simulations does not undo the allocation of HRLs to the system. However, when the concept drastically changes, this new concept would have to go through the different HRLs, starting from the beginning. Another confusing phrase in the HRL v2 was the statement: “Environmental constraints identified.” Some of the experts presumed that the ‘environment’ also includes human needs. According to the CWA, these come later, after the work domain, the tasks, strategies, and social organisation have been analysed. This is contrasting to UCD, where the needs of the user are the point of departure. At Deep Blue, the UCD approach is very prominent. This contrast could also help to explain why the expert with experience in SESAR would want user tests starting from lower HRLs. Nevertheless, one method or design philosophy does not necessarily exclude the other. Users can be involved in identifying the needs, generating scenarios, developing the concept, discussing the function allocation, and so on, while respecting the principle of analysing cognitive constraints after environmental constraints have been identified. These unclarities demonstrate that there is still room for improvement in terms of usability.

Usability and utility of the toolbox. Lastly, the toolbox could be made more user-friendly with the use of the proper classifications. The wheel does not need to be reinvented as there are already existing tools repositories. For example, the SAFEMODE consortium developed the Human Assurance Toolkit for the aviation and maritime industry in which they included 24 methods classified by class (that is, “technique”, “model” or “process”), design process phase, level of HF expertise needed, and level of method maturity (Save et al., 2020). Experts could also be provided with a standard toolbox and the option to customise it further, depending not only on their industry but also on the resources that they have available within their organisation. From the comments of the HF experts and their enthusiasm in thinking along about possible classifications, it was evident that integrating a toolbox into the HRL was perceived as utile. In general, the HRL proved to be very agile. Several concepts, such as the CWA and the Coactive Design method, were integrated without problems. The HRL could function like a ‘coat rack on which we can ‘hang’ important HF concepts.

5.4 Objective 4 and 5: Implications for Deep Blue

This thesis was written in collaboration with Deep Blue. Therefore, this thesis not only aimed to provide useful input for the HRL, but also Deep Blue. One objective was to find out to what extent the current practices at Deep Blue are already in line with the HRL (Research Objective 4). Knowing this can help Deep Blue to understand how easy it would be to adopt the framework. This understanding is valuable in case the HRL would become a standard for European research projects. But even without external pressure, Deep Blue may decide that it wants to adopt the framework because it sees the benefits, especially as a means of maintaining an overview and supporting communication. This section will summarise the conclusions about the alignment of current practices at Deep Blue with the process suggested by the HRL v1 and v2 frameworks.

Based on the analysis of current practices, several HRL v1 sub-questions were identified that were answered as part of the projects. For example, in ALICIA, sub-questions concerning performance requirements, metrics, usage scenarios and tasks had been answered. Also sub-questions concerning user roles, procedures, human-machine function allocations, human-system issues, manpower, personnel, training, supporting human use, design principles, and basic human research were attended by the projects selected to paint a picture of current practices at Deep Blue. Since most projects at Deep Blue strictly follow the SESAR HPAP and E-OCVM, the overlap between these and the HRL was studied as part of the present work. The HRL v1 was redesigned to fit current practices even better by, for example, replacing the HSI v1 domains with the HF pie categories while preserving the sub-questions mentioned before.

Based on the evaluation of the redesigned HRL through a series of interviews with HF experts, additional input was collected regarding the conformance of the HRL v2 to current practices. In a general sense, the HRL v2 was said to be easy to understand because it represents a very well-known process. But also on a more detailed level, the HF experts from Deep Blue identified similarities between the HRL v2 (exit criteria) and the activities they conducted in former projects. Nevertheless, some results indicated that current practices at Deep Blue deviate from the HRL. First of all, certain HF categories are often lagging behind the Human in the System, Working Environment and Procedures, Roles & Responsibilities categories. Whereas Organisation & Staffing, Training & Development, Teams &

Communication are attended throughout the HRLs, these often only start to be considered at a later stage in the projects that Deep Blue was involved in. Secondly, it was not clear whether the timing and the number of testing moments in the HRL are in line with the standards at Deep Blue. However, if validations are indeed starting earlier and contain more iterations in practice than is suggested by the HRL, this would not necessarily be a problem. It would mean that systems are currently subjected to more stringent and intense testing than they need to be according to the HRL. One potentially big gap between the HRL and current practices at Deep Blue might be that, following the HRL, systems development starts by analysing the work domain, rather than the user needs as is standard practice at Deep Blue. This difference makes sense since the analysis of current practices demonstrated that Deep Blue deals with changes to existing systems rather than developing completely new systems from scratch. As discussed before, the CWA helps overcome the task-artefact cycle and is, therefore, explicitly aimed at designing for innovation, whereas descriptive methods such as anthropological analysis, scenario-based design, and rapid prototyping are dependent on a particular device. So, Deep Blue could consider using the HRL for projects that require a clean slate perspective. In order to lower the effort of HF experts even further and make sure the HRL is more achievable from a commercial point of view, prototypical templates could be developed for the analysis of general constraints of certain classes of application domains such as aviation, finance, nuclear energy, manufacturing and healthcare. In this way, HF experts do not need to conduct a new work domain analysis at the start of each systems development process. This is possible because the work domain analyses are event-independent and, therefore, adaptable to idiosyncratic details of particular applications (Vicente, 1999, p. 351). Creating the template might require considerable effort but would pay itself off in subsequent projects. They could even be exploited as a new product.

Apart from these insights, Deep Blue has been provided with the HRL v2, a product of the present work (Research Objective 5). In addition, the mapping of the HRLs on the V-phases will help experts at Deep Blue to use the HRL complementarily to the E-OCVM. The next step will be to integrate also the feedback that resulted from the evaluation of the HRL v2.

5.5 General reflection

The conclusions with regards to the research objectives should be considered in light of the research' limitations. First of all, despite the valuable insights resulting from the analysis of current practices, solely basing one's conclusions on an analysis of current practices has its limitations, as current practices may be suboptimal. In order to limit the implications of this, the present work also looked at systems engineering approaches described by CSE literature, and it included an evaluation in which HF experts actually used the HRL v2.

In addition to this weakness, the results are based on a limited number of projects and experts. Therefore, they could be specific to the aviation industry, the companies involved in the study, or even the HF experts. The conclusion should be tested amongst a wider audience. In follow-up research, the HRL v2 could be further updated based on the insights from this study and then tested through a survey. This time, the focus would not be on collecting qualitative input but on validating the redesigned HRL. Another limitation of the study is that, because of its qualitative nature, the results cannot be compared with those of Handley (2020). However, I believe that a qualitative approach fits the current maturity of the HRL framework better than a quantitative approach. Once a design concept has been completed, it can undergo quantitative testing in studies like the one mentioned before, in which a larger target group is involved.

Furthermore, the present work used a prototype of an HRL developed by the HRL workgroup rather than the version it will publish. It is uncertain how much the HRL that will be published will deviate from the initial version that was used as the basis of the present work. Consequently, the findings might turn out to be partially irrelevant because, for example, some of the proposed points of improvement are already integrated. There is also a risk that the HRL that will be published deviates a lot from the prototype and that it becomes difficult to integrate the suggested implementations from the present work. However, even in this case, the current work is of value because it provides rich insights into current practices. Additionally, for each of the suggestions in this thesis, the underlying reasoning and source were shared. So even though the HRL workgroup might decide not to adopt the suggestions themselves, the input from the analysis and evaluation is invaluable.

Another limiting factor of this study is that the focus was on applying the HRL to activities that already took place. The present work did not spend much attention on how the HRL could be used for planning purposes. For this, a longitudinal study would be more suitable. HF experts in charge of a project could be asked to plan their project according to the HRL and then evaluate the HRL's usability and utility when planning the project, during the project and after finishing the project. The current study also did not challenge the experts too much to explain how they reached the exit criteria. The interviews were simply not long enough to go into this much detail. This could be the topic of a follow-up study. Dr Savage-Knepshield has been identifying which project events provided the answers to which HRL questions (P. Savage-Knepshield, personal communication, November 3, 2020). The outcome of this will probably be published in the special HRL issue of Ergonomics on Design.

One of the strengths of the present work is its practical approach. During the past ten years, the development of the HRL was very conceptual. Moreover, most authors researching the HRL started all over rather than building forth on the work of other authors. The present work took a comprehensive look at the evolution of the HRL so far and the concepts already present. After careful consideration, an existing framework was selected and analysed in the light of established practices. In order to further increase the study's practical value, the HRL was adapted to reflect the outcomes of the initial analysis and subsequently evaluated again. So the current study not only pinpointed issues but also proposed possible solutions and integrated these. In addition, the experts who shared their insights regarding the HRL all had a strong academic foundation, but more importantly, much practical experience. Thus, the present work not only contributed to the HRL as a theoretical concept but also to the HRL as a practical tool. Finally, the research both zoomed in on a specific field of application - European aviation projects - and took a broad perspective. The number of interviewed experts is too limited to conclude that the HRL is generalisable across industries, but the absence of strong disagreements amongst the experts and the shared enthusiasm for concepts such as the HF categories is a promising sign.

6. Conclusion

The present work originated from the belief that considering HF throughout the systems development process increases safety and that HRL could be the tool to structure and promote it. In order to be adopted, the HRL would need to be consistent with the current practices of HF experts while also adding additional value. The HRL was subjected to two testing iterations. First, the framework was analysed by comparing it to current practices at Deep Blue and relevant literature. Based on the input from this analysis, the framework was redesigned and subjected to a second round of testing in which experts from various industries used it. The research succeeded at finding points of improvement, improvement strategies, and important obstacles such as the mapping of the HRL to established models. This thesis has contributed to the development of new insights on systems development and validation from an HF point of view and to the formation of an improved HRL framework. In this way, the present work supports the development of an HRL that can hopefully empower HF experts and project managers and spark the interest of system engineers.

References

Aliciaprojectvideo. (2011, May 3). *ALICIA project new cockpit concepts* [Video]. YouTube. https://www.youtube.com/watch?v=mXXoZU3T6gY

Anderson, M. (n.d.). *Air France Flight 447: What happened?* Human Factors 101. https://humanfactors101.com/incidents/air-france-flight-447/

Billiere, B., Dautermann, T., Dokic, J., Lanzi, P., Liston, P., Maret, T., & Mosquera Benitez, D. (2013). *D11.1-1: Validation plan* (Report No. ACP2-GA-2012-3141501) [Unpublished report].

Bonelli, S., Tomasello, P., Ragosta, M., Granger, G., Turhan, U., Açikel, B., Bayar, S., Ozan, A., Borghini, G., Di Flumeri, G., Sciaraffa, N., Aricò, P., & Drogoul, F. (2017). *D4.1: Validation plan* [Unpublished report].

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101. https://doi.org/10.1191/1478088706QP063OA

Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology, 38*(1), 155-200. https://doi.org/10.1111%2Fj.1467-9531.2008.00203.x

Campbell, D. (2019, May 2). *Redline: The many human errors that brought down the Boeing 737 Max*. The Verge. https://www.theverge.com/2019/5/2/18518176/boeing-737-max-crash-problems-human-error-mcas-faa

CORDIS. (2017, March 31). *HP neurometrics toolbox for highly automated systems design*. https://cordis.europa.eu/project/id/699381

CORDIS. (2019a, July 16). *All condition operations and innovative cockpit infrastructure*. https://cordis.europa.eu/project/id/233682

CORDIS. (2019b, July 16). *All condition operations and innovative cockpit infrastructure*. https://cordis.europa.eu/project/id/233682/reporting

CORDIS. (2019c, August 2). *Advanced cockpit for reduction of stress and workload.* https://cordis.europa.eu/project/id/314501

CORDIS. (2021, January 17). *Safe landing through enhanced ground support.* https://cordis.europa.eu/project/id/890599

Dallat, C., Salmon, P. M., & Goode, N. (2018). Identifying risks and emergent risks across sociotechnical systems: The NETworked Hazard Analysis and Risk Management System (NET-HARMS). *Theoretical Issues in Ergonomics Science, 19*(4), 456-482. https://doi.org/10.1080/1463922X.2017.1381197

Davies, A. (2019, September 26). *Feds say Boeing 737s need to be better designed for humans*. Wired. https://www.wired.com/story/feds-boeing-737s-better-designed-humans/

Deep Blue, Eurocontrol, DLR, European University Institute, LFV, EASN-TIS, EuroUSC, Airholding S.A., & Swiss International Air Lines. (2019). *SAFELAND: Safe landing through enhanced ground support* [Unpublished report]

Dekker, S. (2009). *Flight crew human factors investigation*. Onderzoeksraad. https://www.onderzoeksraad.nl/nl/media/inline/2020/1/21/human_factors_report_s_dekker.pdf

Del Giudice, K., Hale, K., & Johnston, M. (2015). Development of a scale for quantifying a system's human factors readiness. *Procedia Manufacturing*, *3*(2015), 5192-5198. https://doi.org/10.1016/j.promfg.2015.07.580

Diehl Aerospace, Westland Helicopters, Thales, Dassault Aviation, Alenia Aeronautica, BAE Systems, Jeppesen, A-Volute, & IntuiLab. (2010). *State-of-the-art technologies route mapping* (Report No. ALICIA/D/DASP/WP1-0004) [Unpublished report].

Enspire Science. (n.d.). *TRL scale in Horizon Europe and ERC: Explained.* https://enspire.science/trl-scale-horizon-europe-erc-explained/

Eurocontrol. (2007). *Guidelines for the human factors aspects of safety assessment: Human Assurance Level* (Report No. SAF.ET1.ST03.1000-HAL-01-00) [Unpublished report].

Eurocontrol. (2010). *E-OCVM version 3.0: Volume I*. ICRAT. http://www.icrat.org/icrat/seminarContent/2010/European_Operational_Concept_Validation_Methodology.pdf

European Commission. (2014). *Horizon 2020 work programme 2014 – 2015: General annexes.* http://ec.europa.eu/research/participants/portal/doc/call/h2020/common/1617621-part_19_general_annexes_v.2.0_en.pdf

GAO. (1999). *Best practices: Better management of technology development can improve weapon system outcomes (GAO/NSIAD-99-162)*. https://www.gao.gov/assets/160/156673.pdf

Gates, D., & Baker, M. (2019, June 22). *The inside story of MCAS: How Boeing's 737 MAX system gained power and lost safeguards*. The Seattle Times. https://www.seattletimes.com/seattle-news/times-watchdog/the-inside-story-of-mcas-how-boeings-737-max-system-gained-power-and-lost-safeguards/

Graettinger, C. P., Garcia-Miller, S., Siviy, J. M., Schenk, R. J., & Van Syckle, P. J. (2002). *Using the Technology Readiness Levels scale to support technology management in the DoD's ATD/STO environments* (Report No. CMU/SEI-2002-SR-027). Carnegie Mellon University. https://resources.sei.cmu.edu/asset_files/specialreport/2002_003_001_13931.pdf

Hale, K., Fuchs, S., Carpenter, A., & Stanney, K. (2011). A scale for assessing human factors readiness levels. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *55*(1), 2030-2034. https://journals.sagepub.com/doi/pdf/10.1177/1071181311551423

Handley, H. A. (2020, October 5-9). *Understanding Human Readiness: Usability and Rater Reliability* [Conference session]. HFES 64th International Annual Meeting, online.

Haskins, B., Stecklein, J., Dick, B., Moroney, G., Lovell, R., & Dabney, J. (2004). Error cost escalation through the project life cycle. *INCOSE International Symposium*, 14(1), 1723-1737. https://tinyurl.com/y6essavj

Héder, M. (2017). From NASA to EU: The evolution of the TRL scale in public sector innovation. *The Innovation Journal: The Public Sector Innovation Journal, 22*(2), 1-23. https://www.innovation.cc/discussion-papers/22_2_3_heder_nasa-to-eu-trl-scale.pdf

HFES. (2020, November 18). *Human Factors and Ergonomics Society (HFES) applauds Boeing and FAA for resolving 737 MAX human factors design flaws*. Newswise. https://www.newswise.com/articles/human-factors-and-ergonomics-society-hfes-applauds-boeing-and-faa-for-resolving-737-max-human-factors-design-flaws

Hoffman, R. R., & Woods, D. D. (2011). Beyond Simon's slice: Five fundamental trade-offs that bound the performance of macrocognitive work systems. *IEEE Intelligent Systems*, *26*(6), 67-71. https://doi.org/10.1109/MIS.2011.97

Hollnagel, E. (n.d.). *Cognitive Systems Engineering: RIP*. https://erikhollnagel.com/ideas/cognitive-systems-engineering.html

Human Engineering Limited. (2007). *Application of Human Assurance Level to RA Downlink* (Report No. HEL/EC/061578/RT01) [Unpublished report].

Human Factors & Aviation Safety: Hearing on Boeing 737-Max8 crashes: Testimony to the United States House of Representatives (2019) (testimony of Mica R. Endsley, PhD). https://transportation.house.gov/imo/media/doc/Endsley%20Testimony.pdf

Jander, H., Borgvall, J., & Ramberg, R. (2012). Towards a methodological framework for HMI readiness evaluation. *Proceedings of the Human Factors and Ergonomics*

Society Annual Meeting, 56(1), 2349-2353. https://doi.org/10.1177/1071181312561436

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction, 3*(1), 43-69. https://doi.org/10.5898/JHRI.3.1.Johnson

Johnston, M., Del Giudice, K., Hale, K. S., & Winslow, B. (2013). Development of a system for communicating human factors readiness. *International Conference on Human Interface and the Management of Information*, 475-484. https://link.springer.com/content/pdf/10.1007/978-3-642-39226-9_52.pdf

Krois, P., & Rehmann, J. (2005). Assessing human factors risks in air traffic management research. *2005 International Symposium on Aviation Psychology*, 419-424. https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=1062&context=isap_2005

Lo, J.C. (2020). *Gaming simulation and Human Factors in complex socio-technical systems: A multi-level approach to mental models and situation awareness in railway traffic control* (Publication No. 978-94-028-2078-2) [Doctoral thesis, Delft University of Technology]. Islandora.

Mauro, C. (n.d.). *The human factors science behind the 737MAX MCAS problem and why the MAX should be scrapped*. Mauro New Media. https://www.mauronewmedia.com/blog/the-human-factors-science-behind-the-737max-mcas-problem-and-why-the-max-should-be-scrapped/

Mellett, U., & Nendick, M. (2007). *The Human Factors Case: Guidance for human factors integration* (Report No. 978-2-87497-002-3). Eurocontrol. https://www.skybrary.aero/bookshelf/books/4556.pdf

Miller, M., Thomas, S., & Rusnock, C. (2016). Extending System Readiness Levels to assess and communicate human readiness. *Systems Engineering*, *19*(2), 146–157. https://doi.org/10.1002/sys.21344

Morgan, C. C., Pelchen-Medwed, R., Averkova, D., Pinheiro, J. P., Feuerberg, B., Morel, M., Dehn, D., Botezan, A. D. (2018). *SESAR Human Performance Assessment Process V1 to V3 - including VLDs.* SESAR Joint Undertaking. https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/h2020-sesar-er4-26-2019_hp-v1-3_16.06.05_en.pdf

Naikar, N., Hopcroft, R., & Moylan, A. (2005). *Work domain analysis: Theoretical concepts and methodology* (Report No. ADA449707). Defence Science and Technology Organisation Victoria (Australia) Air Operations Division. https://apps.dtic.mil/sti/pdfs/ADA449707.pdf

Newton, V., Greenberg, A., & See, J. (2017). Project management implications and implementation roadmap of human readiness levels. *International Conference on HCI in Business, Government, and Organizations*, 99-111. Springer. https://www.osti.gov/servlets/purl/1424878

Nolte, W. L. (2008). *Did I ever tell you about the whale? Or measuring technology maturity*. IAP.

Nunes, A., & Laursen, T. (2004). Identifying the factors that led to the Ueberlingen mid-air collision. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 48*(1), 20-24. https://doi.org/10.1177/154193120404800142

O'Neil, M. P., Shattuck, L. G., & Sciarini, L. W. (2015). A framework for assessing and communicating human systems integration efficacy across the system lifecycle. *Procedia Manufacturing 3*(2015), 3054–3061. https://doi.org/10.1016/j.promfg.2015.07.851

Phillips, E. L. (2010). *The development and initial evaluation of the human readiness level framework* [Thesis, Naval Postgraduate School]. Calhoun.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-13(3), 257-266. https://doi.org/10.1109/TSMC.1983.6313160

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem. *Safety Science*, *27*(2-3), 183-213. https://doi.org/10.1016/S0925-7535(97)00052-0

Save, L., Licata, V., Cocchioni, M., Cichomska, K., Ragosta, M., Kirwan, B., Kilner, A., Bettignies-Thiebaux, B., Kling, F., Gibson, H., Waters, S., Kurt, R. E., Ramirez, C., De Maya, B. N., Efremov, A., Bieder, C., Alarcon, D. M., Louzi, K., Stamatopoulou, E. A. … MacKinnon, S. (2020). *D3.1: Human Assurance toolkit and guidance* [Unpublished report].

Scaravetti, D., & Montagnier, E. (2009). Human factor early involvement in product design processes. *Proceedings of ICED 09*, 6, 135-146. https://www.designsociety.org/publication/28732/Human+Factor+Early+Involvement+in+Product+Design+Processes

Schraagen, J. M. C. (2018). Naturalistic decision making. In L. J. Ball & V. A. Thompson (Eds.), *The Routledge international handbook of thinking and reasoning* (pp. 487-501). Routledge.

See, J. E., Craft, R., Morris, J., & Newton, V. (2017). Incorporating human readiness levels at Sandia National Laboratories. *2017 Resilience Week (RWS),* 150-153. https://doi.org/10.1109/RWEEK.2017.8088664

SESAR Joint Undertaking. (2012, October 22). *Situation Awareness Global Assessment Technique (SAGAT)*. Eurocontrol. https://ext.eurocontrol.int/ehp/?q=node/1601

60 Minutes Australia. (2019, May 27). *Rogue Boeing 737 Max planes 'with minds of their own'* [Video]. YouTube. https://www.youtube.com/watch?v=aO7_indbfME

SJU (2020). *Project handbook: Programme execution guidance*. https://ec.europa.eu/research/participants/data/ref/h2020/other/guides_for_applicants/jtis/h2020-guide-project-handbook-sesar-ju_en.pdf

SKYbrary. (2018, December 5). *T154 / B752, en-route, Uberlingen Germany, 2002.* https://www.skybrary.aero/index.php/T154_/_B752,_en-route,_Uberlingen_Germany,_2002

Smit, E. (2020). *Facilitating the incorporation of HF practice in a corporate design process through qualitative and consensus analysis* [Master's thesis, University of Twente]. University of Twente Theses.

Terenzi, M., Napoletano, L., Rozzi, S., Granger, G., Imbert, J., Aricò, P., Borghini, G., Graziani, I., & Babiloni F. (2014). *D3.1: Validation plan* [Unpublished report].

Titan Grey. (n.d.). *Boeing 737 MAX*. https://titangrey.com/boeing-737-max/

Usability.gov. (2020, September 30). *System Usability Scale (SUS).* https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html

Van Diggelen, J., & Johnson, M. (2019). Team design patterns. *Proceedings of the 7th International Conference on Human-Agent Interaction*, 118-126. https://doi.org/10.1145/3349537.3351892

Vicente, K. J. (1999). *Cognitive work analysis: Toward safe, productive, and healthy computer-based work*. CRC Press.

Waterson, P., Le Coze, J. C., & Andersen, H. B. (2017). Recurring themes in the legacy of Jens Rasmussen. *Applied Ergonomics, 59*(Part B), 471-482. http://dx.doi.org/10.1016/j.apergo.2016.10.002

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. CRC Press.

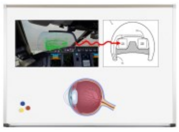
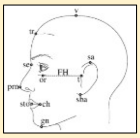
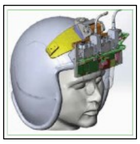
Wright, S., & O'Hare, D. (2015). Can a glass cockpit display help (or hinder) performance of novices in simulated flight training? *Applied Ergonomics, 47*(2015), 292-299. https://doi.org/10.1016/j.apergo.2014.10.017

Appendix A. HRL framework of the workgroup

Basic Research and Development: Scientific research, analysis, and preliminary development on paper and in the laboratory. This phase culminates in a validated proof-of-concept that addresses human needs, capabilities, limitations, and characteristics.	1	Relevant human capabilities, limitations, and basic human performance issues and risks identified	This first level of human readiness represents a broad, high-level exploration of human capabilities and limitations and basic human performance issues and risks relevant to the developing concept or proposed practical application.	Are human capabilities, limitations, and risks relevant to the proposed practical application being identified?	1	Have key user capabilities and limitations been identified?	At this initial readiness level, work is very conceptual, occurring in the laboratory and on paper. Human systems professionals can begin researching human capabilities and limitations (with respect to both traits and states) relevant to the developing concept (these will be further refined in HRL 2) for various potential user roles (operators, maintainer, monitor, inspector, etc.). They can also begin preliminary identification and research of plausible scenarios to develop an understanding of how various types of users might interact with the proposed practical application throughout the lifecycle--training, operation, maintenance, and sustainment. Emerging knowledge of the concept or proposed practical information may prompt relevant human research or highlight potential key human systems risks and issues for further research. Potential human systems risks and issues may include human performance challenges and limitations with the proposed technology or potential user interface features of the new technology. Basic human research begins in HRL 1 and may continue through Levels 2 and 3 as additional information about the proposed technology becomes available.	Human performance relevant to the developing concept or proposed application has been identified and characterized at a basic level.	Document key user capabilities and limitations. [1] Document potential usage scenarios. [2] List potential key human systems issues. [3] Document human research findings relevant to the developing concept, technology, or application. [4] Begin developing overarching human systems plan for the entire program and integrating within broader system plan. [N/A]	At the comparable TRL 1 level, a preliminary concept is being explored in the lab; little is known about eventual applications. Consequently, for HRL 1, human systems experts can begin addressing human involvement at a basic research level to start identifying the characteristics of the people who might use this concept and how (preliminary usage scenarios). The term "practical applications" is used here to align with the wording used later in TRL 2. It refers to the ways in which a concept or idea might be applied for some practical purpose, not to applications in the narrow sense of software applications. At HRL 1, human systems experts conduct basic research on the human capabilities and limitations relevant to the developing concept (to be further refined at HRL 2). This research spans all of the types of users who may potentially be part of the system (e.g., operators, maintainers, monitors, and inspectors). Human systems experts also learn as much as they can about the developing concept itself to understand the technological component of the system and potential impacts on human use.
					2	Are usage scenarios for potential users being identified?				
					3	Have potential key human systems issues and risks throughout the lifecycle been identified?				
					4	Is basic human research relevant to the developing concept or application being conducted?				
	2	Human-focused concept of operations defined and human performance design principles established	As practical applications are being invented or identified, implications for human involvement are analyzed concurrently. Relevant human performance design principles are developed to begin identifying human use requirements and provide inputs for preliminary conceptual designs.	Are human performance design principles relevant for user interactions and performance with the developing technology understood?	5	Have human systems professionals with requisite expertise been engaged and funded to support the lifecycle of this effort?	For successful application of the HRL scale, qualified human systems professionals must be engaged and funded. The number of human systems experts and the types of specialties required will depend on the nature of technology under development. In HRL 2, continued analysis of relevant human capabilities and limitations at both individual and team levels provides the basis to create design guidelines for supporting successful human performance with the technology. Key human performance design principles, standards, and guidance may be used to create an overarching human systems theme to establish consistency throughout later design and development. Task descriptions provide more detail around the preliminary usage scenarios begun in HRL 1 for the range of expected user roles and serve as a basis for future task analyses and human-machine function allocation decisions. Issues in legacy and comparable systems may be identified through interviews with experienced users or historical documentation such as issue reports. Legacy issues (from previously deployed systems and similar technologies) may identify potential sources of human error or mis-uses as well as features that should be included or excluded to prevent or mitigate such occurrences. Analysis of potential mis-uses at this early stage permits mitigation or prevention through subsequent engineered design. Practitioners should begin thinking about appropriate metrics for successful human performance to support activities at future HRL levels, including development of key performance parameters (KPP). Metrics may be objective (reaction time and error rate) or subjective (user perceptions, ratings, and opinions). These metrics will be continuously reviewed and updated as additional system information and human performance data become available.	Key human performance design principles, standards, and guidance for human interaction with the technology have been established.	Document expected level of effort and types of specialties required for human systems professionals throughout the lifecycle. [5] List relevant human capabilities and limitations by user role. [6] List and reference key human systems principles, standards, and guidance. [7] Provide overview descriptions of key tasks by user role and scenario. [8] List pros and cons from legacy and comparable systems to guide design and development. [9] List potential sources of human error or mis-use. [10] Define and document initial plausible metrics for successful human performance. [11] Update overarching human systems program plan to reflect emerging information and continue integrating within broader system plan. [N/A]	HRL 2 occurs as practical applications are being invented or identified during TRL 2. This knowledge gives the HSI team a firmer concept against which to analyze human capabilities and limitations and identify relevant human performance design principles, standards, and guides. Details may not be available at this point, but human systems experts can identify the sorts of things users might do as part of the user-centered concept of operations (e.g., they might use the technology while sitting). The NNSA readiness level calculator places the question about who will perform the work at TRL 2 (Q#10: Is it known who will perform the research or project?). Accordingly, HRL Q#5 addresses engagement and funding for human systems experts. When engaging and funding human systems professionals, it's important to recognize that different professionals have different specialties. More than one human systems expert may be required to provide support. Q#11 (Are plausible metrics for successful human performance being identified?) is included at HRL 2 to serve as a preliminary high-level exploration of human performance that must be accomplished before progressing to HRL 3. These metrics must be initiated at HRL 2 so their feasibility can begin to be addressed once a proof of concept is available at HRL 3. At HRL 4 and above, the HSI team will use the metrics to evaluate human performance at progressively higher levels of technology fidelity as the technology matures. Metrics may be objective or subjective. By HRL 6, objective indicators of human performance are emphasized, though subjective metrics may still be informative.
					6	Have relevant human capabilities and limitations been refined?				
					7	Have key human performance design principles, standards, and guidance been researched?				
					8	Are basic task descriptions for user roles being developed?				
					9	Has human performance on legacy or comparable systems been analyzed to understand key human-technology interactions?				
					10	Have potential sources of human error and mis-use been identified?				
					11	Are plausible metrics for successful human performance being identified?				
	3	Requirements for supporting human performance established	Analyses of human operational, environmental, functional, cognitive, and physical needs completed during analytical and laboratory-based studies of the proof of concept to understand the requirements for supporting each human user role.	Have human user needs, capabilities, limitations, and characteristics been mapped to expected operational and system demands to establish system requirements for supporting human performance?	12	Have usage scenarios been updated, based on human needs analyses for the proof of concept?	Knowledge gained through increased maturity of the concept as well as analytical and laboratory demonstrations of the proof of concept facilitates a broad range of human systems evaluations. Updated usage scenarios support design and testing. Detailed functional, task, and cognitive task analyses are conducted to determine user goals, strategies, and challenges and perceptual, information, and decision requirements for each user role. Scenario and task descriptions should characterize the visual, auditory, and tactile environments in which tasks will occur (luminance, glare, ambient noise, and vibration) as well as abnormal and emergency conditions. Attributes of usage scenario environments also have implications for multiple HSI domains, including environment, safety, and occupational health (e.g., musculoskeletal and chronic use disorders). Increased maturity of the concept permits analysis and evaluation of implications for any relevant HSI domains. With the increased maturity of the concept, human systems professionals can begin considering human-machine function allocations to optimize performance. As part of developing function and task descriptions, human systems professionals can use the proof of concept to identify human user tasks that are particularly critical for achieving system goals and may therefore require additional rigor in the analysis. Mappings of human capabilities, limitations, and needs to expected system demands are used to evaluate human performance implications, identify user requirements, and propose design features accordingly. Human needs may include cognitive needs (decision and information), and demands imposed on human users may be physical or perceptual. As a result of these analyses, specific human performance requirements are identified and flowed into higher level system level requirements. All relevant and applicable human performance data (for both individuals and teams) are collected and evaluated to understand whether metrics for successful human performance are feasible--accuracy, completion times, error types and frequency, workload, situation awareness, user satisfaction, vigilance, fatigue, impacts of PPE, and ergonomics (e.g., lifting, carrying, pinch points, awkward postures, reach, fit, layout, and accessibility). Specific KPPs to support successful human performance are identified. Human systems professionals can use all of the activities at this level to recommend design features that will best support human interactions within the developing system.	The implications of human interactions with key elements of the technology are evaluated analytically or via laboratory studies of a proof of concept. Requirements to support human performance are identified and flowed into higher level system requirements. Specific KPPs to support human performance are identified. Conceptualizations of the human role in the system are used to create preliminary design recommendations to accommodate human capabilities, limitations, and needs. Implications of the proposed technology design for all relevant HSI domains are evaluated (e.g., training needs). An initial determination of the feasibility of the human performance metrics proposed in HRL 2 is conducted, based on the proof of concept.	Document updated usage scenarios and function/task descriptions/analyses and identify tasks critical to system performance goals. [12, 13] Describe preliminary human-machine function allocations. [14] Identify human performance requirements to flow into higher level system requirements. [15] Identify specific KPPs to support successful human performance. [15] Document evaluations of implications for all relevant HSI domains. [16, 17, 18, 19] Document evaluations of implications for maintenance and sustainment. [20] Describe human performance implications derived from mappings of human capabilities, limitations, and needs to expected system demands. [21] Characterize human interactions and performance with the key elements of the technology that have been demonstrated analytically or via laboratory studies of the proof of concept. [22] List the human performance data used to show proof of concept and characterize the feasibility of identified human performance metrics. [22] Provide design recommendations to support human use. [23] Finalize overarching human systems program plan and finish incorporating into broader system plan. [N/A] Begin developing draft test plan for subsequent modeling and part-task testing with mockups. [N/Al]	At TRL 3, a proof of concept backed by analytical evidence is established. Elements of the technology are available, but not yet integrated or representative. Nevertheless, they are sufficient for human systems experts to begin understanding the demands the developing technology will impose on human users. In turn, human systems experts can analyze what the users will need from the system. This information can be used to determine user requirements. With respect to the various HSI domains, the questions specifically call out those HSI domains that are broadly applicable to many different types of systems; e.g., safety, environment, manpower, personnel, and training. Any other potentially relevant HSI domains (e.g., habitability and survivability) are combined and covered in a single question. Note that there is no question specifically for the human factors engineering domain because the topics for that domain are embedded in existing questions throughout the entire scale.
					13	Have function and task descriptions for each user role been updated, based on human needs analyses for the proof of concept?				
					14	Have candidate human-machine function allocations been evaluated, based on human needs analyses for the proof of concept?				
					15	Have requirements to support human performance been identified and included in system level requirements?				
					16	Are safety analyses for human users being completed?				
					17	Are manpower, personnel, and training analyses being completed?				
					18	Are environmental constraints and impacts being analyzed?				
					19	Are analyses for other relevant HSI domains being completed?				
					20	Are maintenance and sustainment analyses being completed?				
					21	Are human capabilities, limitations, and needs being mapped to expected operational and system demands to identify human performance issues and system requirements?				
					22	Have relevant human performance data been evaluated to determine the feasibility of metrics for successful human performance, based on the proof of concept?				
					23	Have preliminary design features to accommodate human capabilities, limitations, and needs been investigated and recommended, based on the proof of concept?				

Technology Demonstration: The technology is repeatedly demonstrated at increasing levels of fidelity, first in the laboratory and later in relevant environments. This phase concludes with demonstration of a representative deliverable in a high-fidelity simulation or actual environment, with evaluation of human-system interfaces provided by representative users.	4	Modeling, part-task testing, and trade studies of human systems design concepts completed	Human systems design concepts are developed and evaluated in basic laboratory environments, using modeling and part-task testing with low-fidelity prototypes that begin integrating key elements. Trade studies are conducted to analyze and identify the most viable human systems design options.	Have human systems designs been established, based on modeling, part-task testing, and trade studies?	24	Have strategies to mitigate safety implications for human users been identified and recommended?	Human systems experts develop prototypes, analytical models, and mockups of candidate design concepts to support modeling, part-task testing, and trade studies for a variety of human systems evaluations, including all HSI domain analyses, analysis of workload and human errors, and identification of error-prone tasks for mitigation. Strategies to mitigate implications involving any relevant HSI domains are recommended. Updated human-machine function allocations and task analyses are used to optimize task flow and sequencing for maximal human performance. Outcomes of these activities in turn support provision of inputs to begin optimizing procedures for human use. All relevant and applicable human performance data (for both individuals and teams) are evaluated in order to address whether human performance metrics and requirements (including KPPs) are successfully met, based on modeling and part-task testing. A positive response to the human performance question indicates that the relevant data have been collected, evaluated, *and* deemed satisfactory by qualified human systems experts. Human performance results are compared against existing systems to gauge the extent of any improvements and determine whether additional modifications are necessary. Trade studies are used to analyze and identify the most viable user interface design options. Modeling, part-task testing, and trade studies are all used to identify and recommend alternatives strategies (such as engineered, administrative, or PPE controls) to support human use, if needed. Human systems professionals verify that preliminary designs conform to key human performance design principles, standards, and guidance identified as relevant at HRL 2.	Human interactions and performance are evaluated, based on modeling and part-task testing with mockups. Task analyses are updated to optimize task flow and sequencing. Modeling and part-task testing are used to update human-machine allocations. Procedures are undergoing initial human systems evaluation. Strategies to address concerns for any HSI domains deemed relevant are recommended (e.g., training delivery approaches). Alternative strategies (including modified design features) to support human use are identified and recommended, based on modeling, part-task testing, and trade studies.	Document strategies to mitigate concerns related to any relevant HSI domains. [24, 25, 26, 27] Document strategies to mitigate concerns related to maintenance and sustainment. [28] Describe updated human-machine function allocations and provide the rationale. [29] Summarize key observations for procedure development. [30] Document results of updated task analyses and implications for system design. [31] Characterize human interactions and performance, based on modeling and part-task testing with mockups. [32] List the human performance data used during modeling and part-task testing with mockups and evaluate ability to meet metrics and requirements for successful human performance. [32] Describe recommended strategies to support human use. [33] Characterize conformance of preliminary designs to human performance design principles, standards, and guidance. [34] Provide final test plan for human systems modeling and part-task testing with mockups. [N/A] Provide draft test report documenting observations from modeling and part-task testing with mockups. [N/A] Begin developing draft test plan for subsequent human systems testing via prototypes in mission-relevant simulations or actual environments. [N/A]	At TRL 4, key elements of the technology are integrated and can be crudely demonstrated via rapid prototypes in the laboratory. HSI experts can use the prototypes to conduct modeling and part-task testing of human system interfaces. Outcomes of these activities can be used to support trade studies of various user interface design concepts. Note that the term "interface" as used in the HRL scale does not refer only to GUIs. Interface refers to any intersection between the human and the rest of the system. The term is used in the same way it is used for interfaces between technological components of a system, wherein the concern is to ensure the components interface properly to support compatibility and interoperability. The same concern applies to human-system interfaces. The questions addressed at HRL 4 are repeated at HRLs 5 to 8 to evaluate human performance at progressively increasing levels of fidelity (fidelity with respect to the technology, environment, and users).
					25	Have strategies to accommodate manpower, personnel, and training concerns been identified and recommended?				
					26	Have strategies to address environmental constraints and impacts been identified and recommended?				
					27	Have strategies to address other relevant HSI domains been identified and recommended?				
					28	Have strategies to address maintenance and sustainment been identified and recommended?				
					29	Have human-machine function allocations been updated, based on modeling and part-task testing?				
					30	Is modeling and part-task testing being used to design procedures for human user roles throughout the lifecycle?				
					31	Have task analyses been updated to optimize task flow and sequencing, based on modeling and part-task testing?				
					32	Have relevant human performance data been evaluated to determine whether human performance metrics and requirements are successfully met, based on modeling and part-task testing?				
					33	Have strategies to support human use been identified and recommended, based on modeling, part-task testing, and trade studies?				
					34	Has conformance of preliminary designs to human performance requirements, design principles, standards, and guidance been verified?				
	5	User evaluation of prototypes in mission-relevant simulations completed to inform design	Human performance is evaluated via prototypes in mission-relevant simulations or actual environments. The fidelity of key elements has increased significantly, and users participating in testing are independent from the design team. This level represents the first opportunity for more realistic users to interact with the system.	Have design recommendations based on user evaluation of prototypes in mission-relevant simulations been provided?	35	Have strategies to mitigate safety implications for human users been updated, based on prototype testing in mission-relevant simulations?	Human systems experts update the design prototypes for use in mission-relevant simulations or actual environments to support continued human systems evaluations, including all HSI domain analyses. Strategies to mitigate implications involving any relevant HSI domains continue to be undertaken. Prototype testing in mission-relevant simulations is used to evaluate the suitability of human-machine function allocations established at earlier HRL levels and update procedures for human user roles. Task analyses begun in earlier HRL levels are further refined to optimize task flow and sequencing and maximize human performance. Evaluation of all relevant and applicable human performance data (for both individuals and teams) continues in order to address whether human performance metrics and requirements (including KPPs) are successfully met, based on prototype testing in mission-relevant simulations. A positive response to the human performance question indicates that the relevant data have been collected, evaluated, *and* deemed satisfactory by qualified human systems experts. Although subjective metrics are informative, objective metrics of human performance become more critical at this level. Human performance results are compared against existing systems to gauge the extent of any improvements and determine whether additional modifications are necessary. Prototype testing in mission-relevant simulations is used to identify and recommend alternative strategies (such as engineered, administrative, or PPE controls) to support human use. Human systems professionals verify that prototypes conform to key human performance design principles, standards, and guidance identified as relevant at HRL 2.	Human interactions and performance are evaluated in the context of more realistic mission-relevant simulations with higher fidelity and users independent from the design team. Analyses of implications for relevant HSI domains, human-machine allocations, procedures, and task flow and sequencing are updated, based on results from mission-relevant simulations. Alternative strategies (including modified design features) to support human use are identified and recommended.	Document strategies to mitigate concerns related to any relevant HSI domains. [35, 36, 37, 38] Document strategies to mitigate concerns related to maintenance and sustainment. [39] Describe the suitability of human-machine function allocations and document any updates. [40] Summarize key updates for human user procedures. [41] Document updates to task analyses and implications for system design. [42] Characterize human interactions and performance, based on prototype testing in mission-relevant simulations. [43] List the human performance data used during prototype testing in mission-relevant simulations and evaluate ability to meet metrics and requirements for successful human performance. [43] Describe recommended strategies to support human use. [44] Characterize conformance of prototypes to key human performance design principles, standards, and guidance. [45] Provide final test plan for human systems prototype testing in mission-relevant simulations. [N/A] Provide final test report documenting observations from modeling and part-task testing with mockups. [N/A] Provide draft test report documenting observations from human systems prototype testing in mission-relevant simulations. [N/A] Begin developing draft test plan for subsequent human systems testing in high-fidelity simulated or actual environments. [N/A]	At TRL 5, fidelity and integration of key elements has increased significantly, with demonstrations in mission-relevant simulations or environments. A key element at HRL 5 is the increased fidelity of the system simulation and the users participating in prototype testing. They may not be fully representative of the intended users, but they are independent from the design team. This independence permits more informative usability assessments from users who have not been involved in designing the inner workings of the technology. HRL 5 may be the first opportunity for more realistic users to interact within the system.
					36	Have strategies to accommodate manpower, personnel, and training concerns been updated, based on prototype testing in mission-relevant simulations?				
					37	Have strategies to address environmental constraints and impacts been updated, based on prototype testing in mission-relevant simulations?				
					38	Have strategies to address implications for other relevant HSI domains been updated, based on prototype testing in mission-relevant simulations?				
					39	Have strategies to address maintenance and sustainment been updated, based on prototype testing in mission-relevant simulations?				
					40	Has the suitability of human-machine function allocations been determined, based on prototype testing in mission-relevant simulations?				
					41	Is prototype testing in mission-relevant simulations being used to update procedures for human user roles throughout the lifecycle?				
					42	Have task analyses to optimize task flow and sequencing been updated, based on prototype testing in mission-relevant simulations?				
					43	Have relevant human performance data been evaluated to determine whether human performance metrics and requirements are successfully met, based on prototype testing in mission-relevant simulations?				
					44	Have strategies to support human use been identified and recommended, based on prototype testing in mission-relevant simulations?				
					45	Has conformance of prototypes to human performance requirements, design principles, standards, and guidance been verified?				
	6	Human systems design fully matured as influenced by human performance analyses, metrics, prototyping, and high-fidelity simulations	Human performance is evaluated with objective metrics in relevant high-fidelity simulated or actual environments, with a functional prototype and representative users.	Have human performance analyses satisfactorily demonstrated that human systems design is fully matured?	46	Has the range of user scenarios and tasks been tested in high-fidelity simulated or actual environments?	HRL 6 represents the final demonstration before verification and validation activities begin. Therefore, it is important to test the full range of user scenarios and tasks in the high-fidelity simulated or actual environments used at this level, including emergency and non-normal events and the proposed system to track and resolve human systems issues after fielding. Strategies to mitigate implications involving any relevant HSI domains continue to be undertaken. Procedures for human user roles are finalized in preparation for verification and validation. Evaluation of all relevant and applicable human performance data (for both individuals and teams) continues in order to address whether human performance metrics and requirements (including KPPs) are successfully met, based on testing in high-fidelity simulated or actual environments. A positive response to the human performance question indicates that the relevant data have been collected, evaluated, *and* deemed satisfactory by qualified human systems experts. Although subjective metrics are informative, objective metrics of human performance are critical at this level. Testing at this level is also used to identify and recommend alternative strategies (such as engineered, administrative, or PPE controls) to support human use. A key evaluation at HRL 6 is to establish that human-system interfaces are fully mature and ready for verification and validation testing. Minor changes may still be made at later HRL levels, but a fully complete solution to support human performance requirements should be in place at the end of HRL 6. Human systems professionals verify that functional prototypes conform to key human performance design principles, standards, and guidance identified as relevant at HRL 2.	Human interactions and performance are evaluated in the context of high-fidelity simulated or actual environments; with a functional prototype, representative users, and the full range of user scenarios and tasks. Evaluations of a system to track and resolve human systems issues after fielding begin. Analyses of implications for relevant HSI domains are updated, based on results from mission-relevant simulations. Procedures for human user roles are finalized. Alternative strategies (including modified design features) to support human use are identified and recommended.	Describe user scenarios and tasks tested with a functional prototype in a high-fidelity simulated or actual environment. [46] Describe the proposed system that will be used to track and resolve human systems issues after fielding. [47] Document strategies to mitigate concerns related to any relevant HSI domains. [48, 49, 50, 51] Document strategies to mitigate concerns related to maintenance and sustainment. [52] Summarize key updates for human user procedures and document status. [53] Characterize human interactions and performance, based on testing in a high-fidelity simulated or actual environment. [54] List the human performance data used during testing in a high-fidelity simulated or actual environment and evaluate ability to meet metrics and requirements for successful human performance. [54] Describe recommended strategies to support human use. [55] Characterize conformance of functional prototypes to key human performance design principles, standards, and guidance. [56] Provide a final test plan for human systems testing in relevant high-fidelity simulated or actual environments. [N/A] Provide a final test report documenting observations from human systems prototype testing in mission-relevant simulations. [N/A] Provide draft test report documenting observations from human systems testing in relevant high-fidelity simulated or actual environments. [N/A] Begin developing draft test plan for subsequent human systems testing with the final development deliverable. [N/A]	TRL 6 occurs at the end of the technology demonstration phase and represents a major step in a technology's demonstrated readiness. TRL 6 concludes with demonstration of a representative deliverable in a high-fidelity simulation or actual environment, with evaluation of human-system interfaces provided by representative users. After TRL 6, the production and deployment phase begins in order to address final verification, validation, and operational use. During HRL 6, the full range of user scenarios and tasks; including emergency and non-normal events, troubleshooting, and resolution; is evaluated to support full maturation of human systems interfaces. Accordingly, assumptions regarding human performance are thoroughly tested (e.g., operators can respond to alarms within required time and accurately diagnose problems in a timely manner). After HRL 6, user interface modifications are still possible if critical issues or risks are identified during verification and validation testing; however, such changes should be more minor in extent at this point.
					47	Has a system to track and resolve human systems issues after fielding been evaluated in high-fidelity simulated or actual environments?				
					48	Have strategies to mitigate safety implications for human users been updated, based on testing in high-fidelity simulated or actual environments?				
					49	Have strategies to accommodate manpower, personnel, and training concerns been updated, based on based on testing in high-fidelity simulated or actual environments?				
					50	Have strategies to address environmental constraints and impacts been updated, based on testing in high-fidelity simulated or actual environments?				
					51	Have strategies to address implications for other relevant HSI domains been updated, based on testing in high-fidelity simulated or actual environments?				
					52	Have strategies to address implications for maintenance and sustainment been updated, based on testing in high-fidelity simulated or actual environments?				
					53	Is testing in high-fidelity simulated or actual environments being used to finalize procedures for human user roles throughout the lifecycle?				
					54	Have relevant human performance data been evaluated to determine whether human performance metrics and requirements are successfully met, based on testing in high-fidelity simulated or actual environments?				
					55	Have strategies to support human use been identified and recommended, based on testing in high-fidelity simulated or actual environments?				
					56	Has conformance of functional prototypes to human performance requirements, design principles, standards, and guidance been verified?				

Production and Deployment: Final testing, verification, validation, and qualification occur, with human performance evaluations based on representative users. This phase concludes with operational use of the deliverable and continued systematic monitoring of human-system performance.	7	Human systems design fully tested and verified in operational environment with system hardware and software and representative users	Human performance is evaluated with the final development deliverable in an operational environment. Recommended strategies to support human use and resolve human performance issues have been incorporated.	Has operational environment testing and verification demonstrated satisfactory human performance with the established human systems design?	57	Has the range of user scenarios and tasks been tested with the final development deliverable in an operational environment?	At HRL 7, the effectiveness of previously identified strategies to address various HSI domain concerns can be evaluated, given that representative users use the final development deliverable in an operational environment during the full range of user scenarios and tasks. For example, within the HSI domain for environment, impacts of realistic operational environmental conditions such as light, noise, vibration, stress, and anomalous situations on human performance can be evaluated. Within the HSI domains for manpower, personnel, and training; effectiveness of planned user training approaches and manuals can be evaluated. The effectiveness of near-final user procedures can be evaluated. At HRL 7, human systems professionals verify whether human performance metrics and requirements (including KPPs) have been successfully met and whether key recommended strategies to support human use have been incorporated into the final development deliverable. A positive response to the human performance question indicates that the relevant data have been collected, evaluated, *and* deemed satisfactory by qualified human systems experts. Human systems professionals must coordinate with the design team to resolve any high-priority human use issues that have not been satisfactorily addressed; e.g., provide a thorough rationale and explicitly identify the consequences of failure to address the issue. Human systems professionals verify that the final development deliverable conforms to key human performance design principles, standards, and guidance identified as relevant at HRL 2.	Human interactions and performance are evaluated with the final development deliverable in an operational environment during the full range of user scenarios and tasks. The effectiveness of a system to track and resolve human systems issues after fielding is evaluated. The effectiveness of strategies to address any HSI domain concerns is evaluated, based on testing with the final development deliverable in an operational environment. The effectiveness of human user procedures is evaluated. Recommended strategies (including modified design features) to support human use have been satisfactorily incorporated into the final development deliverable.	Describe user scenarios and tasks tested with the final development deliverable in an operational environment. [57] Characterize effectiveness of the proposed system to track and resolve human systems issues after fielding. [58] Analyze the effectiveness of strategies to mitigate concerns related to any relevant HSI domains. [59, 60, 61, 62] Analyze the effectiveness of strategies to mitigate concerns related to maintenance and sustainment. [63] Characterize the effectiveness of human user procedures. [64] List the human performance data used during testing with the final development deliverable in an operational environment and evaluate ability to meet metrics and requirements for successful human performance. [65] Characterize the incorporation of recommended strategies to support human use. [66] Characterize conformance of the final development deliverable to key human performance design principles, standards, and guidance. [67] Provide final test plan for human systems testing with the final development deliverable. [N/A] Provide final test report documenting observations from human systems testing in relevant high-fidelity simulated or actual environments. [N/A] Provide draft test report documenting observations from human systems testing with the final development deliverable. [N/A] Begin developing draft test plan for subsequent human systems qualification testing with the production system. [N/A]	At TRL 7, the development deliverable is very close to the final planned operational system, which represents a significant step beyond TRL 6. TRL 7 typically signals the end of development, at which time the engineering design is essentially frozen. The actual development deliverable (actual system hardware and software) is demonstrated in an operational environment, using the full range of user scenarios and tasks. For the DOD, a TRL 7 demonstration may occur as part of a Joint Concept Technology Demonstration during a planned military command post or field exercise. By this stage in the development process, key strategies that have been identified at previous HRL levels and recommended to support human use or mitigate HSI domain concerns should have been incorporated in the final development deliverable. At HRL 7, the primary task of human systems professionals shifts from identification and recommendation of mitigation strategies to evaluation and verification that key recommended strategies have been successfully incorporated and work as intended to support satisfactory human performance.
					58	Has a system to track and resolve human systems issues after fielding been evaluated with the final development deliverable in an operational environment?				
					59	Has the effectiveness of strategies to mitigate safety implications for human users been evaluated with the final development deliverable in an operational environment?				
					60	Has the effectiveness of strategies to accommodate manpower, personnel, and training concerns been evaluated with the final development deliverable in an operational environment?				
					61	Has the effectiveness of strategies to address environmental constraints and impacts been evaluated with the final development deliverable in an operational environment?				
					62	Has the effectiveness of strategies to address implications for other relevant HSI domains been evaluated with the final development deliverable in an operational environment?				
					63	Has the effectiveness of strategies to address maintenance and sustainment been evaluated with the final development deliverable in an operational environment?				
					64	Have human user procedures been tested with the final development deliverable in an operational environment?				
					65	Have relevant human performance data been evaluated to determine whether human performance metrics and requirements are successfully met, based on testing with the final development deliverable in an operational environment?				
					66	Have recommended strategies to support human use been satisfactorily incorporated into the final development deliverable?				
					67	Has conformance of the final development deliverable to human performance requirements, design principles, standards, and guidance been verified?				
	8	Total human-system performance fully tested, validated, and approved in mission operations, using completed system hardware and software and representative users	Human performance is validated with the production system in a representative environment before the final system is fielded. Any remaining human use issues have been satisfactorily resolved.	Has final production system testing validated satisfactory total human-system performance?	68	Has the range of user scenarios and tasks been tested with the production system in a representative environment?	Qualification testing with the production system represents the final opportunity to identify and incorporate elements to support human readiness before operational use. Human systems professionals can use production system qualification activities to conduct a final check on the error tracking and resolution system, procedures, and the effectiveness of strategies to address HSI domain issues. A positive response to the human performance question indicates that the relevant data have been collected, evaluated, *and* deemed satisfactory by qualified human systems experts. By HRL 8, any remaining human use issues should have been identified and satisfactorily resolved. If any high-priority items or KPPs have not been satisfactorily addressed, human systems professionals must work with the design team to coordinate a mutually agreeable path forward. As a final check, human systems professionals verify that the final production systems conforms to key human performance design principles, standards, and guidance identified as relevant at HRL 2.	Human interactions and performance are evaluated with the production system during the full range of user scenarios and tasks. The effectiveness of a system to track and resolve human systems issues after fielding is evaluated. The effectiveness of strategies to address any HSI domain concerns is evaluated, based on qualification testing with the production system in a representative environment. The effectiveness of human user procedures is evaluated. Recommended strategies (including modified design features) to support human use have been satisfactorily incorporated into the production system.	Describe user scenarios and tasks tested with the production system in a representative environment. [68] Characterize effectiveness of the system to track and resolve human systems issues after fielding. [69] Analyze the effectiveness of strategies to mitigate concerns related to any relevant HSI domains. [70, 71, 72, 73] Analyze the effectiveness of strategies to mitigate concerns related to maintenance and sustainment. [74] Characterize the effectiveness of human user procedures. [75] List the human performance data used during testing with the production system in a representative environment and evaluate ability to meet metrics and requirements for successful human performance. [76] Characterize resolution of any remaining human use issues. [77] Characterize conformance of the final production system to key human performance design principles, standards, and guidance. [78] Provide final test plan for human systems testing with the production system. [N/A] Provide final test report documenting observations from human systems testing with the final development deliverable. [N/A] Provide draft test report documenting observations from human systems testing with the production system. [N/A] Begin developing draft overarching program report for the entire human systems program. [N/A] Provide a draft human systems evaluation plan for continued systematic monitoring after the system is fielded. [N/A]	At TRL 8, certification and validation activities are used to qualify the production system (final system hardware and software) and demonstrate that it will work under expected conditions. For the DOD, a TRL 8 demonstration may occur via developmental test and evaluation (DT&E). Any design changes that are made at this point are the minimum required to bring the system up to specification. If testing is successful at this level, a decision is made to approve the system for production. At HRL 8, human systems professionals evaluate and validate that key recommended strategies to support human use or mitigate HSI domain concerns have been successfully incorporated and work as intended to achieve satisfactory human performance. If any high-priority human usability issues have not been satisfactorily addressed, human systems professionals must coordinate a path forward with the design team.
					69	Has a system to track and resolve human systems issues after fielding been finalized and tested with the production system in a representative environment?				
					70	Has the effectiveness of strategies to mitigate safety implications for human users been evaluated and successfully demonstrated with the production system in a representative environment?				
					71	Has the effectiveness of strategies to accommodate manpower, personnel, and training concerns been evaluated and successfully demonstrated with the production system in a representative environment?				
					72	Has the effectiveness of strategies to address environmental constraints and impacts been evaluated and successfully demonstrated with the production system in a representative environment?				
					73	Has the effectiveness of strategies to address implications for other relevant HSI domains been evaluated and successfully demonstrated with the production system in a representative environment?				
					74	Has the effectiveness of strategies to address implications for maintenance and sustainment been evaluated and successfully demonstrated with the production system in a representative environment?				
					75	Have human user procedures been tested and found acceptable to meet task demands with the production system in a representative environment?				
					76	Have relevant human performance data been evaluated to determine whether human performance metrics and requirements are successfully met, based on qualification of the production system in a representative environment?				
					77	Have human use issues been satisfactorily resolved, as evidenced by qualification of the production system in a representative environment?				
					78	Has conformance of the final production system to human performance requirements, design principles, standards, and guidance been verified?				
	9	System successfully used in operations across the operational envelope with systematic monitoring of human-system performance	The qualified system is fielded in the operational environment and operated by the intended users. Human systems professionals continue to monitor the fielded system. Systematic collection and analysis of human performance issues, errors, and accidents occurs to determine needed enhancements.	Is human performance with the fielded system being systematically monitored?	79	Has the qualified system been fielded in the operational environment for the intended users?	Human systems professionals must be involved after the system is fielded to ensure that the demonstrated human performance capability is sustained. Resolution of emerging human systems issues may involve coordination of hardware, software, or process modifications. Any proposed system upgrades must be analyzed for new human systems impacts that may be introduced. The impacts of changes in user knowledge, skills, and abilities must be evaluated. Training efficacy should be periodically monitored. Lessons learned should be documented and archived to facilitate future efforts on similar systems.	A human systems evaluation plan is in place for continued systematic monitoring of the fielded system. Human systems professionals evaluate the fielded system in the operational environment with actual users. Recommended human systems mitigations are evaluated for implementation.	Document customer acceptance and operational use of the qualified system. [79] Conduct ongoing human systems evaluations of the existing system and potential upgrades through observations, tests, and issue analysis and resolution. [80, 81, 82, 83, 84] Document results and lessons learned, as appropriate. [81] Finalize test report documenting observations from human systems qualification testing with the production system. [N/A] Finalize the overarching human systems program report. [N/A] Finalize human systems evaluation plan for continued systematic monitoring after the system is fielded. [N/A]	At TRL 9, an operational system is released for use. For the DOD, initial operational use may occur in the form of final operational test and evaluation (OT&E). If OT&E is not a requirement, a technology can achieve TRL 9 by performing a live operational mission. Once fielded, the system will continue to undergo systematic monitoring to ensure that the demonstrated human performance capability is sustained via a data collection and feedback loop.
					80	Are potential upgrades to the fielded system being evaluated to address human systems issues and impacts?				
					81	Are human systems performance data and lessons learned being documented for future applications?				
					82	Does the system designed to track and resolve human systems issues in the fielded system support these activities?				
					83	Are human systems mitigations to improve performance in the fielded system being identified and implemented?				
					84	Is user training for operation of the fielded system being evaluated for required modifications?				

Level	HMD Example	TRL and HRL HMD Activities
1		**TRL 1: Basic Principles observed and reported** At TRL 1, a concept or idea for a possible technology emerges. Basic scientific research occurs on paper or in the laboratory. At TRL 1, the idea to replace a head-up display (HUD) located at a fixed position within an aircraft with a display integrated into a helmet might be conceived and explored further to understand whether basic scientific principles support the concept. The concept is very speculative and exploratory. At this point, it is not known whether the concept is viable or not. **HRL 1: Relevant human capabilities, limitations, and basic human performance issues and risks identified** At HRL 1, human systems experts explore human capabilities, limitations, and basic human performance issues and risks. For the HMD, basic human performance issues and risks may include an inability to meet system performance requirements if the helmet does not fit properly or weighs too much. Another concern might be the impact of visual fatigue due to insufficient display resolution. Limitations such as the need for spectacles may be explored to begin identifying the implications of potential compatibility issues for eyeglass wearers.
2		**TRL 2: Concept and application formulated** At TRL 2, rough, fuzzy, and speculative applications are being formulated. Pure research gives way to more applied research. The concept begins to take shape in the form of a potential application, but there is no proof it will work. At TRL 2, the concept of a helmet-mounted display for aircraft missions begins to emerge, but it is still only a concept with no associated hardware yet. Assumptions have not been proven in detail. At this level, it may only be an assumption that the form, fit, and function of a HUD can be effectively altered and adapted for use inside a helmet. At HRL 2, human systems experts research relevant human performance design principles to begin identifying human use requirements and provide inputs for preliminary conceptual designs. For the HMD, anthropometric data for human head measurements are obtained for potentially relevant populations of users such as Air Force fighter pilots. A human performance design principle to accommodate the 5th to the 95th percentile with respect to user head sizes may be established, which generates requirements regarding helmet size and weight.
3		**TRL 3: Concepts demonstrated analytically or experimentally** TRL 3 represents the proof-of-concept stage wherein work moves beyond paper studies and analyses. At this level, active R&D occurs via analysis, modeling and simulation, and laboratory studies to validate the feasibility of key elements of the technology. These elements are not yet representative and not integrated in any way. For the HMD, the design team might validate that the holographic waveguide technology researched in earlier TRLs can successfully relay high-quality source imagery. **HRL 3: Requirements for supporting human performance established** With a proof of concept available at HRL 3, human systems experts can begin mapping anticipated demands of the developing technology against human user capabilities, limitations, and needs. For the HMD, human systems experts might analyze whether the image quality afforded by the holographic waveguide technology will meet human operational needs. The interpretability of proposed HMD symbology might be analyzed to understand congruence with human cognitive needs.
4		**TRL 4: Key elements demonstrated in laboratory environment** TRL 4 begins to bridge the scientific research that occurred at previous levels to the engineering efforts at TRLs 4 through 6. At TRL 4, key elements of the technology are integrated to establish that they can work together. The prototype is very crude, with low fidelity at this point, and may have been assembled with ad hoc hardware that is representative of the final product in function only. The intent is to achieve a "good enough" integration to permit a laboratory demonstration of the integrated elements. A laboratory demonstration of the HMD at TRL 4 may be used to test whether the assembled components can successfully transmit intended imagery and symbology and to understand possible approaches to address any issues that occur. **HRL 4: Modeling, part-task testing, and trade studies of human systems design concepts completed** At HRL 4, human interactions and performance are evaluated, based on modeling and part-task testing with rapid prototypes and mockups. For the HMD, human systems experts might use part-task testing and experimentation to collect accuracy and timing data for a representative task, compare results against existing HUD or HMD systems, and evaluate the effectiveness of the current image quality for user tasks and scenarios. If results suggest metrics for successful human performance may not be met, human systems experts may conduct trade studies to evaluate and recommend HMD design alternatives.
5		**TRL 5: Key elements demonstrated in relevant environments** At TRL 5, both the fidelity of the technology and the fidelity of the environment in which it is demonstrated have increased substantially. The integration of key elements is more realistic and more representative of the final product. At TRL 5, the demonstration may still occur within a controlled laboratory environment, but the conditions more closely approximate the expected operational environment outside the laboratory. At TRL 5, the HMD may be demonstrated during realistic scenarios in a fixed-base laboratory flight simulator that provides simulated imagery inputs. **HRL 5: User evaluation of prototypes in mission-relevant simulations completed to inform design** HRL 5 represents the first opportunity for more realistic users to interact with the system, which has significantly increased in fidelity. For the HMD, human systems experts may evaluate display lag with users independent from the design team during scenario completion in mission-relevant simulations. Key human performance metrics might include mental workload, situational awareness, and user satisfaction, with outcomes used to further inform design and support human usability.
6		**TRL 6: Representative of the deliverable demonstrated in relevant environments** TRL 6 occurs at the end of the technology demonstration phase and represents a major step in a technology's demonstrated readiness. The prototype system is capable of performing all of the functions required in the operational system. Although it is close to the desired configuration in terms of weight and volume, it is not quite fully representative in form. TRL 6 concludes with demonstration of a representative deliverable in a high-fidelity simulation or actual environment. At TRL 6, the HMD may be demonstrated during realistic scenarios in a motion flight simulator that adds another degree of realism. **HRL 6: Human systems design fully matured as influenced by human performance analyses, metrics, prototyping, and high-fidelity simulations** HRL 6 represents the final demonstration before verification and validation activities begin. At HRL 6, evaluations are conducted with representative users during the range of user scenarios and tasks in order to establish the human systems design is fully mature and ready for verification and validation testing. For the HMD, demonstration in a motion simulator provides opportunities to evaluate the effects of vibration on objective metrics such as image quality and display jitter and collect user feedback.
7		**TRL 7: Final development version of the deliverable demonstrated in operational environment** TRL 7 represents the start of the production and deployment phase. The development deliverable is very close to the final planned operational system in both form and function. By the end of TRL 7, the engineering design is essentially frozen. For the DOD, a TRL 7 demonstration in an operational environment may occur as part of a Joint Concept Technology Demonstration (JCTD) during a planned military exercise. At TRL 7, the HMD might be demonstrated and further tested in an exercise like Roving Sands, wherein Army, Marine Corps, and Air Force units simulate combat operations, without the dangers associated with real-world operations. **HRL 7: Human systems design fully tested and verified in operational environment with system hardware and software and representative users** At HRL 7, the primary task of human systems professionals shifts from identification and recommendation of mitigation strategies to support human usability to evaluation and verification that key recommended strategies have been successfully incorporated and work as intended to support satisfactory human performance. For the HMD, humans systems experts might participate in a military exercise like Roving Sands to collect additional data from representative users who use the HMD across a range of scenarios and tasks in a realistic operational environment. The intent is to verify that the final development deliverable supports human usability.
8		**TRL 8: Actual deliverable qualified through test and demonstration** At TRL 8, certification and validation activities are used to qualify the production system and demonstrate that it will work under expected conditions. For the DOD, a TRL 8 demonstration may occur via developmental test and evaluation (DT&E). If testing is successful at this level, a decision is made to approve the system for production. For the HMD, the production helmet is evaluated to determine whether it meets all its design and operational requirements. At that point, any design changes are limited to those necessary to bring the HMD up to specification. **HRL 8: Total human-system performance fully tested, validated, and approved in mission operations, using completed system hardware and software and representative users** At HRL 8, final production system testing is used to validate satisfactory total human-system performance. This level represents the final opportunity to identify and incorporate elements to support human readiness before operational use. If any high-priority items have not been satisfactorily addressed, human systems experts coordinate a mutually agreeable path forward with the design team. For the HMD, human systems professionals evaluate the final production helmet to validate that metrics for successful human-system performance have been met.
9		**TRL 9: Operational use of deliverable** At TRL 9, an operational system exists. The technology is applied in its final form and used operationally. For the DOD, initial operational use may occur in the form of operational test and evaluation (OT&E) to determine operational effectiveness and suitability. TRL 9 does not include ongoing or planned product improvement of reusable systems. To achieve TRL 9, the HMD might undergo final testing during OT&E, or it might be used in a live operational mission. **HRL 9: System successfully used in operations across the operational envelope with systematic monitoring of human-system performance** At HRL 9, the system is used in its intended operational environment by the intended users to successfully accomplish the required mission. Human systems experts periodically evaluate human-system performance, using the metrics established throughout design and development, to determine whether modifications are needed and to document lessons learned. For the HMD, human systems experts might be consulted as part of a team to troubleshoot and resolve intermittent degradations in night vision acuity.

Appendix B. TRL

TRL 1 – Basic principles observed

TRL 2 – Technology concept formulated

TRL 3 – Experimental proof of concept

TRL 4 – Technology validated in lab

TRL 5 – Technology validated in relevant environment (industrially relevant environment in the case of key enabling technologies)

TRL 6 – Technology demonstrated in relevant environment (industrially relevant environment in the case of key enabling technologies)

TRL 7 – System prototype demonstration in operational environment

TRL 8 – System complete and qualified

TRL 9 – Actual system proven in operational environment (competitive manufacturing in the case of key enabling technologies; or in space)

Appendix C. HSI domains

HSI domain	Definition
Manpower	The number and mix of personnel (military, civilian, and contractor) authorized and available to train, operate, maintain, and support each system acquisition.
Personnel	The human aptitudes, skills, knowledge, experience levels, and abilities required to operate, maintain, and support the system at the time it is fielded and throughout its life cycle.
Training	The instruction and resources required to provide personnel with requisite knowledge, skills, and abilities to properly operate, maintain, and support the system.
Human Factors Engineering	The comprehensive integration of human capabilities and limitations (cognitive, physical, sensory, and team dynamic) into system design, development, modification and evaluation to optimize human-machine performance for both operation and maintenance of a system. Human factors engineering designs systems that require minimal manpower, provide effective training, can be operated and maintained by users; and are suitable and survivable.
Environment	The factors concerning water, air, and land and the interrelationships which exist among and between water, air, and land and all living things.
Safety	The design and operational characteristics that minimize the possibilities for accidents or mishaps to operators which threaten the survival of the system.
Occupational Health	The design features that minimize risk of injury, acute and/or chronic illness, or disability, and/or reduced job performance of personnel who operate, maintain, or support the system.
Survivability	The characteristics of a system that reduce risk of fratricide, detection, and the probability of being attacked; and that enable the crew to withstand manmade or natural hostile environments without aborting the mission or suffering acute and/or chronic illness, disability, or death.
Habitability	Factors of living and working conditions that is necessary to sustain the morale, safety, health, and comfort of the user population which contribute directly to personnel effectiveness and mission accomplishment, and often preclude recruitment and retention problems.

Note. By Phillips, 2010, p. 4-5

Appendix D. Overview HRL Phillips

HRL	Description
1. Activation of Human Systems Integration: base-lining and commitment.	Lowest level of socio-technical readiness. HSI infrastructure is setup within Systems Engineering planning. Total system analysis from both functional relationship and organizational perspectives occurs. Activity examples include front-end analyses, preliminary functional allocation, and initial HSI assessment and plan.
2. Human Systems Integration analysis suite in support of component technology development.	Significant HSI input to acquisition development and documentation occurs. Activity examples include initial human-machine interface assessment, generation of Target Audience Description, and threat/hazard assessment.
3. Component human touch point resolution (i.e., human-system interaction, to include hardware and software): refining requirements thresholds.	Multiple needs analyses and studies are conducted in support of requirement definitions. HSI domain assessments inform ongoing development actions, as well. Activity examples include human-in-the-loop analyses, sub-system hazard analysis, and HSI plan update and revision.
4. Component human touch point engineering parameters and human performance indicators.	Iterative evaluation and analysis of each HSI domain takes place and provides critical items of consideration bearing on system design. Activity examples include usability testing, development of human-centered source selection, and updating the human-machine interface assessment.
5. Limited system human performance parameters/demonstration.	Various HSI assessments and testing are performed to support the significant increase in fidelity of breadboard technology. This includes supporting “high fidelity” laboratory integration of components. Activity examples include examining safety and occupational health design features and cognitive task analyses.
6. Field (relevant environment) validation of human performance prototypes.	Representative model or prototype system is tested in a relevant environment. Evaluations of human performance embedded in demonstration system occur and HSI predictive models are updated. Activity examples include testing human reliability and usability of prototype in a high-fidelity laboratory environment or in simulated operational environment.

7. Final Developmental Test & Evaluation/human performance parameters.	Significant HSI participation in system test events occurs. Iterative evaluation and analysis of each HSI domain continues as well. Activity examples include error and fault analysis to cover human error performance, equipment operability, safety procedures, and error recovery mechanisms.
8. Operational Test & Evaluation/human performance parameters.	Special human-centric analyses are conducted to update thresholds, objectives, and evolving criteria for Operational Test & Evaluation. Iterative evaluation and analysis of each HSI domain continues as well. Activity examples include system hazard analysis and HSI domain tradeoff studies.
9. Sustainment: Initiation of capability gap feedback cycle.	Extensive and iterative review and verification of fielded system begins, as well as post-product improvement evaluations for the next incremental builds. Activity examples include post-fielding training evaluation analysis and sustaining a hazard analysis for the fielded system.

Note. By Phillips, 2020, p. 34-36

Appendix E. HF Anchors

Allocation of Function	Assigning those roles/functions/tasks for which the human or equipment performs better while enabling the human to maintain awareness of the operational situation.
Anthropometrics and Biomechanics	Accommodating the physical attributes of its user population (e.g., from the 1st through 99th percentile levels).
Computer-Human Interaction	Employing effective and consistent user dialogues, interfaces, and procedures across system functions.
Communications and Teamwork	Applying system design considerations to enhance required user communications and teamwork.
Culture	Addressing the organizational and sociological environment into which any change, including new technologies and procedures, will be introduced.
Displays and Controls	Designing and arranging displays and controls to be consistent with the operator's and maintainer's tasks and actions.
Documentation	Preparing user documentation and technical manuals in a suitable format of information presentation, at the appropriate reading level, and with the required degree of technical sophistication and clarity.
Environment	Accommodating environmental factors (including extremes) to which the system will be subjected and understanding the associated effects on human-system performance.
Functional Design	Applying human-centered design for usability and compatibility with operational and maintenance concepts.
Human Error	Examining design and contextual conditions as causal factors contributing to human error, and consideration of objectives for error tolerance, error prevention, and error correction/recovery.
Information Presentation	Enhancing operator and maintainer performance through the use of effective and consistent labels, symbols, colors, terms, acronyms, abbreviations, formats, and data fields.
Information Requirements	Ensuring the availability and usability of information needed by the operator and maintainer for a specific task when it is needed, and in a form that is directly usable.
Input/Output Devices	Selecting I/O methods and devices that allow operators or maintainers to perform tasks, especially critical tasks, quickly and accurately.
Knowledge, Skills and Abilities	Measuring the KSAs required to perform job-related tasks, and determining appropriate selection requirements for users.

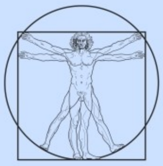
Operational Suitability	Ensuring that the system appropriately supports the user in performing intended functions while maintaining interoperability and consistency with other system elements or support systems.
Procedures	Designing operation and maintenance procedures for simplicity, consistency, and ease of use.
Safety and Health	Preventing/reducing operator and maintainer exposure to safety and health hazards.
Situational Awareness	Enabling operators or maintainers to perceive and understand elements of the current situation, and project them to future operational situations.
Special Skills and Tools	Minimizing the need for special or unique operator or maintainer skills, abilities, tools, or characteristics.
Staffing	Accommodating constraints and efficiencies for staffing levels and organizational structures.
Training	Applying methods to enhance operator or maintainer acquisition of the knowledge and skills needed to interface with the system, and designing that system so that these skills are easily learned and retained.
Alerts	Designing visual and auditory alerts to invoke the necessary operator and maintainer response.
Workload	Assessing the net demands or impacts upon the physical, cognitive, and decision-making resources of an operator or maintainer using objective and subjective performance measures.
Work Space	Designing adequate work space for personnel and their tools or equipment, and providing sufficient space for the movements and actions that personnel perform during operational and maintenance tasks under normal, adverse, and emergency conditions.

Note. Based on Krois and Rehmann, 2005, p. 420-421

Appendix F. HRL Factsheet

HRLs Explained

HRL Working Group Core Team



Proposed HRLs would:

- Complement and supplement existing TRLs
- Focus on readiness for human usability
- Fully incorporate human element of a system throughout the lifecycle
- Minimize costs of design changes by early identification of human issues
- Mitigate issues early to reduce human error in fielded systems
- Distill results from detailed HSI methods
- Provide leading indicators of human readiness
- Use a single number to communicate human readiness

HRLs are described here with the example of a ***helmet-mounted display*** used to project information to aircrew members' eyes and enable tasks such as aircraft control, targeting, and weapon cueing.

The HRL working group was established in August 2019 to mature the HRL scale and test utility in a range of scenarios, leveraging inputs from 20+ members representing the broader HSI community throughout the DOD, DOE, other federal agencies, industry, and academia.

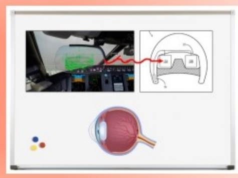
Basic Research and Development

HRL 1

High — Human Performance Risk — Low

Definition

Relevant human capabilities, limitations, and basic human performance issues and risks identified.



What That Means

This first level of human readiness represents a broad, high-level exploration of human involvement for the developing concept or proposed practical application.

Helmet-Mounted Display Example

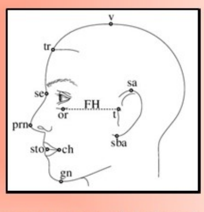
Basic human performance issues and risks include poor fit, excessive weight, or visual fatigue due to insufficient display resolution.

HRL 2

High — Human Performance Risk — Low

Definition

Human-focused concept of operations defined and human performance design principles established.



What That Means

Implications for human use are analyzed. Human performance design principles inform human use requirements and preliminary conceptual designs.

Helmet-Mounted Display Example

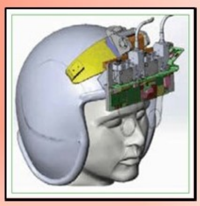
Anthropometric data for human head measurements are obtained for the 5th to the 95th percentiles in relevant user populations.

HRL 3

High — Human Performance Risk — Low

Definition

Analyses of human operational, environmental, functional, cognitive, and physical needs completed, based on proof of concept.



What That Means

Human systems evaluations of the proof of concept are used to map human user needs to expected system demands.

Helmet-Mounted Display Example

The image quality afforded by the developing technology is analyzed to determine if it meets human operational and functional needs.

SAND2019-15036 O

Technology Demonstration

HRL 4

High — Human Performance Risk — Low

Definition
Modeling, part-task testing, and trade studies of user interface design concepts completed.

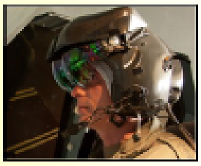
What That Means
User interface design options are analyzed via trade studies and evaluated in basic laboratory environments.

Helmet-Mounted Display Example
Accuracy and timing data are collected for representative tasks to evaluate effectiveness of current image quality and gauge the extent of any improvements over legacy systems.

HRL 5

High — Human Performance Risk — Low

Definition
User evaluation of prototypes in mission-relevant simulations completed to inform design.



What That Means
The fidelity of key elements has increased significantly, and users participating in testing are independent from the design team.

Helmet-Mounted Display Example
Display lag is evaluated in a fixed-base laboratory flight simulator, using human performance metrics such as mental workload, situational awareness, and user satisfaction.

HRL 6

High — Human Performance Risk — Low

Definition
Human-system interfaces fully matured as influenced by human performance analyses, metrics, prototyping, and high-fidelity simulations.

What That Means
Human performance analyses with a functional prototype are used to evaluate whether human-system interfaces are fully matured.

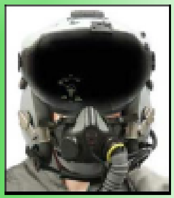
Helmet-Mounted Display Example
The effects of vibration on objective metrics such as image quality and display jitter are demonstrated in a motion flight simulator.

Production and Deployment

HRL 7

High — Human Performance Risk — Low

Definition
Human-system interfaces fully tested and verified in operational environment with system hardware and software and representative users.



What That Means
The final development deliverable is evaluated to verify whether metrics for successful human performance have been met.

Helmet-Mounted Display Example
Data are collected during a military exercise, with representative users and a range of HMD scenarios and tasks in a realistic operational environment.

HRL 8

High — Human Performance Risk — Low

Definition
Total human-system performance fully tested, validated, and approved in mission operations, using completed system hardware and software and representative users.

What That Means
Human performance is validated with the production system before the final system is fielded.

Helmet-Mounted Display Example
The final production helmet is evaluated to validate that metrics for successful human performance have been met, possibly during DT&E.

HRL 9

High — Human Performance Risk — Low

Definition
System successfully used in operations across the operational envelope with systematic monitoring of human-system performance.

What That Means
The qualified system is fielded in the operational environment and operated by the intended users.

Helmet-Mounted Display Example
After the HMD is fielded, human systems experts are consulted to troubleshoot and resolve intermittent degradations in night vision acuity.

For more information, contact Judi See (jesee@sandia.gov), Holly Handley (hhandley@odu.edu), and Mike O'Neil (michael.oneil@nps.edu)

Appendix G. Elaborated description of comparison between projects and HRL v1

ALICIA

Together with 40 European partners, Deep Blue participated in ALICIA (CORDIS, 2019a). ALICIA is a project that ran from 2009 to 2014 and addressed the aim of increasing time efficiency within the future Air Traffic Management (ATM) system. In order to contribute to this objective, the research project focused on developing innovative cockpit applications that improve operations of an aircraft under degraded (weather) conditions, or as they called it: ‘All Condition Operations’. The project has been labelled as confidential, which was still a common practice at the time (S. Bonelli, personal communication, January 18, 2021). Consequently, details about the results of the project cannot be shared. However, this will not pose a problem as we are interested in the validation strategy rather than the outcomes. The documents that will be used (i.e. S. Bonelli, personal communication, November 19, 2020; S. Bonelli, personal communication, January 7, 2021; Diehl Aerospace et al., 2010) are not accessible, but several dissemination materials can be found online (Aliciaprojectvideo, 2011; CORDIS, 2019b).

In ALICIA, several developed technologies were integrated and tested in a simulation exercise: The Global System Evaluation. The technologies were supplied by the partners and were at different stages of development. The rationale behind ALICIA is that a lot of new concepts, such as technologies that aim to contribute to an aircraft’s All Conditions Operations capability, will be introduced within the next 15 years and that the introduction of these technologies would be very costly if it would be done according to the ‘classical’ approach (S. Bonelli, personal communication, January 7, 2021). Rather than developing individual technologies and integrating them into the flight deck design one by one, ALICIA bundles technologies into new flight-deck core concepts and tests to what extent these can facilitate a more expansive range of operational conditions. The approach in ALICIA is very interesting for the HRL as the HRL starts from a concept and then supports the development of technologies that subsequently have to be integrated and validated as one system. In ALICIA, a concept had to be formed based on the technologies that had already been (partly) developed. It is unclear whether and how the HRL could be applied in such a situation

Before the Global System Evaluation, the technologies were validated individually by the different partners to ensure they would be ready for integration. Because of the complexity of the project's validation activities, it was decided that a structured approach was needed for the development of the validation plan (S. Bonelli, personal communication, January 7, 2021). This is good news for the HRL as it means there is an expressed need for a framework that can help structure the validation process. The validation in ALICIA was based on the E-OCVM (S. Bonelli, personal communication, January 7, 2021). This framework contains three elements that, together, structure the operational concept development and validation. The first element, the CLM, is a model that tracks the advancement of a system throughout the different phases of development, called V-phases, from a concept up to a fully operational system. Secondly, the Structured Planning Framework (SPF) structures the activities that need to be carried out in each CLM V-phase. Finally, the Case-Based Approach (CBApp) is there to merge supporting evidence from multiple CLM V-phases in order to create a consistent source of information concerning areas of particular interest to the stakeholders, such as safety, cost and benefits, and HF. Mainly the SPF was used in ALICIA to develop the validation plan (S. Bonelli, personal communication, January 7, 2021). The SPF consists of six steps related to planning, executing and reporting validation activities on both project and programme levels:

1. Identify the acceptance criteria and performance requirements for the exercise;
2. Refine the evaluation objectives;
3. Refine evaluation exercise requirements;

4. Identify indicators and metrics;
5. Develop the evaluation scenarios;
6. Produce the evaluation exercise plan.

There are several noticeable similarities with the HRL. Specifically, HRL sub-questions 15 and 21 focus on the identification of performance requirements, sub-question 11 also demands the identification of metrics, sub-questions 2 and 12 ask for the development of usage scenarios, and the HRL also requires the development of a test plan (see Supporting Evidence for HRL 3-8). Although the HRL addresses systems requirements to support HP, it does not include acceptance criteria and exercise requirements.

Apart from the E-OCVM structure, there was no further guidance from SESAR at the time,f and there was no established framework to support validations (S. Bonelli, personal communication, January 18, 2021). A Common Testing Framework (CTF) was created to ensure that all evaluation partners would be on the same wavelength considering the evaluation of their prototype. The CTF decomposes the project objectives into lower-level goals and Key Performance Indicators (KPIs) that can be directly tested (S. Bonelli, personal communication, February 23, 2021). Based on the maturity level of the prototype, the CTF advises which evaluation methods could potentially be used. An additional document provided an overview of the current TRL of every technology (Diehl Aerospace et al., 2010). It is interesting to note that there were also some references to the impact of HF on readiness in this document. Take, for example, this extract, saying that certain concepts "are technically possible but have not been developed beyond the prototype stage (and therefore are currently somewhere between TRL levels 3 and 5), mainly due to HF issues … rather than technical ones" (Diehl Aerospace et al., 2010, p. 33).

The HRL could have been applied here to indicate better the discrepancy between the readiness of the concept in a technical versus operational sense.

After the evaluation of the different technologies, synthetic environments and bench testing were used to test the feasibility of the integrated concept (S. Bonelli, personal communication, January 7, 2021). Not all technologies were included in the Global System Evaluation, as some were not mature enough (S. Bonelli, personal communication, January 18, 2021). Before the demonstration, HMI components were identified so that the validation exercises could focus on these as they were expected to have the most significant impact on pilots' tasks. Furthermore, this information was used to help the HF experts prepare for the validation and assemble training material for the pilots. Next to the All Conditions Operation, there was a baseline configuration, namely a cockpit with legacy functionalities (S. Bonelli, personal communication, November 19, 2020). The latter functioned as a benchmark against which the new layout could be compared to gauge relative performance. In the HRL, there is no specific mention of the use of benchmarks. It could be part of conducting trade studies (HRL 4). In HRL 2, legacy and comparable systems have to be analysed (sub-question 9), but this is meant to support the understanding of key interactions between human and machine rather than to assess the benefits of the proposed solution compared to the benchmark. In HRL 2, practitioners should also start thinking about appropriate metrics for successful HP. In order to define these, the performance of a legacy system could be used as a benchmark.

Task performance was defined in relation to socio-technical elements, including flight context, task environment, task functions, task procedures, task activity, team, information (flow), and process. Additionally, elements such as regulation and the airline's objectives, business model and safety culture were explored. Whereas the HRL also includes sub-questions regarding tasks (sub-questions 8, 13, 42, 46, 57, 68, 75), it does not address information flow and team factors, nor does it concern commercial objectives and organisational culture. In general, it could be stated that the HRL focuses a great deal on the task and not so much on the context or environment in which the tasks have to be conducted

The methods used in ALICIA were a mixture of qualitative and quantitative methods (S. Bonelli, personal communication, January 7, 2021). Quantitative data included task performance times, number of errors, number of events detected, and ratings of workload (e.g. NASA TLX), situation awareness (e.g. SART), and system usability (e.g. SUS). The qualitative analysis focused on interactions between Liveware and Liveware, Software and Hardware, as described by the SHEL model. ‘Liveware’ stands for the human in the system. Qualitative methods included interviews, observation, debriefings, and heuristic analysis (S. Bonelli, personal communication, January 7, 2021). The HRL does not mention communication or any other dependencies between operators, so it seems to focus on the interactions between Liveware and Software/Hardware. In line with HRL sub-questions 2, 12 and 46, usage scenarios were designed and subsequently tested in a scenario-based simulation (S. Bonelli, personal communication, November 19, 2020). These were used to evaluate the benefits brought by the ALICIA technologies. For the structuring of the scenarios, a scenarios framework, called the Supply Context Organisation Process Effect (SCOPE) framework, was used. The framework helps to capture and analyse operations systematically, in line with the SHEL model. It starts with ‘setting the scene’: identifying *who*, *what information* and *which tools* are involved. Subsequently, it says to define (1) what competencies, procedures and interactions are supporting the action, (2) what happens within the process and between the process and the broader system, and (3) what environmental, organisational and personal factors influence the action. It ends with looking at the outcome in terms of short term as well as long term results. This framework could be used to answer the HRL questions about usage scenarios. Interestingly, the scenarios were only developed after the technologies were evaluated individually by the partners and ready to be included in the Global System Evaluation. Furthermore, Work Package (WP) 3 focused on technical evaluations, WP 4 on functional evaluations and WP 5 on operational evaluations (S. Bonelli, personal communication, January 7, 2021). Thus, while hardware and software received attention in WP 3 and 4, HF were attended only later in WP 5 (S. Bonelli, personal communication, January 18, 2021). Despite the evaluation exercises based on the CTF completed before the Global System Evaluation, there was little attention for HF, also evidenced by the fact that Deep Blue, functioning as the HF evaluation specialist in this project, was only invited to intervene in one of the preliminary evaluations. This emphasises again the need for more awareness about the early inclusion of HF in systems development.

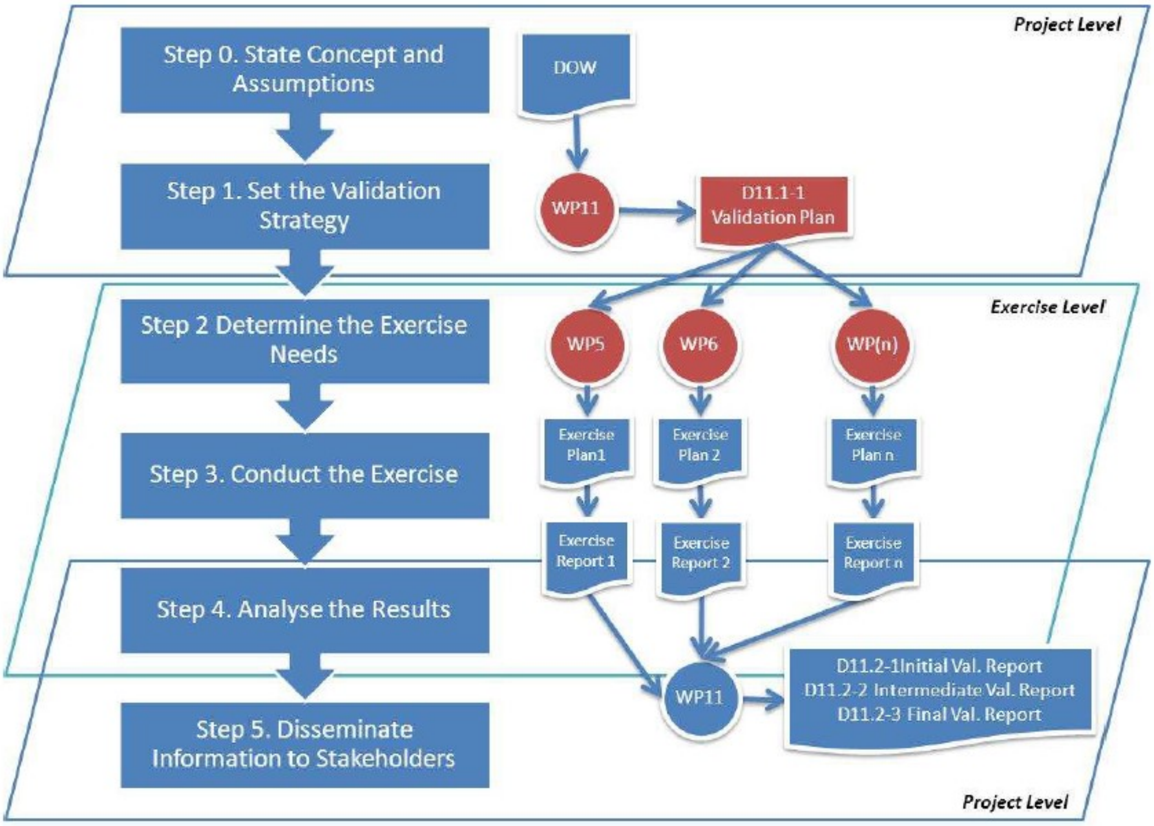
ACROSS

The project Advanced Cockpit for Reduction Of StreSs and workload (ACROSS) ran from 2013 until 2016 (CORDIS, 2019c). In total, it had three objectives (Billiere et al., 2013). First of all, it aimed at supporting the management of peak workload situations. Secondly, its goal was to allow for reduced crew operations. Lastly, the aim was to identify open issues that stand in the way of the successful implementation of single-pilot operations.

Although the *topic* is different from that of ALICIA, the ACROSS project had a very similar *structure*. For example, the validation in ACROSS was also based on the E-OCVM. Furthermore, just like ALICIA, ACROSS integrated different technologies into an operational concept, called the 'Global Operational Concept' (P. Lanzi, personal communication, January 8, 2021). In Figure G1, the validation process of ACROSS is shown (Billiere et al., 2013). The figure shows how the ACROSS deliverables relate to the SPF steps. In the DOW, the scope of ACROSS was defined. WP11 produced the validation plan and report. The research and development activities were grouped in four WPs, each corresponding to one of the main task categories: Aviate/Fly, Navigate, Communicate, and Manage. Two additional WPs corresponded to flight crew incapacitation and monitoring, corresponding to new systems to be developed by ACROSS.

Figure G1

Validation Process in ACROSS



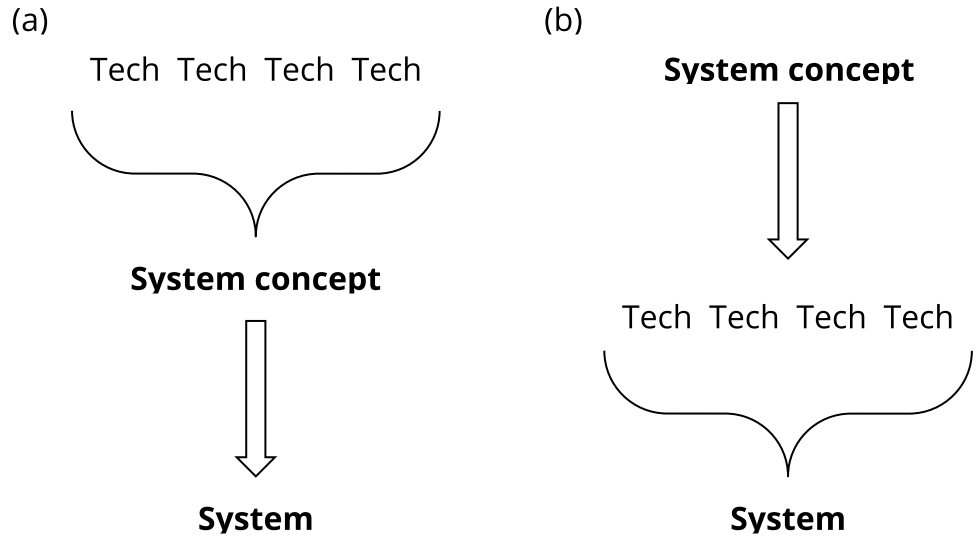
Note. By Billiere and colleagues, 2013, p. 18

Although the E-OCVM was followed more precisely in ACROSS than in ALICIA, the CBApp was not applied. This is related to the fact that ACROSS technologies were relatively mature (in a technical sense) compared to those used in ALICIA, but the project still fell within the exploratory rather than industrial research phase, and so it was allowed more degrees of freedom (P. Lanzi, personal communication, January 8, 2021). SESAR distinguishes fundamental exploratory research and applications-oriented exploratory research, and ACROSS is an exemplar of the latter.

Four operational situations, or ‘configurations’, were developed to work towards the project objectives: (1) fully capacitated flight crew under peak workload conditions, (2) intentionally reduced flight crew (with one pilot at rest), (3) partial flight crew incapacitation, and (4) full flight crew incapacitation (Billiere et al., 2013). So, in line with the HRL, usage scenarios were defined. At the start of the project, current and target maturity, expressed in both CLM V-phases and TRLs, were determined for each of the configurations and at the project level. Although the technical maturity of the individual technologies was relatively high, the Global Operational Concept was at a much lower level in terms of maturity (P. Lanzi, personal communication, January 8, 2021). Initially, the project was supposed to work with the IRL. However, it was soon realised that combining the technologies could lead to potential confusion for the operator. For example, the same category of action could be initiated with a green button on one technology but with a violet button on another, potentially leading to confusion and consequent errors. Just like ALICIA, this project shows there is a need for structuring a complex validation process, something for which the HRL is particularly suitable. Meanwhile, the project also highlights again that HF or operational concerns often get to play a more active role later in the development cycle (after individual technologies have already been developed). Whereas the HRL assumes that system development starts from one concept, branches off in several technologies or technological components, and later integrates these to form the system as it was envisioned, both ALICIA and ACROSS started from many separate technologies for which a system concept had to be defined and evaluated. See Figure G2 for an illustration of this.

Figure G2

Visual representation of development process (a) in ALICIA and ACROSS versus (b) in the HRL. In the representation of the left, several technologies ('Tech') are combined to form one system concept, that is then developed into an actual system. In the representation on the right, a system concept is defined and, subsequently, several technologies are developed and integrated to form the system envisioned at the beginning.



The validation objectives were divided into technical and operational validation objectives (Billiere et al., 2013). The first category focused entirely on system performance, whereas the latter considered the impact of the proposed solution on three main performance areas: cost-benefits, HP, and safety. This structure is in line with the division between TRL (technical) and HRL (operational), although HRL focuses more specifically on HP and does not explicitly address cost benefits. Although Safety and HP form separate areas, they are often intertwined. In the context of ACROSS, Safety focuses more on hazards and risks, whereas HP focuses on the following three dimensions:

1. The operational impact of procedures, roles, responsibilities, and working methods;
2. Interactions between the human and the system;
3. Team structure and communication.

All validation objectives were captured in a table with expected evidence, validation activities, validation methods, and key performance indicators. The key performance indicators are an example of identifying *metrics*, as included in the HRL (sub-question 11). In the HRL, there is currently no sub-question about the identification of methods, nor does it offer examples of methods that could be used. To get a better idea of the validation in ACROSS and how it relates to the HRL, we have a closer look at the validation objectives.

The validation objectives concerning procedures, roles, responsibilities, and working methods capture elements such as the acceptability and HP impact of changes introduced in roles and proposed operating methods. The HRL treats user roles too (sub-question 8, 13, 30, 41, and 53). Operating methods are more abstract and general than tasks and include user activities, responsibilities and roles (P. Lanzi, personal communication, January 8, 2021). They can be compared to procedures, which are also mentioned in the HRL (sub-question 30, 41, 53, 64, and 75). The HRL also includes specific questions related to tasks (sub-question 8, 13, 31, 42, 57, and 68). HP has been split into workload, situational awareness, task execution, and human error (Billiere et al., 2013). Methods used to assess these are interviews, the Instantaneous Self-Assessment of workload questionnaire (ISA), NASA-TLX, debriefings, observations, data recording and the Situation Awareness for SHAPE tool. In ACROSS, both acceptability for the operator and impact on HP were considered because the impact on HP may be low while acceptability is also low (P. Lanzi, personal communication, January 8, 2021). A telling example of this can be found in the literature (Wright & O'Hare, 2015). Wright and O'Hare evaluated a glass cockpit display and found that performance was lower when the glass cockpit display was used than the traditional display. However, participants had a preference for the glass cockpit display and also associated it with higher performance. The HRL does not mention acceptability and seems to focus on the impact on HP instead.

The validation objectives concerning human-system interactions focused on the acceptability of human-system task allocation, the impact of system performance on HP and the usability of new system functionalities (Billiere et al., 2013). The HRL also covers the first one as it includes sub-questions about human-machine function allocations (sub-questions 14, 29, and 40). However, ACROSS is more specific and looks at workload, situational awareness, trust in automation, and easiness to build a correct mental model. The second validation objective compares information needs (of the human) with the information provided (by the system). In the HRL, there are sub-questions regarding the strategies to support human use and the conformance of preliminary design to HP requirements but not specifically related to information presentation, and HP is not defined as detailed as in ACROSS. For the usability validation objective, the system was divided into components, namely: input devices, output devices and alarms. The system was validated as a unique instance, but it was classified into these components of interest to understand better the interaction between the system and the operator and to support the analysis of data much like hypotheses are used in scientific experiments (P. Lanzi, personal communication, January 8, 2021). This is comparable to ALICIA, where the validation was focused on system elements that interacted with the operator. Next to questionnaires, debriefings, and observations, also heuristic evaluation and cognitive walkthrough were used (Billiere et al., 2013).

The validation objectives concerning team structure and communication focused on the acceptability and impact of changes in the team structure and communication between team members. Next to the methods mentioned so far, gaming techniques were also introduced to collect HP data to achieve the validation objectives. The HRL does not touch upon team structure and communication apart from function descriptions and allocations (e.g. sub-question 13 and 14).

In addition to the HP validation objectives regarding roles and responsibilities, human-system interaction, and team structure and communication, there are some validation objectives that are categorised as Safety and Cost-Benefit but are also included in the HRL. For example, one Safety validation objective is to identify possible hazards. HRL sub-questions 3, 10, 24, 35, 48, 59 and 70 likewise stimulate the identification of potential human systems issues. However, the HRL does not include the assessment of the frequency and severity of potential risks. In ACROSS, methods such as focus groups and HAzard and

OPerability study (HAZOP) are used to this end. The validation objectives labelled as Cost-Benefit or Transitional factors focus on identifying implications of technological solutions on competence requirements, potential regulatory issues and security concerns. Competence requirements concern knowledge, skill and experience, operator licensing, and training needs. These overlap with the Manpower, Personnel and Training sub-questions (Q17, Q25, Q36, Q49, Q60, Q71) of the HRL. In ACROSS, training needs were part of the Cost-Benefit area because there is a trade-off between HP, costs and training (P. Lanzi, personal communication, January 8, 2021). Regulatory as well as (cyber) security aspects fall outside of the HRL's scope. In conclusion, the HRL mostly overlaps with the HP area, although it misses some aspects such as communication. On the other hand, the HRL does include some aspects that fall in the areas of Safety and Cost-Benefit in the validation structure of ACROSS, such as training and potential risks. Furthermore, whereas the HRL focuses on human-system *performance*, ACROSS considers it essential to take into account the *acceptance* of the developed concepts to the stakeholder communities (Billiere et al., 2013). At the start of the project, it was determined what expertise was missing in the ACROSS consortium so that relevant stakeholders could be identified and included in an External Experts Advisory Group. Stakeholders had multiple roles in the project. During the initial phases of the project, the stakeholders participated in meetings and provided information that was used to identify needs. In later project phases, stakeholders were involved in Part Task evaluations and provided their feedback via questionnaires and facilitated group sessions. Typically, an accident results from a system failure rather than simply an operator's error, so it is crucial to identify all the decision-makers who may be involved in preparing the condition in which an accident may occur.

Interestingly, in ACROSS, the partners used Validation Assumptions, following the E-OCVM (Billiere et al., 2013). The ACROSS consortium partners responsible for the validation activities could identify and propose validation assumptions depending on their needs. An assumption is “a proposition that is taken for granted, as if it were true, for the purposes of performing demonstrations or assessments in specific context” (Billiere et al., 2013, p. 150). These validation assumptions had to be communicated to the WP leader to ensure they would remain traceable. The HRL v1 framework does not mention the identification of assumptions. The validation assumptions could be an essential link between the HRL and the TRL as they often pertain to a technical capability.

NINA

NINA stands for Neurometric INdicators for ATM and is composed of two strands of work (Terenzi et al., 2014). The first strand of work explored how the cognitive state of an air traffic controller (ATCO) can be described based on the monitoring of several physiological parameters. The second strand of work explored how adaptive HMI solutions could be triggered by the cognitive states of the ATCO. In total, the project lasted a bit over one year, starting in 2014.

The measurements corresponding to the first strand of work took place in a laboratory setting and a simulated ATM environment, with actual ATCOs. As part of the simulation, ATCOs were asked to fill in several questionnaires (ISA, NASA-TLX) and participate in debriefings. Meanwhile, direct over-the-shoulder but non-intrusive observation took place to collect insights about the ATCOs performance, including aspects like workload, stress, drowsiness, satisfaction, difficulties faced, recovery actions, and safety-related events. Additionally, log files were extracted and used as performance indicators. An example of the quantitative data collected by the prototyping platform is the ATCO’s order input. Next to being able to detect different workload levels, the physiological measures also aimed to assess the improvement resulting from training over a medium versus long period, and to identify the main cognitive features that characterised skilled ATCOs so that these could be used to indicate when an ATCO student has reached a sufficient level of training.

The second goal of NINA was to develop early prototypes of adaptive solutions. These solutions were based on HF considerations and guidelines (M. Terenzi, personal communication, January 8, 2021). More specifically, the HF pie was used. The HF pie contains the following HF categories: Human in the System, Working Environment, Teams & Communication, Training & Development, Organisation & Staffing, and Procedures, Roles & Responsibilities (Mellett & Nendick, 2007). Adaptive solutions were intended to be triggered by different levels of cognitive control (Terenzi et al., 2014). Essentially, this meant that an ATCO received different automated solutions depending on whether the instruments identified the ATCO's cognitive activity to be taking place at a skill-, rule-, or knowledge-based level. After the preliminary prototypes were developed, they were subjected to user testing (M. Terenzi, personal communication, January 8, 2021). NINA defined HP much more precisely than the HRL does, and the HF pie could be an interesting input for the HRL framework. Perhaps more interesting than the HSI domains (Safety, Manpower, Personnel, Training and Environment) that were not attended to as part of the development of NINA's adaptive solutions. Neither were maintenance and sustainment. Although elements similar to the HSI domains were included in ALICIA and ACROSS, it is worthwhile to note that none of the projects used the HSI domains. Perhaps this reflects a difference between industries or even between Europe and the US. Interestingly, also human user roles, task flow and human-machine function allocations were not updated. Instead, the validation seemed to focus on evaluating whether HP metrics and requirements were met, supporting human use and ensuring the conformance of preliminary design to HP guidelines (corresponding to HRL sub-questions 11, 15, 21, 22, 32, 33, 34, 43, 44, 45).

It is an interesting thought experiment to imagine what HRL could have been assigned to the system that was developed in NINA. The first strand of work focused on a monitoring system that does not interact with the human but merely measures human parameters. This system had a low TRL as it was very invasive, and substantial time and effort were necessary to calibrate it (M. Terenzi, personal communication, January 8, 2021). The focus of NINA was not on the operational efficiency but rather on the experimental efficiency, so choosing the right tasks and physiological indicators. Thus, it seems that the efforts in NINA were mainly increasing technological maturity, improving the hardware and software, and that no special attention was paid to increasing maturity from an HF perspective. Alternatively, NINA could be said to respond to HRL sub-question 4: “Is basic human research relevant to the developing concept or application being conducted?” and sub-question 8: “Are basic task descriptions for user roles being developed?” (see Appendix A). When it came to the adaptive solutions, operational efficiency played a more prominent role. In line with HRL 2, 3, and 4, relevant HF guidelines were identified, needs and requirements for supporting HP were established, and the prototypes were defined in collaboration with experts. There was even a preliminary user test, corresponding to HRL 5.

When drafting the validation plan structure for NINA, the E-OCVM methodology was taken into consideration (Terenzi et al., 2014). NINA was determined to start from V-phase 0, from the identification of an ATM need. The identified need was to monitor the state of operators innovatively by applying neurometrics and physiological science (Terenzi et al., 2014). NINA aimed to advance the knowledge concerning this need to the feasibility phase (i.e., V2), or more specifically, to advance the monitoring of ATCOs’ cognitive workload and create a proof of concept for an adaptive interface. While the HRL states that a proof of concept usually is available at HRL 3, prototypes were already established and even tested in NINA, implying that system development was working towards HRL 5. This maturity level is more or less in accordance with the targeted maturity, as V2 corresponds to HRL 4 (see Figure H2). Apart from the application of the CLM V-phases, there was no further reference to other aspects of the E-OCVM in NINA's validation plan.

Apart from the CLM V-phases, the E-OCVM was not formally followed. This was the case because NINA, like ALICIA and ACROSS, was still at an exploratory level (M. Terenzi, personal communication, January 8, 2021).

NINA also reports on the risks identified at the beginning of the project, their severity and the strategies to mitigate them (Terenzi et al., 2014). At first glance, this seems to be a similar risk assessment to the one in ACROSS. However, it pertains to risks concerning the study, not with regards to the system. HRL contains neither a risk assessment at the system's level nor at the project level. However, this does not mean that it should necessarily be added. The HRL framework is meant to identify maturity from an HF point of view and to plan development and validation activities. Existing risk assessments could be used complementary to the HRL.

STRESS

From 2016 to 2018, project STRESS took place, coordinated by Deep Blue (CORDIS, 2017). STRESS is about HP neurometricS Toolbox foR highly automatEd Systems deSign. As the title already partly gives away, STRESS concerns a project that is very similar to NINA. In STRESS, there were also two validation activities (Bonelli et al., 2017). First of all, the project developed and verified indexes to assess attention, stress, workload and cognitive control. Secondly, the project studied the impact of automated tools on HP. STRESS also focused on neurometrics and ATCOs, just like NINA.

In STRESS, the E-OCVM was (loosely) applied like in previously discussed projects. STRESS started from CLM V0, the identification of ATMS needs. The identified need is the same as in NINA. Validation activities in STRESS were divided into four phases. First of all, preliminary activities took place. Future usage scenarios were selected, HP indexes were developed and tested in the laboratory, and the simulation platforms were prepared. In the HRL framework, usage scenarios are also defined early on (HRL sub-question 2). In STRESS, preliminary scenarios were discussed in a focus group with operators. The HRL does not specify the active engagement of users, and, in fact, sub-question two could be answered by the project team without interaction with the end-user or other stakeholders. The development of indexes and laboratory tests could be an example of basic human research (HRL sub-question 4). After the preliminary activities were completed, the validation was prepared. This preparation included the generation of final scenarios, the recruitment of participants, and the set-up of data collection tools and the platform. In STRESS, psychological measures, as well as task performance, behavioural and self-report data were collected to assess HP. The third and fourth phases consisted of executing the validation and analysing the data, respectively. In total, two validations took place, corresponding to the two aims of the project.

The objective of the first validation was to identify neurometrics that could be used to assess ATCOs' level of stress, attention, workload and cognitive control in a realistic ATM environment (Bonelli et al., 2017). This identification of metrics is in line with sub-question 11 of the HRL. The specific HF were chosen based on a literature review about the impact on HP and the research conducted in NINA (P. Tomasello, personal communication, January 11, 2021). A simulation took place where newly-graduated ATCOs had to perform control activities on a realistic training platform (Bonelli et al., 2017). During the simulations, different levels of stress, attention and workload were induced. Several neurophysiological measurements were taken, namely electrical activity in the brain, heart rate, eye blink, eye movement/fixation, cortisol levels, and sweat gland activity.

The validated neurophysiological indexes were then used in the second validation to investigate the impact of automation on ATCOs' stress, workload, attention and type of cognitive control (Bonelli et al., 2017). The simulation took place on a realistic platform of Ecole Nationale de l'Aviation Civile (ENAC) with a prototype of a highly automated system. Both fresh graduates, as well as experts, participated in the experiment. During the simulation, the ATCOs got to experience different levels of automation as well as automation failures. It is interesting to highlight that both NINA and STRESS looked at novices and experts. In the HRL, these are not mentioned, and the more general term 'user' is adopted. The result of STRESS is a neurometric toolbox for objectively measuring HP. Furthermore, the results of the research provide input for the recommendation of automation design guidelines. Thus, STRESS answered HRL sub-questions 7 ('Have key HP design principles, standards, and guidance been researched?') and 11 ('Are plausible metrics for successful HP being identified?'). Although not the project's goal, STRESS provided an excellent opportunity for ENAC to further develop and validate its automated system (P. Tomasello, personal communication, January 11, 2021). Considering the validation activity conducted - a simulation with a prototype in a relevant environment with realistic users - this system could have attained HRL 5. However, because the development of the automated system was not the project's primary purpose, it is not clear whether and to what extent other HRL sub-questions, such as those related to function allocation and user role procedures, were addressed.

SAFELAND

The last project that will be discussed in the context of this study is the project SAFELAND. SAFELAND is also coordinated by Deep Blue and started in 2020. The project is planned to be completed at the end of 2022. Apart from reviewing available documents, I spent six months observing the activities in project SAFELAND. The observed activities included meetings with consortium partners, informal discussions of Deep Blue colleagues and stakeholder workshops.

SAFELAND aims to create a solution that can provide the SAFE LANDing of incapacitated single-piloted aircraft through enhanced ground support (CORDIS, 2021). SAFELAND answered SESAR's call for application-oriented research. This implies that the maturity level at the programme's start was TRL 1 (Deep Blue et al., 2019). The target maturity level in SAFELAND is TRL2/TRL3 or V1 of the CLM. By now, we can carefully conclude that determining the current and target maturity level, not only from a technical but also an operational point of view, is a common practice in these European aviation research projects.

The project started with a CWA to model work functions and function allocations in single-pilot operations (SPO). More specifically, a Work Domain Analysis (WDA), Control Task Analysis (CTA) and Social Organisation and Cooperation Analysis (SOCA) were applied (S. Bonelli, personal communication, May 7, 2021). The model was developed considering pilot incapacitation, and it was used to explore different possible alternative allocations of the flight functions to the single pilot, air traffic controller, airline operational control centre, and automation. Current dual-pilot operations functioned as a baseline against which the new concepts could be compared in terms of safety. Results from past projects and the literature were used wherever it was relevant. For example, the results from ACROSS were taken into account and literature on the integration of remotely piloted aircraft systems (drones) was consulted (Deep Blue et al., 2019). A recent high-fidelity human-in-the-loop SPO simulation conducted by NASA in collaboration with the FAA was of specific interest. The study indicated that, to fulfil safety and regulatory requirements, all flight deck functions and indications should be available on the ground via remote control in case of pilot incapacitation. Furthermore, ACROSS listed pilot incapacitation as a significant open issue for SPO and proposed a possible role of ground support. Thus, SAFELAND stands on the shoulders of these outcomes. SAFELAND took a different approach than that promoted by the HRL. In the HRL, the first levels start by focusing on the user and its capabilities, limitations as well as potential usage scenarios. Rather than conducting basic human research, SAFELAND takes a broader perspective when defining the concept, namely that of the work domain. Only after conducting the WDA, SAFELAND started to look into the tasks and how these could be divided among the different actors. Moreover, the project also looks into the social organisation and cooperation that constrain the SPO system, which is currently missing in the HRL.

The SPO concept was not developed in isolation by the SAFELAND consortium. Multiple workshops were organised to develop the operational concept further and explore various function allocation alternatives with the Advisory Board (S. Bonelli, personal communication, May 7, 2021). The Advisory Board consisted of SMEs affiliated with aircraft manufacturers, system industries, air navigation service providers, research associations, pilot associations, regulatory and institutional bodies and occupational healthcare companies. Through an interactive process, the initial operational concepts were refined based on the ideas and needs of the SMEs. The workshops were also organised to collect requirements to be tested in the simulation that will be conducted later on in the project. These activities are in line with HRL 3: “Analyses of human operational, environmental, functional, cognitive, and physical needs completed during analytical and laboratory-based studies of the proof of concept to understand the requirements for supporting each human user role” (see Appendix A). One important difference is that, in the case of SAFELAND, human needs and system requirements were collected via workshops with stakeholders rather than analyses and lab studies. Next to operational and HF aspects, the different concepts were also considered in the light of regulatory, technological, and economic aspects (Deep Blue et al., 2019). These were believed to be very important as they directly link to HF-related decisions. For example, from an HF perspective, it could be determined that the safest option (with the lowest workload) for SPOs is to have a ground pilot available for every plane at all times. However, this would be very costly and altogether cancel out the value of running SPOs instead of dual pilot operations. For this reason, HF should be considered in its context and not in isolation. In the HRL, HP is considered separately from regulatory or economic constraints. This could be a critical gap between the HRL v1 and current practices. Ideally, safety decisions are not dependent on regulatory or financial factors, but in case this is an inevitable reality, it would probably be better if the HRL addresses this interrelation. In conclusion, the HRL could be improved by taking the broader work domain as the point of departure and involving the social organisation surrounding the system. Meanwhile, detection of incapacitation, bandwidth and availability of data link, ethical aspects and acceptability regarding the public, Trade Unions, and pilot organisations were all not considered in SAFELAND. Additionally, apart from pilot incapacitation, other anomalies were excluded (S. Bonelli, personal communication, May 7, 2021). So, just like ACROSS, SAFELAND makes several assumptions.

In order to achieve the target maturity level, SAFELAND plans to evaluate the most promising concepts in a real-time simulation with mock-ups and real pilots and controllers (Deep Blue et al., 2019). Data will be collected through debriefings, questionnaires (e.g. NASA-TLX) and over-the-shoulder observations. After the simulation exercise, another workshop will take place to discuss the results of the HF, safety, regulatory, economic and cyber-security analyses. Considering the level of advancement of the test environment, it would make sense to estimate that project SAFELAND will bring SPOs to HRL 5. The operational concept will also include procedures for the handover of control to ground support in case of incapacitation (Deep Blue et al., 2019). This corresponds with HRL sub-questions 30 and 41 (see Appendix A). The procedures, system and automation will be validated in terms of acceptability, (shared) situational awareness, workload, and control of flight tasks. As with previous projects, HP was divided into several factors that were assessed and a baseline was identified, which could be seen as the identification of metrics (HRL sub-question 11, 22, 32, 43). Furthermore, the allocation of flight functions was investigated (Deep Blue et al., 2019). In the HRL, function and task descriptions for user roles (e.g. Q13) are distinguished from the human-machine function allocations (e.g. Q14). Based on the analysis of the simulation and workshop results, SAFELAND will also identify potential additional systems that could be developed to support the ground personnel in their effort to carry out the flight management tasks (S. Bonelli, personal communication, May 7, 2021). For this, the relevant information flows between the different actors in the system (including the automation) will be addressed. In the HRL, no particular attention is paid to information flows between system actors. Potential hazards will be identified and classified with the Functional Hazard Assessment, which is part of the Safety Assessment Methodology of Eurocontrol (S. Bonelli, personal communication, May 7, 2021). Possible mitigation measures will then be discussed. These mitigation strategies could take place at three different levels: Engineering (e.g. system redundancy), Operational (e.g. operating restrictions), and Personnel (e.g. personnel training requirements). Cyber-security issues and their impact on safety will be identified using the SESAR ATM security risks assessment methodology (Deep Blue et al., 2019). The outcomes of this assessment will be used to identify safety requirements for the system. Cyber-security falls outside of the scope of the HRL and could perhaps be considered part of technological maturity.

One thing that makes SAFELAND an interesting project is that it advances SPOs towards an HRL 5, which is higher than the target TRL. The concept is developed and validated taking some technical abilities, such as the data link, for granted. So far, this does not seem to form a problem. Nevertheless, to proceed on the HRL scale, the technological part will have to be developed since HRL 6 demands the validation of a fully-matured system rather than a prototype. As stated in the proposal, SAFELAND will integrate the results of all evaluation exercises and provide conclusions about the next steps on the road to higher maturity. The HRL could be used to this end since it clarifies which exit criteria need to be fulfilled to reach the next maturity level and those that follow.

Appendix H. Elaborated description of comparison between validation frameworks and HRL v1

HF Case

The HF Case represents a process by which the structural identification and treatment of HF issues can be managed from the beginning to the end of an ATM project (Mellett & Nendick, 2007). Eurocontrol developed it to provide a system design approach that integrates technical, human and procedural performance. It was first launched in 2004, targeting project managers who could use the process to reach their operational objectives. A human-centred philosophy underlies the HF Case. It promotes the idea that a system should be based on operational requirements that ensure the human strengths and capabilities are used to the best advantage, and the human limitations are compensated as much as possible.

The HF Case consists of five stages: Fact-Finding, Issues Analysis, Action Plan, Actions Implementation, and HF Case Review (Mellett & Nendick, 2007). During Fact-Finding, it should be identified what will change, who will be affected by the change, and *how* they will be affected. HF issues are identified at a high level using the HF Pie categories. These were also used in NINA. See Table H1 for the definition of each of the six categories. In the next stage, Issues Analysis, the potential impact of the design on HP should be considered by using a What-If analysis in combination with the HF Impact Wheel. The HF Impact Wheel consists of 12 essential HF impacts on HP: cognitive processes, acceptance, comfort, fatigue, error, job satisfaction, situational awareness, motivation, skill change, trust, stress, and workload. See Appendix I for the definitions of all HF impacts. Project-specific HF issues with regard to the HF impacts need to be identified and prioritised. After the Issues Analysis, mitigation strategies should be defined during the Action Plan stage. In order to identify the HF actions that are required, data can be collected to gain a better understanding of the HF issues. Possible studies are literature review, cognitive task analysis, workload assessment, inspection of the workplace and equipment, direct observation of operations, user surveys (e.g. SHAPE), investigation of hazard and health issues reports and incidents, error and near-miss investigation. In the subsequent stage, these actions will be implemented. Finally, the HF Case should be reviewed by an independent HF expert.

Table H1

HF pie categories defined

HF Issue category	Description
Working Environment	The working space, general equipment and furniture used, and physical environment in which people work.
Organisation & Staffing	Organisational management, people management, and personal factors.
Training & Development	The systematic development of the knowledge, understanding, skill and attitude behaviour patterns required by an individual in order to adequately perform a given task.
Procedures, Roles & Responsibilities	The organisation's accepted working methods, the purpose that someone has in an organisation, and his/her duties.
Teams and Communication	How people work and communicate with each other on shared goals and tasks.
Human in System	This category emphasises that the human is a key part of the system. One main concern is human-machine interaction.

Note. Based on Mellett & Nendick, 2007, p. 57-73

The HRL prescribes a similar process as the HF Case. However, in the HRL, the analysis of issues and the identification and implication of mitigation strategies takes place multiple times in validation environments of increasing fidelity and with prototypes of increasing level of detail. This difference is partly due to the fact that the HF Cases is project-oriented, whereas the HRL takes the system as the point of departure. The HF pie categories used in the HF Case are comparable to the HSI domains in the HRL. In both structures, Environment is included, and Organisation & Staffing and Training & Development capture the same elements as the HSI domain of Manpower, Personnel and Training. Procedures, Roles & Responsibilities and Human in the System are also considered in the HRL and would fall under the HSI domain of HFE. HFE is not mentioned in the HRL, but this domain is considered to be answered through all of the sub-questions that do not specifically address one of the other HSI domains (P. Savage-Knepshield, personal communication, November 3, 2020). The HRL could potentially benefit from applying categories or domains consistently rather than mixing sub-questions on domain-level with sub-questions that pertain to specific elements of an HSI domain. The HF pie categories could be used instead of the HSI domain since it also captures elements that are not promoted by the HSI domains, like team and communication factors. Furthermore, the Impact Wheel could be an excellent addition to the HRL in order to stimulate the consideration of all relevant HF. For example, they could be used to identify not only potential HF issues but also relevant HP metrics, requirements, and design principles. Another benefit of the HF Case compared to the HRL is that it concludes with a Review stage. In the HRL, it is required to write validation plans and reports, but an independent review is not part of the HRL process yet.

The HF Case also defines roles for managing and executing the validation process. First of all, a coordinator is responsible for implementing the HF Case (Mellett & Nendick, 2007). Additionally, a facilitator should be available to facilitate workshop sessions and expert interviews during the Issues Analysis stage. Finally, a key stakeholder team should be established at the beginning of the project. The HF Case is very much in favour of actively incorporating stakeholders early on in the project. The HRL does not explicitly mention the involvement of stakeholders. Users are expected to be part of the simulations, but it is not said that they will be asked for their opinion or ideas. Theoretically, HRL 9 could be achieved by merely measuring HP without speaking to the humans themselves.

The HF Case can be applied to existing, modified and new systems (Mellett & Nendick, 2007). This is another difference compared to the HRL since the HRL is created explicitly to support the development of new systems. Perhaps (part of the) HRL could also be used to validate modified systems. Just like the HRL, the HF Case is supposed to be initiated as early in the project life cycle as possible so that HF issues can be identified and resolved before it becomes (financially) unfeasible.

E-OCVM

A framework that was used to at least some extent in all five projects is the E-OCVM. This framework was first released in 2005. It is considered to be of proven value to the validation of large projects and therefore mandatory for Eurocontrol and European Commission ATM validation activities (Eurocontrol, 2010). The E-OCVM is to be used by managers responsible for setting up systems development programmes and the managers that supervise those programmes and the resultant projects, as well as systems developers and validation practitioners. Apart from these target users, the E-OCVM methodology also supports the understanding of ATM research and development amongst the broader stakeholder community, such as system users, regulatory bodies and technology providers. In general, the target group of the E-OCVM is very similar to that of the HRL, although there is no mention of stakeholders in the HRL apart from system users.

The E-OCVM aims to achieve a more consistent R&D process across collaborating organisations while simultaneously leaving these organisations freedom to execute the specific validation activities (Eurocontrol, 2010). Moreover, the E-OCVM aims to support validation activities at any scale, ranging from single studies addressing a local operational change to large international programmes. This implies that it does not provide instructions that can be applied blindly. Instead, the E-OCVM should be used to create a customised validation process suitable for the scope and the complexity of the project and concept being developed. Like the HRL, the E-OCVM has been created to offer more structure and transparency throughout the development and validation phases of operational concepts. The E-OCVM furthermore distinguishes validation and verification, where validation treats the question of whether the right system is being built, whereas verification deals with whether the system is built in the right way. The E-OCVM includes maturity criteria that assess both these concept- as well as more system-related aspects. The same goes for the HRL. Especially the first couple of levels look, for example, at possible usage scenarios and alternative strategies to deal with HP issues, whereas the later levels focus on validating the proposed designs and the final system. When the E-OCVM was introduced, several pitfalls were presented that could be tackled by the E-OCVM. These pitfalls were based on experience with validation projects. They included the failure to use the results from past projects and the failure to document activities and the corresponding assumptions, leading to the avoidable repetition of work. The HRL could also support the avoidance of these pitfalls. For example, the HRL structure makes it easier to attach a specific activity to certain sub-questions, thereby keeping an overview of the effort that has already been completed. The HRL could be extended to capture also the assumptions that will inevitably be made during projects, as we have seen in ACROSS and SAFELAND. Other pitfalls were the failure to adequately assess the maturity level and consequently the planning of inappropriate validation activities or the tendency to focus on quick wins and leave more complex problems for later on. These are two problems that the HRL tries to address by providing a maturity scale that promotes HF validation activities throughout the system development life cycle.

As explained before, the E-OCVM consists of three components, the CLM, the SPF and the CBApp. Since the E-OCVM bears a resemblance to the HRL in terms of aim and target group, it is worthwhile to have a closer look at each of the components. The CLM consists of 7 phases in total. At V0, the need for change is established. This means that ATM problems are identified and prioritised. Usually, this is the phase leading up to a call for projects. At V1, managerial activities such as scoping the project aims and creating a validation strategy and plan take place. In V2, the focus is on analysing, modelling and running simulations. V3 subsequently focuses on integrating prototypes and procedures into a representative environment, and during V4 regulatory activities take place. V5 focuses on the deployment, and V6 implies that the system is being used, maintained and monitored. The last phase, V7, stands for the removal and replacement of the system. The CLM views systems development as a circular process, where every system inevitably gets replaced at some point, initiating a new cycle starting with identifying the need for change (P. Lanzi, personal communication, January 8, 2021). According to the CLM, there is always a default system or situation that is being changed by the new system's development. This default situation is taken as a baseline, and the change, which can be a relatively minor change but also a completely new system, is analysed in relation to the baseline. Apart from the reference to the analysis of legacy or comparable systems (sub-question 9), the HRL does not refer to the current system as a baseline. When assessing the change to a new system at V2, the CLM covers many aspects, including allocation of people and training. In that way, it is similar to the HSI domains, which also cover a vast area of HF elements. However, the CLM has an ever broader perspective as it also includes, for example, regulatory aspects. The E-OCVM focuses mainly on V1, V2 and V3 (E-OCVM). After V3, the system development typically transfers from the R&D community to the industry. V1, V2, and V3 are comparable to HRL 1 to 6. This entails that the HRL provides more granularity to the development process. Whereas the HRL goes more into detail when it comes to analysing, modelling, and simulation, the CLM pays more attention to the system's deployment, taking into account not only HF but also regulatory aspects.

As explained in the paragraph about ALICIA, the SPF consists of six steps. For each of these steps, the SPF proposes critical activities at the programme, project and exercise level. See Figure H1 for the breakdown of SPF step four as an example (Eurocontrol, 2010). In general, the SPF is focused on the managerial side of systems development, while the HRL is focused on the system validation itself. This makes sense as the SPF was designed to be performed (at least once) for each CLM V-phase. The HRL does not specify managerial considerations apart from sub-questions regarding the drafting and finalising of plans. Whereas the E-OCVM separates the SPF from the CLM, the managerial and maturity-related sub-questions are lumped together in the HRL.

Figure H1

Key activities at programme, project and exercise level for step four of the SPF

Step	Sub-Step	Name	Activities – Programme Level	Activities – Project Level	Activities – Exercise Level
4. Analyse the Results	4.1	Analyse the data as planned.			Analyse exercise data.
	4.2	Prepare analysis contributions.			Prepare exercise analysis contributions.
	4.3	Prepare the validation report and cases.	Integrate results of projects. Build programme-level cases reports. Consolidate project validation reports.	Integrate results of exercises. Build cases. Produce project validation report. Pass results up to programme level. Train participants (where required).	Identify validation exercise shortcomings. Produce exercise validation report(s). Pass results up to project level.

Note. By Eurocontrol, 2010, p. 44

An essential element of the SPF is to select the appropriate validation tools for the activities (Eurocontrol, 2010). This selection depends on the concept’s maturity. However, it is up to the validation practitioner to determine which tool is most appropriate to the needs and objectives of the validation activity. Several techniques and tools are mentioned as examples: literature review, SME interview, fast-time modelling, gaming, real-time simulation, shadow mode trials, and live trials.

The third element of the E-OCVM is the CBApp. The CBApp facilitates grouping all information (mostly evidence) concerning particular perspectives (e.g. HF, Business, Standardisation and Regulation, Technology) into dedicated cases. These cases can be used by project or programme managers and other stakeholders to make well-informed decisions. Moreover, the cases can be used to capture risks and open issues.

The E-OCVM also describes a maturity assessment process to decide whether a project can transition to the next CLM V-phase. If the achieved maturity of the concept is too low to progress to the next CLM V-phase, the SPF steps should be repeated. On the other hand, if the maturity assessment indicates that the concept has already reached a higher level of maturity, life cycle phases may be skipped. At the basis of this assessment are the life cycle transition criteria, which can be found in Table H2 up until V4 (Eurocontrol, 2010, p. 38-42). It becomes again apparent that the E-OCVM covers a more extensive range of aspects than the HRL. The HRL complements the TRL by adding HF considerations, whereas the E-OCVM combines HF with technical feasibility and affordability. Although some transition criteria are beyond the scope of the HRL, several criteria are also relevant to the maturity framework. Take, for example, ‘potential alternatives’. In SAFELAND, several alternative concepts were considered for development and validation. In order to reach HRL 4, trade studies are used to identify the most viable design options and alternatives strategies to support human use.

Table H2

CLM V-phase life cycle transition criteria

V1-V2 Transition Criteria	What are the operational concepts/enablers to meet the needs?
	What are the potential contexts of use / application / deployment?
	What are the related operational concepts and their possible implications?

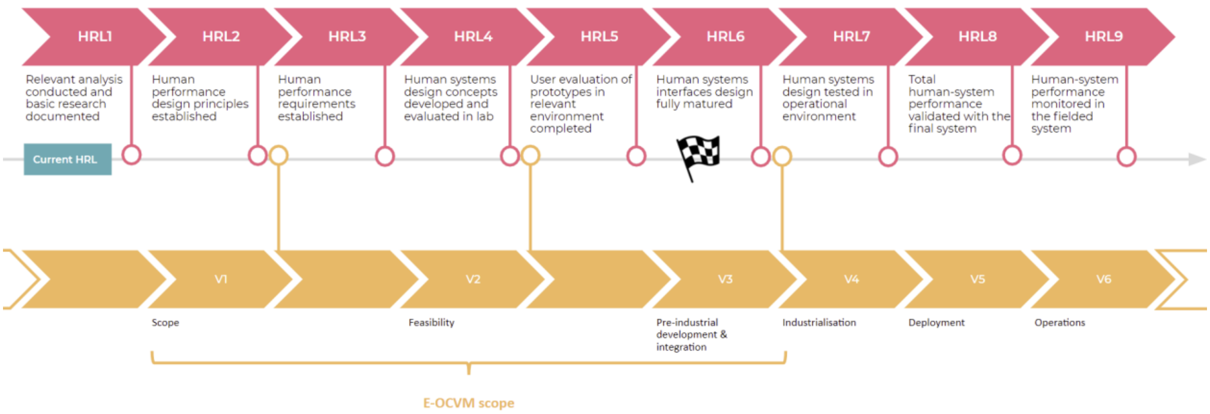
	What are the potential benefits/costs?
	What are the potential alternatives?
	What needs deserve to be validated (R&D needs)?
	What are the potential risks (solution risks)?
V2-V3 Transition Criteria	Operational feasibility (user acceptance, safety)?
	Technical feasibility (preliminary assessment based on research prototypes)?
	Transition feasibility (including institutional issues)?
	Potential benefits validated for concept options?
	Affordability for stakeholders?
	Alternative solutions compared?
V3-V4 Transition Criteria	Operational feasibility when integrated into the target system with other concept elements?
	Pre-industrial feasibility?
	Transition feasibility: Possible refinements?
	Potential benefits for alternatives/options sufficiently validated?
	Affordability for stakeholders: possible refinements?
	Availability of mature specification material to enable standardisation & regulation, if so intended?

Note. Based on E-OCVM, 2010, p. 38-39

Together, the CLM, SPF and Cases, supported by the maturity assessment, provide a means of supervising and synchronising different exercises and workstreams throughout the development life cycle. This is very useful as the analysis of Deep Blue projects also showed that the development of a new system is divided over various programs, projects, and work packages. The HRL can offer this same benefit but is more in accordance with the TRL and provides more granularity. Although the HRL has a finer structure than the E-OCVM's CLM, the levels and phases can easily be related to each other. A visual representation of this relation can be seen in Figure H2. This interrelation is based on the mapping of the TRL on the CLM, assuming that the TRL and HRL should ideally be aligned (SJU, 2020). However, there are some inconsistencies in this regard. For example, another document assumes that V0 is equal to TRL 1, that V1 covers TRL 2 and TRL 3, and that V2 is reached at TRL 5 (Morgan et al., 2018). Yet another document, namely that of ACROSS, differentiates from the first one as it supposes that V2 is reached at TRL 3 (Billiere et al., 2013). Since there is no complete agreement, the first document was followed when creating Figure H2 because it corresponds largely to the information presented in the ACROSS report. For the only phase in which all three documents differed, the first one offers the middle way (V2 is reached at TRL 4 rather than TRL 3 or TRL 5).

Figure H2

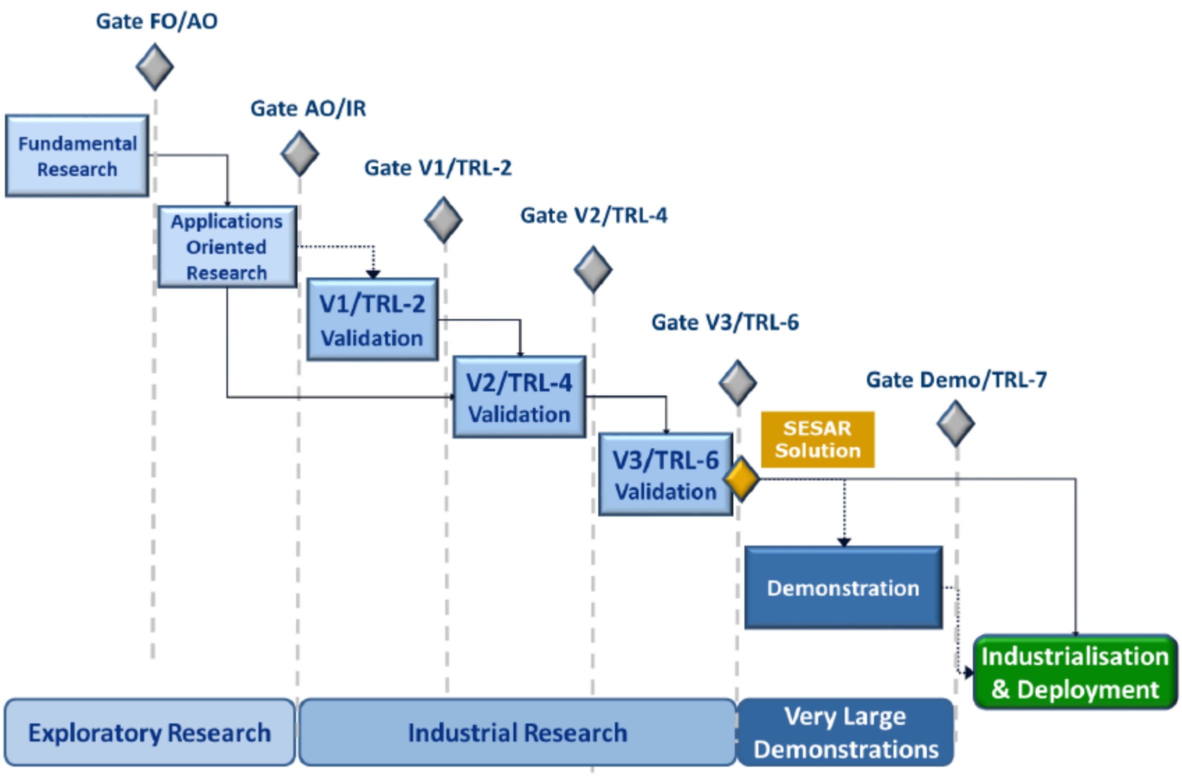
HRL in relation to the CLM V-phases



Most of the projects reviewed as part of this thesis were labelled as 'exploratory research'. As shown in Figure H3, these types of research are taking place before V1, corresponding to HRL 2. This figure comes from the SESAR Project Handbook, a guiding document outlining the deliverables that are required to be produced for SESAR solutions at each maturity level (SJU, 2020). FO stands for fundamental research, AO for applications-oriented research and IR for industrial research. Since most of the Deep Blue projects included simulations in relevant environments, they would be assigned an HRL much higher than HRL 2 (provided that all sub-questions were answered). That could mean that Morgan et al. (2018) were more accurate than the Project Handbook (SJU, 2020) and Billiere et al. (2013) after all about the relation between TRL and V-phases. For now, the relation between the V-phases and the TRLs, and thus also the HRLs, remains unclear.

Figure H3

Gates and phases of SESAR development life cycle



Note. By SJU, 2020, p. 23

Human Assurance Level

Eurocontrol developed the HAL framework to complement its existing Safety Assessment Methodology with guidelines dedicated to HF (Eurocontrol 2007). This framework can be used to answer the question of which HF need to be addressed in a project, the depth and scope of assurance required, and the evidence needed to prove that HF have been satisfactorily addressed. In a similar way that the HRL is inspired by the TRL, the structure of the HAL was based on the Software Assurance Level framework. In total, there are four HALs, which all contain Human Assurance Objectives. Currently, the HAL framework applies to operations room personnel only. The HAL framework was developed in five steps (L. Save, personal communication, January 7, 2021).

The first step was to identify how other system components influence the human. The HAL described five categories of factors that should be considered: People, Equipment, Operations, and Integration. People factors include both factors that are internal to the operator (e.g. skills) and factors that refer to other people or the social organisation (e.g. supervision, culture). Equipment factors refer to characteristics of the tools that are used as well as the working environment. Operational factors are factors such as procedures. Interaction factors cover the total system performance that emerges from the interaction between all the aforementioned factors. Since considering the system's characteristics alone may not be sufficient to manage all human-related risks, process factors were added to the list of factors. Process factors refer to the managerial and methodological processes followed throughout the different development stages to assure that people, equipment and operational factors are covered. Whereas the HRL looks at internal People factors, such as human capabilities and limitations (sub-question 1, 6, 21), it does not attend to aspects related to the external social world, apart from manpower, personnel and training factors. It seems that the HRL focuses more on the relation between the operator and the equipment. Procedure factors are included in the HRL (see Appendix A; sub-question 30, 41, 53, 64, 75), and integration takes place from HRL 4 onwards. The HRL is, in essence, a process and therefore covers Process factors. It also contains specific managerial considerations such as recruiting the right experts and preparing the test plan.

The second step of developing the HAL was to define the HAL grading policy (L. Save, personal communication, January 7, 2021). The grading policy describes the characteristics of each HAL, such as the severity of potential hazards, different potential mitigation strategies and examples of evidence that assure successful mitigation of human-related hazards. Next, a table was created to indicate the objectives for every HAL and each *life cycle stage* (i.e. Definition, Design & Validation, Implementation, Transfer in operations, Operation). These objectives have been developed with reference to the HF Case. The HAL framework ties the level of effort spent on managing and assessing HF to the level of risk. The life cycle stages are similar to the HRLs as they are also related to the system's maturity, although the HRL offers more levels and thus more granularity.

At step four of the HAL's development, a matrix of the severity and likelihood of a potential hazard has been created to allocate the HAL. For example, if the severity of the effects is 2 (out of 4, 1 being the most severe) and the likelihood is 'possible', HAL 2 will be allocated. The outcome of all steps is a table where the following information is included for each of the HAL objectives: title, short description, category of factor (as described at step 1), life cycle stage at which this objective must be met, rationale, examples of methods and techniques that can be used to achieve the object, examples of the depth and scope of the level of evaluation, examples of the level of evidence required, and the corresponding HAL.

There is an increasing stringency of assurance at each level, HAL 1 being the most stringent one. Furthermore, the levels are additive, which means that the objectives that need to be achieved for HAL 2 also include those for HAL 3 and HAL 4. HAL 1 is extremely stringent and should be reserved for exceptional cases. The HAL can be assigned to a project as a whole but also to various scenarios (Human Engineering Limited, 2007).

The HAL framework should be applied as soon as the information about the project and the system are detailed enough to determine the HAL. Practically, this means during the design phase, towards the end of the Functional Hazard Assessment and the beginning of the Preliminary System Safety Assessment. However, it is acknowledged that the initiation of the HAL might take place at a later stage of the development for several reasons. When reviewing the Deep Blue projects, this was also a recurrent theme as HF expertise was often engaged once the technologies were already quite mature in a purely technological sense. The HRL does not offer a solution strategy for this situation, whereas the HAL framework does. The HAL advises, in these cases, to address at least the HAL objectives of the Definition phase as these are fundamental in addressing human-related safety issues.

The most significant difference between the HRL and HAL is that the HAL is all about risk. In essence, the HAL framework provides a very extensive answer to HRL sub-question 3: “Have potential key human systems issues and risks throughout the life cycle been identified?” (see Appendix A) and then suggests, based on this answer, objectives similar to the HRL sub-questions. As stated in the report by Eurocontrol (2007, p. 4): “The aim [of the HAL framework] is to provide a straightforward framework such that effort associated with assessing and managing HF is appropriate to the level of risk inherent in the system or change". This means that the framework can be meaningfully applied to systems of different scopes. Instead, the HRL is relevant mainly for highly safety-critical systems, perhaps those that would be assigned HAL 1.

SESAR Human Performance Assessment Process

More recently, SESAR published the SESAR HPAP to ensure the systematic consideration of HP (HP) in SESAR projects (Morgan et al., 2018). The HPAP is compatible with the E-OCVM. HP is one of the six transversal areas used to develop the Business Case (see the E-OCVM CBApp). The HF Case is the bedrock of the HPAP (P. Lanzi, personal communication, January 8, 2021). Compared to the HF Case, the HPAP is more detailed.

HP, defined by SESAR as “the human capability to successfully accomplish tasks and meet job requirements", depends on various factors: procedures, task design, technical systems and tools design, physical work environment, individual competencies, training, recruitment and staffing, as well as the management of social factors and issues related to changes and transitions (Morgan et al., 2018, p. 9). These factors are studied within the discipline of HF.

The scope of the HPAP comprises CLM V1, V2, and V3 (Morgan et al., 2018). The HPAP for V1 mainly focuses on defining the preliminary operational concept (from an HP point of view). More specifically, in V1, descriptive models of the human component are developed, the impact of the operational concept on roles/tasks/constraints/needs as well as system performance are explicated, the development of a preliminary operational concept and system architecture are supported, side effects that negatively influence HP are identified, and human-centred models are linked to system models to ensure coherence. The result of V1 should be an operational concept mature enough to be prototyped and tested at the next V-phase. The following techniques can be used for concept exploration: focus groups, task analysis, issues analysis, gaming techniques, use cases, and storytelling. This is not an exhaustive list, but examples of methods provided by the SESAR HPAP. Additionally, low-fidelity mock-ups and prototypes may be used in V1. From this description, it seems that V1 overlaps with HRL 4 rather than 2, as was visualised in Figure H2. The HPAP for V2 focuses on improving and validating the operational concept from an HP perspective. This happens through an iterative process of design, evaluation and redesign. At V2, the operational concept is detailed enough for thorough task analysis techniques and for cognitive walkthroughs, usability assessments and performance measurement supported by mock-ups, prototypes and (real-time) simulators. Based on this description, one would expect V2 to correspond to HRL 5, which is in contrast to Figure H2, but in line with Morgan et al. (2018). The HPA for V3 mainly focuses on pre-industrial validation and integration. HP activities in

V3 should occur in an environment that is more realistic than in V2, which means that low fidelity mock-ups and prototypes are not recommended anymore. In addition to real-time simulations, operational trials (shadow mode trials) can be used. These activities are in line with HRL 6, as also presented in Figure H2.

Several topics are central to the HPAP, that is (1) Roles, responsibilities, operating methods and human tasks, (2) Technical support systems and Human-Machine Interface, (3) Team structures and communication, and (4) Potential transition factors (Morgan et al., 2018). Most of these topics also receive considerable attention within the HRL framework. For example, a great number of sub-questions relate to task description, analysis and flow as well as user roles and procedures (i.e. 8, 13, 30, 31, 41, 42, 46, 53, 57, 68, 75; see Appendix A). Also human-machine function allocations (i.e. 14, 29, 40) and support for human use (i.e. 15, 33, 44, 55, 66) are covered by the HRL. The sub-questions concerning task description partly cover team structure, but communication is not mentioned at all in the HRL. Potential transition factors relate to "acceptance and job satisfaction, changes in competence requirements and impact on staff levels" that need to be considered (Morgan et al., 2018, p. 17). The HRL also captures this topic with the sub-question regarding manpower, personnel, and training analyses (i.e. 17, 25, 36, 49, 60, 71).

At each V-phase, the HPAP follows four steps (Morgan et al., 2018). Firstly, the ATM concept should be understood, and then the HP implications have to be identified. Subsequently, the concept will be improved and validated, and lastly, the findings are collated and it is concluded whether the transition to the next V-phase is possible. The HPAP of a system is assessed against a 'reference scenario'. Instead, the 'solution scenario(s)' refers to the system change that is proposed in order to solve or answer an ATM problem or need. The HPAP also looks at the constraints, or the limitations on behaviour, that the new solution yields. Moreover, during V1 and V2, the HPAP requires the identification of related solutions in order to identify potential interactions. This is similar to the HRL sub-question that asks to analyse comparable and legacy systems (see Appendix A; sub-question 9), although the goal of the sub-question is to identify lessons learned rather than ensuring successful integration. Just like the E-OCVM, the HPAP offers a strategy in case the HP maturity of the system is not equal to the overall maturity. In these situations, the HPAP states that an HP specialist should refer to the HPAP for the previous V-phases and decide which validation activities should be conducted before the activities of the given V-phase (Morgan et al., 2018). To understand the HP implications of the proposed concept, the HPAP offers an argument structure. By going through this structure, it can be systematically decided on which HP topics the new solution has a potential impact. The argument structure contains higher-order arguments directly related to the four topics mentioned before and breaks these further down into lower-level arguments and activities. This structure is very similar to the structure found in the HRL, with core and sub-questions. The HPAP distinguishes itself from the HRL as it not only looks at issues but also at potential benefits. Something that was also done in ALICIA. Furthermore, the HPAP refers to an HP tools and methods repository that can be found online (https://ext.eurocontrol.int/ehp/). Here, it is possible to find tools for all four topics mentioned earlier. However, they are not sorted by V-phase or maturity in general. The HPAP also emphasizes that one activity can be used to support several HP arguments (Morgan et al., 2018). The HPAP also shows how parts of the outputs obtained during the four steps directly feed into the SESAR documentation. Where the HRL is a process tool that is not made for a specific industry or project, as long as the object of analysis is a complex socio-technical system, the SESAR HPAP is highly adapted to SESAR projects.

On top of the argument structure, the HPAP contains a maturity criteria checklist that can be used to determine whether all required information for a certain V-phase has been collected. One of the questions on the maturity checklist is: “Have relevant arguments for V1 been addressed and appropriately supported?" (Morgan et al., 2018, p. 93). Additionally, there are questions concerning, among others, the assessment of benefits and issues, the identification of potential interaction with related concepts, the selection of different concept options, the assessment of preliminary solutions with experts, the revision of relevant documentation, and the preparation of the next V-phase. Although the HRL contains criteria and assessment questions similar to the HPAP's arguments and maturity criteria, it does not separate these. The maturity checklist criteria relate to HP but are more abstract than the arguments in the argument structure. Take, for example, the following maturity checklist questions: "Are the benefits and issues in terms of HP and operability related to the proposed solution sufficiently assessed?" and compare this with one of the arguments: "Human actors can acquire an adequate mental model of the machine and its automated functions." (Morgan et al., 2018, p. 93). The latter treats HP more specifically. Next to the mental model, the HPAP arguments zoom in on situation awareness, workload, level of trust, alarm and alerts knowledge, and skill and experience.

Appendix I. Practitioners' definition of HF impacts

Acceptance	The fact to consider something or someone as satisfactory.
Cognitive processes	How people think, make judgements and problem solve on the job. This includes information processing capability, memory, decision-making, vigilance and attention span.
Comfort	How people physically perceive and experience their working environment.
Error	A generic term to encompass all those occasions in which a sequence of mental or physical activities (intended or unintended) results in an undesired outcome.
Fatigue	The need for recuperation of the resources being used for the task in hand. Our focus is on fatigue and 'alertness' and how it affects HP, not physical or 'mental' fatigue.
Job Satisfaction	A term used to describe how content an individual is with their job.
Motivation	Enthusiasm for doing something. The reason a person has for acting in a particular way.
Situation awareness	The accurate perception of what has happened, what is currently happening, and what is therefore likely to happen next.
Skill change	Skill is the ability to do an activity or a job well, especially because you have practised it. Skill change is the gaining or losing of skills, mostly through practice or the lack of practice.
Stress	When perceived demands exceed performance capability. A subset of subjective workload where it is appraised as negative.
Trust	The extent to which the user is willing to act on the basis of the recommendations, actions and decisions of a computer-based tool or decision aid.
Workload	The effort invested by the human operator into task performance. Varies as a function of ability, skill, training and experience. Workload relates to objective workload (task demand) and subjective workload (individual perceptions).

Note. By Mellett & Nendick, 2007, p. 75-76

Appendix J. Original transcripts of the HRL v2 evaluation interviews

Below are the original transcripts of the interviews conducted to evaluate the HRL v2. Small talk at the beginning and end of the interviews and the part in which the experts describe there background are not included to respect privacy concerns.

Interview with aviation expert from Deep Blue

Interviewer: I will first tell you a bit, actually everything is on the slide. The human readiness levels, they are like to Technology Readiness levels, but they focus on human factors. And so just like the TRL, they are meant to show the maturity of a certain technology, but then more from an operational point of view. And the two frameworks are complimentary, so they can be used together is not a replacement for the TRL. It can be used by project managers like yourself, or by the people that are doing the system development and validation, which I think you also do so you are part of both of the target groups. And it's meant really for very complex high risk, social-technological systems of large scale. So, the ones I would say that we are working on in Deep Blue and also in SESAR, and what else is good to know, it should be aligned as much as possible with the TRL. So, if you are in TRL 1, you should be in HRL 1. That's the ideal situation, and the earlier the HRL is being applied, the better because, as you know, it's better to find Human Factors issues at the beginning, when it's still easy to solve them instead of waiting till the end, which, sadly, is often the case. And then, if you do that, if you apply it and you apply it as early as possible, you can improve the project planning and decision making, you can mitigate human factor risk, and you can prevent that you will have a lot of costs and delays later on in the in the project. So that is really shortly what the HRL is about. And I would like to ask you, is there a project that you were involved in or that you are involved in, that would fit this description? There are probably many but is there one example that comes to your mind that you think the HRL could be useful for that? Based on this short description?

Interviewee: I have one question before. When we say that TRL is focusing on technological and ... Okay, technological part and HRL on operational. Okay. Then operational meaning that if the system is built for being adapted to support actually the human capabilities, or if the human is ready to... how is the...?

Interviewer: Yeah, that's a good question. Yeah. If I understand your question correctly, it's both. So it's both whether the system is ready to be used by humans, so whether it supports the human use, but also whether the human received the correct training, so also whether the human is ready to use it, so it includes both. And the TRL is really about it: Does the technology work? Like when we make this plane go in the air? Does it collapse? Or does it fly? So it's very simple just the technology, whether the technology works.

Interviewee: So I understood that it's both cases. The system if it's build right and in supporting the human capability also the human operator which means if it's ready to operate, to perform with that kind of system, right?

Interviewer: Exactly. Yeah.

Interviewee: Okay, so then I think that most of the SESAR projects can fit on it. Also past projects from exploratory research. If you want to start from early developing phase. Now, for example the project 35 that can be useful to use it in this sense maybe, it's now aiming at V2, I'm not sure... It is relevant that is really close to the actual level of maturity? In that case I need to check it. I don't remember if it's actually V2 or ...

Interviewer: We will go through the HRL to see what HRL we would assign to it, so we don't necessarily need to know the V phase. Let's keep it easy.

Interviewee: Okay. For the project 35 the maturity level is V3, is it suitable? It's aiming at V3.

Interviewer: Yeah, it can be any maturity level, it doesn't matter. Because of course, you didn't apply this HRL. Everything is fine. So, can you explain a bit about the project?

Interviewee: Yes. So, this project, further exploits the capability of the multiple remote tower concept. So, the idea is to validate a new setting on which two or plus module, two or three models, which represent each of them a single controller position, and can of course, implement more than one airport for each. So each module represents a potential multiple remote tower position. On which up to three airport can be swapped. And the idea in this environment, this multiple remote tower environment, is to provide another role, test another role, the supervisor role, which actually is not a new role itself, but it's new the function that is expected to be performed by the supervisor, which is to allocate the airport to the different modules according to specific criteria, like for example, workload of controllers, or in engagement phase or engagement state of controllers. He is supported by specific tools, which is monitoring in some way the status of the single modules, so the single airport as well, according to this he can test and then apply the different allocation.

Interviewer: Okay, so this is a person, it is a person doing this supervision?

Interviewee: The supervisor role is done by the supervisor, is a person, is a controller to which this more... with this role, which is more focused on the monitoring not on the traffic control management itself.

Interviewer: Okay. All right. And let's move on to the to the next slide.

Interviewee: Yes.

Interviewer: Because there are a couple of steps of using this HRL. So we can go through them and then we can try to find the HRL for this pj 35. So you start with defining the system that you want to develop, which we just did, then you have to determine where the, whether the exit criteria associated with HRL one and two were fulfilled. So we will go later to the next page to see what they are. If they are not, then we first have to start planning these activities to fulfill these exit criteria. If they are, we can continue to step three. And at step three, we have to determine the relevant cluster. So there's three clusters in the HRL; research and development, technology demonstration, and production and deployment. And every cluster also has three HRLs. So we first chose the cluster and then the HRL. Then we look at all the exit criteria that are associated with that HRL and in case they are not fulfilled, we move one level lower until we found the level where all the exit criteria are fulfilled. And that would be the current HRL. And then from there, you can start to plan the next Human Factors activities to go to the next level. But yeah, let's start with step two, I would say. So, we can go to the next page.

Interviewee: Yes, okay. The bigger one.

Interviewer: Yes, exactly the big page. Yeah. So, the first step was to check whether the criteria of HRL one and two are fulfilled. So just...

Interviewee: Is their mental constraint identified and design principle established right?

Interviewer: Yes

Interviewee: Okay, this first row uh this first column is you want in the system, work environment, okay. For each of them, for each of this [exit criteria], I will call argument because they are very similar to the ones of the HP material let's say, are covered by the concept right now, right?

Interviewer: Yes, correct.

Interviewee: So, for I start from HRL 1, basic human parts documented and conducted, yes, functional structure of the work domain identified, yes, input output the constraint of the control task identified and process and how task can be done described, right, done. Yeah, has been conducted the workshop before focus group but also validation activity, but also field study, several activities has been during these years, because the concept is very well known now, it's studied for different years so.

Interviewer: Okay, so it's probably more mature than that?

Interviewee: Yes than one yes for sure. Then two, let's see. Human performance ... Yeah, because the concept in terms of.... I mean, we are not talking about the remote tower concept itself no? Because that concept is well mature. No, but we are talking about the allocation, the multiple remote tower with allocation of airport. In that sense, we can say that the concept has been studied previously. So, the HRL one can be considered achieved because some validation has been conducted. Honestly, I was not involved in that activity. But working now, in this stage, we of course, had to look back to potential dependability with previous activity. And in that sense, we find out that there were some work done, but actually the current technical implementation in which the supervisor is provided with a specific tool to manage this swap was not done before. So, this is..., we can consider this one, the focus of this activity now, which is the focus of the activity also in SESAR actually.

Interviewer: All right. So...

Interviewee: HRL one and two are achieved we said, okay?

Interviewer: Can you say that, in general, the design principles were established? Or do you have to go through the different exit criteria to determine whether the HRL two was finished? Just to know how you use it. I'm just trying to see how people use this, because this was developed by some scientists that maybe they didn't keep in mind, how do you actually use this in practice. So I'm just trying to see how you would use it. So, would you go through all these exit criteria, for example, human system performance design principles researched? Or can you already tell from the title of the HRL, whether it was completed or not? Like what makes you...

Interviewee: No no, I would like go through them because from my point of view and for the work that we are doing now, some specific aspects related to the supervisor tool, for example, need to be addressed. This is the phase on which we are doing that. Particularly the design principle, really design principle for the development of that tool. For example, how this tool behave, so also the enabler part for that part of activity, so the technical part I'll say, but also then the interface with the controller, with the supervisor. These are things that are currently under developing in that phase. Now, we actually reach a good level of maturity, we are almost ready, of course, but it was work that has been done during that phase. So reaching v3 phase working on a v2 right now. Okay?

Interviewer: Okay, good. So we did step two, we confirm that HRL one and two were fulfilled. So now we can move to step three, and then we have to choose the relevant cluster. So we have to choose whether this falls in the category of research and development, in the category of technology demonstration, or production and deployment.

Interviewee: Okay, wait just a second, for that category, for the categories of HRL 2, human system performance design principle... Okay, we... Okay we say that the activity of the HRL

2 are achieved, which is one of the targets of the study in reality of course, most of the things are actually evaluated but of course they need to be validated so the effectiveness of this assessment, of the design principle that has been applied at that phase in in this work, need to be tested in an iterative way, no?

Interviewer: Yeah I understand. So you only established the principles but they still have to be updated and tested... yeah okay

Interviewee: Yeah we established the principle and implemented some of them in our platform, let's say at the general level no not only the supervisor tool but also the assessment of the way of working, the whole concept. So the supervisor, the controller, how the controller need to behave with the different modules and so on but then this assessment will need to be tested to verify the effectiveness and to verify also if the application of that concept has been well done, fits with the human capability and limitations okay.

Interviewer: Yeah, exactly, yeah. That makes completely sense, yes. And it will actually also come later in HRL so that's good. Okay so which category of which research phase would you say it's in, or development phase i should say?

Interviewee: Okay where they are listed?

Interviewer: That's good so that I know they are not so obvious. They are in the blue on the left.

Interviewee: Ah okay blue on the left, let me see...

Interviewer: Like blue marked but black letters, all the way on the left...

Interviewee: Ah okay, technology demonstration, research development, technology demonstration and production and deployment. But of course this depends on... these are three different maturity phase actually no? Because if we're talking about research and development, it's something that's still under research activities, research and validation phase. No sorry research phase no? The technology demonstration to me is something that already passed the research phase, it's something that is more close to the real application, it's focusing more on the technical aspect, it's a technical demonstration meaning that... I don't know if my background in SESAR is affecting too much...

Interviewer: No no no it's good, that's why I wanted to talk to you because I think this is very similar to the for example E-OCVM and I think it's good to see the differences yeah. So you feel like technology demonstration is more talking about technical aspects and doesn't apply so much to the HRL?

Interviewee: Yeah i feel that... I mean if you talk to me of technological demonstration in my mind, maybe due to my background, E-OCVM and so on, it's something that represents a further step compared to the research and development let's say and exposing more all the technological support, the development of technological support which is demonstration that means which is something well-stable, well-mature that can be actually considered something quite close to the real operational environment that can be tested and in which the part related to the operational aspect have been already consolidated in the previous phase so in that phase we are looking at the technical part, more on the technical part, still human in the loop, but the part related to the operational aspects are already tested, it's a further iteration phase but more mature.

Interviewer: Okay so do you think we should call the category operational and technology demonstration or do you feel that the HRL should just focus more on the first phase so the research and development and then... or do you have no opinion on that, that's also okay?

Interviewee: Because for research and development it's from TRL one until TRL three? Okay. Then from HRL 4 design concept completed... okay... but i mean if you... Why you ask to cluster the concept at TRL 3? At TRL 3 actually according to this three cluster it's still on research and development so the concept has been classified according to this three cluster before, not at HRL 3 no? This is my doubt.

Interviewer: Sorry, can you repeat your doubt?

Interviewee: You are now asking we are HRL 3, no?

Interviewer: Yeah we confirmed that HRL one and two were completed.

Interviewee: Okay but you already defined at the beginning if the concept was in research and development, technology demonstration or production deployment right?

Interviewer: Yeah so maybe we can jump some levels. If you say it was already in the technology demonstration phase that we can jump from two to five for example.

Interviewee: No no no, I thought here that in HRL you were asking me to classify the concept according to research, technology or deployment no?

Interviewer: Yeah, yeah.

Interviewee: Okay. I was thinking why at HRL 3? Actually this classification should be made at the beginning.

Interviewer: You mean the research or technology demonstration... yeah okay, okay.

Interviewee: So not at HRL three because at HRL three I'm just saying that this should actually be the final let's say level or the final gate for the research and development. Because I'm in line with the process that you describe here, research and development from HRL one till HRL three and then start the technological demonstration until HRL six maybe? And then to seven, from seven to nine is the deployment phase let's say no? Production and deployment phase. So I'm aligned with that but if you go through the phase HRL 123, automatically you are within that cluster

Interviewer: Yeah that is actually has a reason I will explain to you because maybe you can tell me then how it can be done differently because actually HRL one and two they are very important to make sure that you develop the right system, not that you develop the system right but that you develop the right system. So you should in all cases first finish HRL one and two and then you can say okay we finished HRL one and two but we are now doing a simulation so we're actually in HRL, I'm just saying something, so in fact you skipped HRL three and four but that is possible, you don't have to finish all the exit criteria if you already have a mature concept. You just have to make sure that at least at the beginning you got it right because otherwise the risk is that you're doing a simulation, everything is super good but then you realize that you made the wrong concept because it's not solving the problem. So that was the underlying thoughts but I can really understand that this is confusing because you go from a lower level back to more abstract level, right? You go from HRL and then back to the development phases. So I can understand that this is confusing.

Interviewee: Okay, but I understood what you say now. So you can actually, okay. Yeah, I fully agree with the relevance of the first two phases. Yeah. And then you can, okay, assess if, actually your concept need to go through the other two steps, or it's already mature enough to, let's say, jump to the phase. Okay.

Interviewer: Yeah, exactly. Yeah. If you have any tips on how that could be visualized better than you can tell me, because I think it's a good point.

Interviewee: Yeah, if you are going to use this kind of visualization, yes, here it's not the evident. This is for sure. So you have to find the some uhm yeah, different visualization to explain that. Yeah. Yeah, because in this case people go to a linear way of course. And I don't know if it is... no, it's not specified, neither in the different steps no? In the explanation of this column, no?

Interviewer: No, in the matrix itself no. So something could be added to show that at least HRL one and two have to be satisfied before you can move to another one. Okay. And going back to the development phases, the research and development, technology demonstration... Do you think, because I know in SESAR, we use definition, development and deployment, right? These three phases we use? Do you think something like this would make more sense?

Because I already heard you saying that technology demonstration for you is more about the technology, but we want to have general phases that contain both the TRL and HRL. So...

Interviewee: Because it's the technology word which make maybe a little bit...

Interviewer: So just demonstration would be better?

Interviewee: Maybe yes... You can say... You can say something like... not prototype demonstration, concept demonstration? or... It should be something that puts together the technological and operational part, right? I don't know. But I think it's the technology that makes more the focus on the technological part, obviously. Because it looks like as if in the first part, you are focused on research and development, which is both actually, you don't have any clear address for a clear focus on that. But then you focus on technology demonstration. So clearly focusing on technology.

Interviewer: Okay. So maybe demonstration would be better or concept demonstration?

Interviewee: Concept demonstration or design... no, no. Or system? Very banal!

Interviewer: Definitely, if it works. And do you think it's better than if we would change it to definition, development and deployment? So research and development would be the definition phase, then technology demonstration would be the development phase... What do you think? Or do you think these categories fit better? I'm curious for your opinion.

Interviewee: Research and development then you say demonstration and then deployment?

Interviewer: Yep. For example.

Interviewee: Research and development will be the concept definition actually, no? Okay. Both can fit, yes. To me both can fit. Because actually the stages are those that are mentioned in the title actually. Research and Development makes more clear than concept definition which is really high level. Here we start from a research so we can have also something really new. From research. Then development. That is new or something that is still under development. Maybe it's not so new, but it's on definition phase. Here you can have both. Demonstration is something under more accurate work phase, investigation phase, validation phase. And then deployment, obviously, it's something that is ready to be... it's on the way to be in operation.

Interviewer: Yeah. Okay. Good. It's good to have your opinion on this one. Okay. So, and would you say that this pj 35 is in the research and development phase? Or is it in demonstration?

Interviewee: Let me see how do you... I would say that...

Interviewer: Tell me, if you have any doubts just let me know.

Interviewee: Yeah, I have one doubt. Because if I look at the HRL. I didn't go honestly into the detail of the description. But I cannot see the validation aspect highlighted. Because if I had to say that the HRL two or the HRL three for example has been accomplished, I have to test it, I have to validate that part, no? I have to see if... because I set something, I defined something, I defined for example the design principle, no? And I design the setting, my system, according to the performance design principle, so I make that the... for example, for the 35, that the controller cannot manage more than three airports because it's out of the human capability, no? Research told us. Human system performance and comparable system analyzed... Okay, so this is something that is related of course to the first claim and then I also identify the metrics, make some benchmark, make some comparison, I identify indicators that can give me a feedback on the goodness of the things... but then I have to test it, no?

Interviewer: Yeah, okay.

Interviewee: A gate for that, to know if I actually go through the next level. It's something that, again, maybe my background in SESAR make it... I'm like this by myself. But I mean, if we consider this phase as just paper-based let's say, in HRL two, okay, so I accomplished the phase but then, in HRL three, I have to test it. I have to!

Interviewer: Your design principles?

Interviewee: Yes. I would expect that... I don't know... Maybe I'm not really in the loop of this.

Interviewer: No no, I am really interested in these things because actually I don't care which HRL project 35 is in, but I really want to know how you use this thing and what is not clear. So, yeah, the idea of the HRL is that you first establish design principles, you come up with requirements for the system, and then you're going to make a prototype and then you test whether the requirements and the design principles work or whether they have to be updated.

Interviewee: But in a time when you... I mean, please, since the... When you define the design principle, in some way, you also have to have an initially idea of requirements and system and so on, no? I mean, this phase maybe goes quite in parallel or in any case... Wait, let me have a look into the details. Because in HRL 5, you arrive to the prototype. So in HRL 5 you validate actually the prototype that you develop here, until here, right? But before you didn't test anything, no?

Interviewer: Yeah, I guess maybe in HRL 4...

Interviewee: Because until HRL four it's all paper work. I'm a right or am I missing something?

Interviewer: Well, in HRL 4, the design concept is completed. So maybe you complete the prototype in HRL 4 and then in HRL 5, you test it in a simulated environment.

Interviewee: But do you foresee any step in which you are going to involve users or potential final users in one of these steps? At some part there is workshop.

Interviewer: Workshop, yeah. So the workshop starts already at HRL one because when you have an idea of a concept you can already ask pilots or air traffic controllers or whoever is going to use the system, you can already ask: What do you think? But you are correct that at that moment you don't have a prototype yet.

Interviewee: No, okay. But you have, of course, at that time the workshop helps you to define the idea, the concept, of the system, the system you want to develop. Then, you define the initial, you apply the design principle, okay? And you start to think how the design principle can fit into your concept, in your system. So you use use case analysis, storytelling, okay. Okay, to me also that part can be in some way shared with the users, it's sort of a second iteration, no? So, this is the use case with maybe a storytelling etc., I can have feedback on the goodness of the design principle that are applied but also on the needs of my user that have been covered, initially covered at least, no? That they the initial requirement can be in some way on the good way to be covered, let's say, so, at that stage. Then I go through HRL three - system requirements established...human capability and limitations investigated. When you say investigated, to me these aspects are very relevant, and something that at that time are investigated already three times, like in three iterations! When you say "investigated" here, what do you mean? It's investigated... here there is no reference to uhm...

Interviewer: But how would you investigate them? Would you do a workshop? What would you do to investigate the needs?

Interviewee: We are in HRL three. If we think that it should be in line with the TRL three - experimental proof of concept - that means that at that level TRL three uses a proof of concept to assess an initial, not of course real time validation, but something more light, of course, but still that allows me to investigate if the human capability and limitation and needs has been taken into account, so, something that involve the users, for example, it could be a workshop with a low level prototype, Wizard of Oz... Which I am not sure if it's... yeah, it should be at that level, it's a prototype. It could be also not a working prototype, but something that allows an initial interact interaction with the users. And here, of course, we are talking about some things that are integrating the system, I'm thinking that systems that are potentially integrated in the Aviation/ATM/ATC field, no? We are talking about project 35, but it can be something uhm a system useful for a controller, for example, something high-level. So our first validation, our first experimental proof of concept could be something that allows controllers to interact with PC-based prototype or...

Interviewer: So instead of doing that at HRL four or five, you would start doing that at HRL three?

Interviewee: In which sense I start...? Yeah, yeah, yeah! I would implement at that phase something that is more... that allows me to actually investigate the things that I listed here.

Interviewer: Okay.

Interviewee: You see, *feasibility of metrisc for successful human system performance evaluated* How do you measure the feasibility? How do you know that that metrics are feasible? By research, by paper? Or by something more concrete? This is my doubt but it's not a critique, I just try to understand how it's supposed to work. Because there is a... I'm more familiar, I have to say, with the E-OCVM scale, no? But I know that the TRL is quite overlapping for some part and this is supposed to be for E-OCVM the core part. V one, two and three are considered the core part, the more relevant aspects, no? In that phase, in E-OCVM you have several iterations at different levels. You know, it's a continuously iteration made with the concept which is actually already...in the worst case, it's at V1, no? But at V1 you already start with something that can be validated, with qualitative results, with workshop, focus group, and so on, because it's not mature enough of course but then at v2 and v3 you start with interacting with the concept to validate and test if it's feasible, if it's workable, if it's supportive for controller and so on. It's something that in this description at least, it's seems to not be reported. So my doubt is that it's not reported here but maybe it's hidden below...

Interviewer: So it's not clear whether it's iterative or...?

Interviewee: Yeah it's not clear if it's iterative, it's not clear how they measure that it's mentioned here in the description of human in the system for example how this metrics, this capability, needs, limitations are investigated or, maybe it's my fault, at that phase there is no need according to the HRL to investigate in concrete, it's just a description.

Interviewer: Yeah, yeah... that's a good point because I think that's the confusion because they just want you to come up with some initial requirements, I don't think you already have *the* requirements, just something that you can then make a prototype with because to make a prototype you first need some principles, requirements,...

Interviewee: Yeah this is fine but then this requirement, especially if they are initial requirements, goes in parallel maybe with the design principle establish and cannot be... I mean, if you set both of them then you can define a preliminary feature to accommodate the human capability limitation and invest the need, investigate the feasibility of metrics.. you can then test this part no? It is also this for example *potential limitation ... human system performances risk* I think... I don't know, uh, yeah I think it should be, is there the need to test this part in your view, because you are building this scale, it should be highlighted because here it's not clear, I mean to me it is not evident. It seems that uhm, it is like a very detailed scale but actually this step or this level of maturity is not so uhm is not so I mean they are of course different steps but they should be arrived at one step in which all these things are validated to be then say okay this is the following step, this is the more mature step. You know what I mean? I say it in a very bad ha-ha.

Interviewer: So you mean that it's not clear that the next step is more mature than the step before?

Interviewee: Yeah because I couldn't yeah, I think that for each step, I mean at least for HRL three you have to plan something that gives you a gate that okay here I need to test with the users, the final user, with something, with someone, that the things that I performed until now are correct. If they are in line with expectation, fulfill the requirement, the needs, and the metric that I established are correct for my target, then I can make this initial validation, define you the level. There is no need to have, of course, a real time no? But an experimental phase that need to be tested if I'm on the right way, and then I can start again. But in that time, I can say, okay, I reached my initial gate, which is maybe the research and development gate no? And then I can start to do the demonstration phase, the demonstration part. So maybe you can add some sort of gate on which you collect all the inputs that you, all the outputs that you had in that three phases before, make a gate, validate at that time, because maybe the validation before has been made only by workshop and focus group which is correct because it's an initial and early phase no? And then from that, once the gate is achieved, you can start with the demonstration phase, on which the relevant part is of course, now the validation part, because we are going to a more concrete part of the concept definition, consolidated, there you have to validate it, for sure, which concept, and then you can also apply some setting, which is more close to the real one. Until the deployment.

Interviewer: Yeah, and how, because, in fact, the criteria for HRL three are the gate criteria to go to HRL four. So.. uhm because I don't know if you noticed that at every level, actually the topics are the same, the questions are the same. But instead of testing it in a workshop, you will test it in a simulation then you will test it in operational environments, etc. So they should follow each other up. And, and actually HRL three, if you finish that, it also includes kind of HRL 2 and HRL 1. Because to develop user scenarios, you first need to define who are the users. And yeah, if you if you get what I mean, so I'm trying understand what...

Interviewee: Yeah yeah yeah I am reading now yes, yes. Okay. I got it. Yeah, so maybe you just need to make more evident the gate and maybe to make it concrete, associating the gate to some activities, some toolbox, as you call it here, which is something that needs to be assessed, like experimental concept test, or initial validation, early validation of the concept or something like that.

Interviewer: So, like the outcome, after HRL 3, what do you have?

Interviewee: After HRL three, you have the... what is made... what is written here in some way. So you have that the concept you have an initial input that the concept is the right one, that it is good for your requirement and needs that you defined at the beginning in one, two, and also three. So goodness of requirement, not goodness, achievement of the initial requirement that you define, goodness for the needs of the of the users that are in some way fulfilled that the indicators and the area that you are going to test and investigate are the right one and investigate with the correct metrics that give you some reliable input. And you can also have aspect of the initial risk identified and maybe tested in that phase.

Interviewer: And testing doesn't belong more in the demonstration phase? Because we are still.. until HRL 3 we are still in just the R&D phase.

Interviewee: Yeah but at that time you can have this first testing and risk in terms of, I don't know,... if the system is completely and understandable by controllers for example. You can, you know in SESAR we identify specific issue for each of the area of our investigation no? Or the role, responsibility, communications, task sharing and you can identify for each of them risk - and I'm not talking about the risk assessment, safety risk assessment or something like that, just potential issues that can be identified at that phase and tested at HRL 3 for example, which is mature enough to be addressed this kind of potential issue as well as benefits no?

Interviewer: Yeah, okay. And if we go back to project 35 then what would you say where is it now? Which HRLs did it complete? Did it also complete HRL three or four or...

Interviewee: Uhm, let me see. I would say that we are currently evaluating a prototype but at quite a high level in a simulated environment. So, if I had to read the description of the HRL I would say that we are at least on five. Evaluation prototype in simulated environment no? Because the design concept has been completed yes *system require...* yeah yeah yeah *design fully matured* I will say no, because it's something that of course, needs to be validated. And so HRL 5 maybe it's the stage that fits well with the...

Interviewer: And if you look at HRL five, did you do all the things that are stated there?

Interviewee: Okay, let's say *potential human system performance issues and risks identified* *conformance of prototype to human system performance requirements and principles verified* these are for me a sort of exit criteria no?

Interviewer: Yeah, they are exit criteria. So to complete HRL five you have to fulfill all those criteria. Yeah.

Interviewee: Okay. So, what we are doing now is to assess, plan and assess, all the activity that allow us to do that yes. *...potential human system performance...* yes, *conformance ... verified* yes, *strategies to support human use recommend*... Yeah not only recommended but identified, tested, integrated into results and then define the requirement for that, for the users, and recommendation yes. *Human system performance metric successfully met* yes performance metrics means that I identified the area of my investigation, identified the indicators of my performance of course, and I found out the rating that I achieve are completely feasible, are completely acceptable. So the performance can be supported, guaranteed in some way.

Interviewer: Yeah for example, the workload is not too high or these things..

Interviewee: Exactly, the work is at an acceptable level and then we can discuss about acceptable words but okay. Yeah, this are the categories no? The human factor categories?

Interviewer: Yeah.

Interviewee: On which is ... no *strategies to address environmental constraints updated* How we can define this claim? *strategy to address environmental constraint updated, strategy to accommodate manpower and personnel updated* So this is yours? These are also let's say exit criteria for each of these areas. Okay. *strategies to address environmental constraints updated*... Okay... Okay, yes, this part of training and personnel, manpower blah, blah blah, for example, are according to the E-OCVM and the assessment that we used to do in SESAR are something that are investigated at a higher level, like starting from a v3 no? Because it's something that of course that need that the concept is well consolidated. Uhm *procedure for operator and maintainer role updated, task flow and sequencing optimized, mechanisms to support ...* Okay, yes!

Interviewer: Okay. And that is what you have done, or what you are planning to do?

Interviewee: This is what we are planning to do at that phase. But it's also things that we plan to do in the previous phase, the things change, is the deeper detail of evaluation of the investigation performed no? In the previous phase, you can do it using qualitative measurement, and not maybe addressing specific quantitative indicators, but maybe, aiming at v3, or the demonstration phase, let's say but, to be more aligned with the current activity at v3, you have to do it at qualitative, quantitative level, you have to address also specific areas that are impacted by the concept, of course, at quantitative level, and you have to be within the framework, the target framework that has been established by the project, okay? So you have to demonstrate that the level of capacity is reached according to the target, or the level of safety is not below that target, for example, and so on.

Interviewer: Okay. And before you did that in workshops, and now you do a simulation or...

Interviewee: Before we did also simulation but according to a different level of maturity of the data collection, for example, no? And early before we can do it at workshop level, for example. It always depends on the maturity of the concept, of course, no? And in SESAR with ENAV it happens maybe, maybe really at the beginning, so more than 10 years ago, some time we perform workshop and then validation within the same phase of project no? Because sometimes it can happen - usually happens actually, but sometime it's more relevant you have to do something, more evident you have to do something - some aspects of the concept are less mature than others. So, what you can do, you can combine the approach and investigate for the less mature aspects use the different technique like for example, a workshop or focus group involving the user and so on, and from the others or once collected the output of that you can implement it in the more mature methodology using the real time simulation or make it a sort of parallel path for both of them: one more mature and one less mature. Some time it happens.

Interviewer: Yeah. And the idea in HRL 5 is that you do a simulation with prototypes, so could be individual technologies. And then at HRL 6, you really, you test the system as a whole. So you do the real time simulation with all the features, all the aspects. So, yeah, I think that is what is... maybe that is not clear enough...

Interviewee: No, no, no, this is clear. That's why I told you the project maybe is in HRL 5, no? Because it's clear that here you are going to the evaluation of prototype in a simulated environment, and to me it means that you are going to a classical, let's say, real time validation, for example. And all of these points here are something that need to be reached, that need to be achieved, to reach this level, the HRL five. Okay? So it is the work that has been done, that needs to be done within that level, to reach this level.

Interviewer: And so are you saying that maybe HRL 5 consists of multiple iterations? It's not one simulation, but it's multiple simulations after one another? or?

Interviewee: Yeah, I would say that it is part of a different iteration, no? Maybe I would expect to have something before, a lower level of maturity, but that gives me an input on the goodness of my system and on the goodness of the way on which I'm capturing requirement needs uhm performance aspect of the concept and so on. Yeah, I would expect an iteration before again at lower level that there is no need to have of course, all the system, all the work environment working properly and so on, but just to test the aspect on which I work on it until then, until that time, and then I will, with that output that will go further defining a more mature system working on the HRL5 and test it.

Interviewer: Yeah, okay. So you just need more... Yeah, you need more feedback, more testing earlier on in the in the process?

Interviewee: Yeah, yeah. To me at least one in TRL three... yeah... three and four.

Interviewer: So related to that, what do you think are the weaknesses and the strengths of this framework? We can start weaknesses because you...

Interviewee: Weaknesses is that one, that maybe we will need more iteration and more test, more validation part because I detailed in uhm... the strength is that it's very well detailed and each phase and for each phase you have a detailed description, no? It's well detailed each step but no? This is good because of course you have clear which are the expectation for that phase no? But what I think maybe can be improved is more... it's to provide more testing phase, real testing phase that involve the user, at different level of course according to the maturity of that phase no? And some gate, the let's call it gate, we can call with a different name just to differentiate from SESAR.

Interviewer: And maybe it's good to call it gate because you.. people know that word.

Interviewee: You can also adopt this approach, yes of course. So introduce some gate that clearly identify the potential end of one stage, so one cluster, the ones that we mentioned at the beginning: research and development and then this demonstration phase, technology demonstration phase, and then the deployment phase. But that also clearly then allow us to identify the work that has been done has been done correctly and in the good way, in the right way. Because it's yeah, it's a, you know, there is a two-fold benefit. Also for the researcher, that needs to have feedback on that, on the activity performed.

Interviewer: Yeah. Okay, some kind of assessment to see whether you did the right thing.

Interviewee: Yeah.

Interviewer: Yeah. And if those weaknesses would be resolved? Would you use this? Because I know you already have the E-OCVM and the HPAP? Would you use something like this?

Interviewee: Yes, it could be used, within SESAR with a very, very strong dissemination campaign. We have to convince people to use it. Usually in this field they are quite conservative. But, yeah, to me, maybe uhm... Because what we have... I would say that the focus on the maybe to... maybe the focus on the... the added value of this scale that could be this detailed part of each step and also the focus on the human could be more highlighted no? Because this is the difference maybe compared to the other scale, which are looking more on the system, the TRL.. Because E-OCVM is looking to both as well, no? System and human. Maybe it's a little bit more balanced on the system part. TRL is looking at technologies, and this should look on the human part no? Maybe all the activities should have a major focus on the human part, the human operator no? In terms for example of this suggestion to have a continuous test and iteration involving users could be a part of characteristic of this scale, because it's a human-centered let's say you know, compared to the others. It's something that could be highlighted and also for the toolbox that you took into consideration. Something should be more focused on... I mean, some of them, or maybe most of them should focus on the input from a human or to test the human part. So all the tools and methodology that focus and that put the human in the loop obviously they should be the toolbox preferential. Just to, you know, to differentiate because of course, this story of the conservative people in that field is quite true and I guess that the use of E-OCVM is quite consolidated. They are introducing in the recent years this TRL in that field because TRL is...

Interviewer: Okay, they're still introducing.

Interviewee: Yes, because they have to fill this gap with the more mature concept no? In order to, instead to readapt the E-OCVM etc. They make it some stage like reference to the TRL part no? So I think that to use in SESAR, then for others of course this could be a great reference because it's a scale that can be used for also different aspects, but to be used in SESAR for example it should propose something more, that makes some difference, provide some added value compared to the other no? This can be true in all fields but in SESAR particularly mainly because they have this consolidated use of the E-OCVM methodology.

Interviewer: So, you feel like the benefit of the HRL is that it focus more on, really on the human. Which is good, which is something more unique, but it should be highlighted more in the framework. The methods should be more human-centered.

Interviewee: Yeah.

Interviewer: Yeah, okay.

Interviewee: Yeah, it should be real more human centered, no? It should be... maybe to facilitate this discussion about iteration and validation blah, blah, blah, could be one point, the use of specific toolbox could be another point. Also, the... For example, here, these categories are all focusing on the human actually *team and communication, procedures roles responsibility, training and development, organization and staffing, working environment.* Yes. Yes. *Task, aspect related to the task, task capability, potential limitation, ...*

Interviewer: I think tasks now fall under procedures, role and responsibilities.

Interviewee: Ah okay. Okay. And also interaction aspect falls here. Procedures, roles and responsibilities.

Interviewer: Yeah. And teams and communication also talks a bit about interdependence, interaction. Yeah.

Interviewee: Okay.

Interviewer: The interaction with other team members would be teams and communication and the interaction with the system would be human in the system. I would say. And if we look at the toolbox, because you talked about it, are there any tools... because you said it should be more human centered, are there any tools that you say for example: this is not human-centered ditch, or: I am missing something?

Interviewee: Yeah, no, no. Gems?

Interviewer: *GEMS*... I don't know from the top of my mind to be honest. I would have to Google.

Interviewee: Okay, okay okay. Yes, those tools, questionnaires and methodology, because I see also focus group... yeah, are, all of them, it seems to, no are involving humans of course, yes. Also... Yes.

Interviewer: So, they are quite human centered right now?

Interviewee: Yes. Yes, according to me yes. For example, at the beginning also you can in the early phase, you can also add interview. You can also add cause uhm... I don't remember uhm... sort of like field study, but there are some specific methods. I don't remember...

Interviewer: A bit like more ethnographic like...?

Interviewee: Yes, they are usually enclosing the so-called field study, but

Interviewer: Yeah, okay, I can look it up. But I know what you mean, it is when you go and you observe the operator, how does he do his work?

Interviewee: Yeah, yeah, yes, yes. Like task analysis for example, I don't know if you put it in some stage.

Interviewer: Task analysis should be there at the beginning.

Interviewee: Maybe this one *ConTA*. Yeah, at the beginning to identify the task flow. Yes. Potential criticalities and so on.

Interviewer: Yeah. Okay. And is there anything else you would change to the toolbox? Because the philosophy right now is really like, okay, these are the exit criteria that you have to answer. And these are tools that you could use, but you are the expert. So you decide how you do it. That is the philosophy behind it. But is there anything you would like, for example, there in the toolbox? Are? They are maybe more questioners, they are, like, for example...

Interviewee: But questionnaires you know there are a number of... There is also for example, for again, for aviation, of course, it's my field, so I can make a reference on that, there are a sort of repository provided by Eurocontrol and SESAR, maybe jointly, I don't know, I don't remember, which report several, a number of tools and questionnaires for each categories of investigation and concept of investigation, like workload, situation awareness, and so on. So they are quite big work done, also in the past, to collect all these references.

Interviewer: So you don't need that here?

Interviewee: Yeah. Uhm made maybe it's not necessary, it's not really worth it, to provide the same here, to make the same job.

Interviewer: Yeah, maybe would it be nice to have it integrated in here, so that the repository of Eurocontrol is somehow integrated in the HRL?

Interviewee: Yeah, yeah. Make a reference on that? Yes. Yes.

Interviewer: Okay. That's good. But yeah, so no need to classify these tools and everything because it has been done...

Interviewee: But the classification now, if I remember well, in the platform of Eurocontrol was made with several cross references, namely that you can have a look to the concept and see how which are the potential to investigate it, but you can also have a look to... not indicator... how they... wait... I try to see... I mean, it was built quite well because you can make research from different point of view. Now I really don't remember the topics, the cross references that they made, but you can search the, for example, the question of your interest can be... or the area of your interest can be investigated from a different point of view. So it's made quite well. And yeah, here uhm, I mean, the cluster or classification that can be used here at work as we say, as I said before, they look at the human part. But actually here, all the methodology are already on that cluster. So already belong to human cluster so I don't know if maybe you can classify qualitative, quantitative or you can uhm... But I don't know if it's necessary. I don't know.

Interviewer: Okay. Clear. All right, I will go to the last questions, because I see that we are over the time. So I want to ask you, do you think this framework is easy to use or complex to use?

Interviewee: It's definitely easy to use, because the philosophy on which it is build is clear. Here, the step, here are the exit criteria that as we said the tool that you can use to address that aspect and the work that you have to do within each phase to reach that. Because the work you have to do within each phase is not... is defined in term of target, right? In terms of reaching of achievement of each criteria. So to me, the philosophy is quite easy to understand, because it's actually straightforward, because it's the work that we do for each time, it's a very well-known process.

Interviewer: Okay. It fits well with the practice that you're already doing?

Interviewee: Yeah.

Interviewer: Okay.

Interviewee: Maybe I would highlight that I'm really familiar with this. Maybe someone which is less familiar with this could ask you: Yes, but for each area, I would like to know, which are the activity, for example, foreseen to reach that target. And this is something that, for example, for the E-OCVM has been detailed for the HP material in SESAR for example. For each phase you have the entry criteria, the work that you have to do and the exit criteria like...

Interviewer: Yeah, exactly. It takes you more by the hand?

Interviewee: Yeah, yeah. Exactly. Yeah. To me, this is feasible as well, it's clear what you have to do. But again, maybe there is this bias to me, so yeah, less familiar people can ask you, I will need at least a schema for each categories. For each category, which are the expected activities. It could be an idea to improve no? To make it more workable, usable for a larger audience.

Interviewer: Yeah. And what is the larger, because you are let's say senior expert...

Interviewee: Just think of people that are starting to do this kind of work. That's why I highlighted several times maybe my background can make some input that are too much linked to SESAR, to E-OCVM, etc. But people that are that are starting with this activity now... In our... let me think...

Interviewer: Like for example, take me, I am a human factors expert, but I am super junior. Of course now I understand this framework, but imagine that I didn't do my thesis on this, so I have no idea and then, like, how long would it take me to learn this or what do I need to know before I could learn this?

Interviewee: Yeah, exactly. Maybe you need to learn which activity can be performed to achieve each target or each categories? Yeah, because you are structuring by categories. So for example, *procedures, roles and responsibility* in SESAR, what has been done is to define a sort of leading questions for each category no? So your task is try to reply to each macro question and then sub questions that allow to go into more detailed part of and explore more into detail that area. Or you can also foresee maybe uhm make a sort of a mapping of for each category, what kind of activity can be, but I don't know if it's feasible by category.

Interviewer: Yeah, I like the idea of a leading question that can also clarify what the category is about for people that are not so familiar with them. Yeah, I think it's a nice idea. And if we go, for example, to HRL five, because I want to understand one more thing, because you say it's not clear what is the activity, it is something you have to infer yourself, but, for example, HRL 5 says that you evaluate a prototype in a simulated environment, which I think is also a form of activity, then there's the exit criteria, so the targets that you want to achieve, and then there's the methods. So if you talk about activity, what do you mean? Because a method could also be seen as an activity, like doing a cognitive walkthrough could be seen as an activity. Just to be clear, because I think you're right, it's more defined in targets, but I want to know what would be the activity that should be added

Interviewee: We are talking by category or in general?

Interviewer: Per level I would say. So, what would be an activity in HRL 5?

Interviewee: In HRL five if we are aiming to evaluate prototype in simulated environment no? If we consider uhm... ah no... if we consider uhm... but maybe I was thinking we can consider it category by category...

Interviewer: Yeah, just however you prefer.

Interviewee: As activity you had to... no maybe no. For the HRL five, the focus is on the evaluation no? So, the main activity that I foresee to cover, to reach this HRL and achieve these criteria, things that I will think about is to assess a real time valuation for example, because in this case, I can... *human systems performance metrics identified* so I have to identify the scenarios, identify the metrics, identify the reference and solution scenario for example no? And *environmental constraints, accommodate manpower* and within, this is a higher level no? At lower level, I will go through all these aspects: the working environment, organizational staffing, training, procedures and communication for each of them I can define tailored questions for example, or for procedure, role and responsibility tailored scenario to investigate the specific new procedures that for example, I have to introduce with my concept or specific scenario for exploiting, to test the team and communication for example, in the 35 for example, you have to solve them, you have to test and have a look to some specific aspects related to team and communication and uhm... So, the activity could be to... all the steps that then bring me to the validation activity as well, the step before. So: define the metrics, define the requirement that I want to test at that stage...

Interviewer: But that happens in HRL 1, 2, 3, 4 right?

Interviewee: Yes, yes, but you define there the early requirements no? And then you test it. You got some output and maybe some requirements need to be redefined, some others has been closed, some more still open. Because the concept was not mature enough to be addressed at that stage. So, at this stage I will collect all my requirement, my previous ones and I will start from that. For each of them, I will define impact in terms of area that can be investigated; could be HP, but could be also aspects related to safety could be aspects related to... maybe we go out of the path in this way, but in any case, okay. Related to HP and so on, then I will go through uhm try to identify potential human system performance issue and risk, no? You have to. So, identify issue and benefit for each area. How? Through a workshop involving also controllers, then, once I identify that I can define objective okay? The objective of my activity. First the expert will do a draft then again check with the expert and with the... maybe also with your operational, the technical people because the system should enable the validation of this aspect no? How? Again, workshop or a focus group. Once the objective has been identified I can look at the matrix identification of indicators and so on. This can be done internally by the expert for example, without the need to actually involve others because this required specific expertise. And then of course, there will be... this of course implies the investigation of all these aspects no? Because if you define the area, for each of them, you have to define how they will be analyzed. And then I would expect to have some technical session to verify the system will work properly. And then I will define the experimental plan, as we said before, with all the actors involved, all the scenarios, all the metrics, all the actors, the users, of course, that are the actors, of course, and blah, blah, blah. I will run the validation, I will collect the results. This is the activity, step by step, that you can perform here. And the focus should be of course, then on the results because it's expected to provide also a refinement of the requirements that has been provided as input no? And provide requirements or recommendation for each of them. Then go through the steps, to the following step maybe no?

Interviewer: So to see if I get you correctly, with the activities you mean more in detail what do I do and it also contains more than project management steps like okay I contact stakeholders, I draft a plan for the validation, ...

Interviewee: But maybe the ones that I mentioned are not really at manager level but are more technical ones I would say because yeah the involvement of the users are directly involved in the exercise, they are part, they should be part of the validation of the assessment.

Interviewer: Okay so it's more the practical things that you have to do?

Interviewee: Yeah it's more technical part that maybe can be useful to use it. Because this is not at the management level no? This scale is a tool, it's more for practical aspects...

Interviewer: Yeah, I was wondering because you say the activities are not there, only the target and the methods, so I was trying to see what are the activities. And I feel it's more all the steps that you take to... like you say, you look again at your requirements, you update them if it's necessary, then you...

Interviewee: Yeah but all at the practical level. It is something that the HP expert for example can be done technically to perform the activity within that stage, to perform also the validation exercise for example. At higher level the activities is more on coordination part, it is not...uhm yeah maybe really at the beginning, at the higher level you can involve of course stakeholders, you can involve managers from dependent project or similar activity to define - and technical part of course, technical supplier - to define the initial requirement at that stage but then for the application of the system requirement, the users requirements is something that needs to be maybe defined at a lower level, not really a technical level, meaning with the involvement of the HP expert, the technical expert, not at the manager level. In this case maybe I would refer to the definition of this aspect at project level in SESAR for example or at least, or maybe also at SESAR level really when they are going to define the different general target for the project, for the area and so on. It's true also that the during the project of course you can involve also the stakeholder, you can define the... uhm you have a table at the beginning of each OSED or VALP in which it's listed the needs of stakeholders, of space users, of the technical company and so on. But they are not directly involved in that part.

Interviewer: They don't take place in the simulator or...?

Interviewee: No, no no, just as a guest if they want.

Interviewer: Perfect, okay, well I have had a lot of feedback so it's really good, so my only question is whether there's something else you want to add, something that I haven't specifically asked about?

Interviewee: No, my only concern is if we talk too much about... having SESAR as a reference. I don't want to that maybe this have a strong impact because of course it's something which is not really uhm I mean that can be viewed also from another different point of view, maybe less related to the SESAR and to aviation so I don't know if you had the occasion to talk also for example with uhm because the framework on which this project has been defined is aviation itself right?

Interviewer: Actually it was built based on, I think military and nuclear power plant, that was the background of the researchers and I took the HRL that they developed and I changed some things and adapted it more to aviation because I was in project Safeland so I was more concerned with aviation.

Interviewee: Okay so you want to stick in that field?

Interviewer: No, I am talking to you but I'm also talking to [other colleague] so...

Interviewee: Ah okay, I was going to suggest her.

Interviewer: Ah that's perfect. Yes a different perspective. Yeah that's why I didn't mind to talk with you a lot about the comparison with SESAR because that is your expertise so it's really something that I can get from you. And then I will also try to talk to people from industrial partners to see also what they think about it because I guess it's completely different from Deep Blue.

Interviewee: Yeah okay okay, so that's fine because yeah [colleague] was the person that I would have suggested you to have talk with.

Interviewer: That is good, I speak to her on Friday. And why would you suggest [colleague]?

Interviewee: I would expect that she can have a different approaches. She has a different background compared to mine; human interaction, design and aspect that I approached but more latterly let's say because they are part of some activity that I used to do or that I studied as well but are not my daily work directly. While for her it's something that she daily manages. And then she also has to do with aspects related to human machine interaction and validation but in a completely different field, both in manufacturing in recent years but also in European projects that usually has a more wide eye on this on this aspect, and not only focus on the aviation which is my specific field. So, okay good, you will collect also different eye on that.

Interviewer: Yeah yeah. I think that's good to compare. Yes, all right, thank you so much. I mean we spent like two hours of talking.

Interviewee: Thank you, it was very interesting. I hope it would be useful for you.

Interviewer: Yeah, definitely. It is very useful.

Interview with aviation/manufacturing expert from Deep Blue

Interviewer: Okay. So I will just go quickly through what is it again, the human readiness level. So it's a scale that is supposed to help with structuring the process of validation and development of systems. And it's based on the TRL but it's supposed to be the human factors alternative for the TRL. Alternative is not exactly the right word, because they are complimentary, you can use both TRL and HRL. And it's supposed to be used by project managers, or program managers, and system developers, human factors practitioners, validation practitioners, these kinds of people. And it's applicable to very complex high risk, social technical systems. So not, I don't know, a smart glasses, but really like a system, a complex system. It is supposed to go in line with the TRL. So in the ideal case, where you are TRL three, you're also supposed to be in HRL three, just to make sure that you apply this human factors perspective from the beginning of the development onwards and you don't end up doing first the technology and then the human factors, which we know that sometimes happens. So the earlier it is applied, the better also, because then it can have the maximum benefits; it can prevent that you will spend a lot of money or time later on in the process because you have to change things; and it's helps you also with planning, like project planning, making decisions; and it can help you to mitigate the Human Factors risk. So that is really shortly what it's about. And I would like to ask you for a project that you are involved in or that you were involved in, that fits this description. So it was about a complex, high risk socio technical system. Yeah, is there any... I mean, probably you did many projects that could fit in this category, but is there any project that's springs to mind that you think the HRL could be applied to that?

Interviewee: Now I'm thinking a bit because I had a bit of a problems with my network, I switched to another one. So I heard you bit and pieces, but I okay. So, you said the type of projects that you can apply this human readiness levels method should be complex projects, right?

Interviewer: Yeah, should be about a complex system.

Interviewee: So would a project like uhm research project like uhm a single pilot or a complex operation in all weather conditions...? Because probably these are the projects that are more similar to this type of definition. Because if I think to other projects, I was involved... okay... I mean excluding the consultancies, the TetraPak consultancies, which I don't think fit... but then it's probably different levels. So probably the best is to think about one of those projects, which are a bit old probably, but that relate to the all condition of operations. And or maybe management of incapacitation of the crew in the flight deck.

Interviewer: Yeah. So which project are you thinking about? Like, is there a specific project? I mean, for example, SAFELAND fits also in this category that you described...

Interviewee: Yeah, it's probably the grandfather of SAFELAND. It can be ALICIA or ACROSS. You might have heard of that.

Interviewer: Yeah. Yeah, I know them. Okay. So...

Interviewee: Somebody already went through them?

Interviewer: No, actually I went through them. I analyzed some of the projects in deep blue to see how they follow this structure. But it is good, I think, we can go with that. Okay, so ALICIA and ACROSS and they were both focused... You have to refresh my memory...

Interviewee: Yeah. Alicia was focused on all condition operations. So basically being able to fly an aircraft, even when weather conditions were not perfect. With the support of a number of enhancements view technologies and tools, software and tools, like, you know, those head up panels to enhance, let's say, to augment the reality. And that was Alicia, whereas Across was, again, technologies to somehow manage the possible incapacitation of the crew. Due to, again, bad weather conditions or problems on board due to the medical problems to one of the two pilots. So also incapacitation of both pilots.

Interviewer: Okay, let's take one, for example Alicia. And then we go through the steps of the HRL to see how we could assign an HRL to that. So here are the steps; define the system. Well, that's the Alicia system. I think there were multiple configurations right? There were multiple concepts. So I think for this exercise, we would take one concept and not Alicia as a whole, and then the first next step is to determine whether the exit criteria of HRL one and two are fulfilled. If they are not, these activities have to be planned to achieve these exit criteria. If they are, then we can continue with step three. Then step three is to determine what cluster this concept can be categorized in. So there's three clusters; research and development, technology demonstration, and production and deployment. And all three of these clusters have three corresponding human readiness levels. So we first choose the cluster and then we choose the readiness level that is appropriate for this Alicia concept. And then the next step is to consider whether the exit criteria that belong to this level that we that we selected are all fulfilled because if they are not, we have to move back one level lower to see if the exit criteria of that level are fulfiled and we continue until we found the level in which the Alicia concept fulfills all the criteria.

Interviewee: Okay.

Interviewer: That's how we can determine the HRL. Now you can move to the next page. And I will also keep the steps here so I can help you a bit. So the first task was to check whether the exit criteria of HRL one and two are fulfilled. So you can try to check this and just let me know what you think, if there's any problems for you to find out where you have to look, these things are interesting for me to improve...

Interviewee: Now what I'm thinking about is one specific concept and that is the augmented panel, that is a panel that is based in front of the pilots, and it allows to project on the reality some additional information that are not possible to see because, for example, of bad weather. So, fog or snow or rain can impede the full visibility, it can be a problem when an aircraft has to land. So having this augmented reality system can help landing the aircraft even in bad weather conditions. This one thing, if you... I mean, otherwise, we also had other concepts for the fossilization of the sound, and that was applied to rotorcraft in particular for search and rescue operations where the possibility of providing to the pilot and to the person that was accompanying the pilot, specialized sound could help search and rescue operations without compromising or limiting the compromission of the safety of the entire operation. I mean, yeah, these are ideas but maybe we can go with the first one. That was to share a bit what I remembered from this project. It has been a while.

Interviewer: Yeah I understand. And you said specialized sounds or spatial sounds?

Interviewee: Uhm distributed in the space. So for example if there is a mountain approaching... it's like in the car basically when you have to park...

Interviewer: It makes a sound when you come closer.

Interviewee: Yeah exactly.

Interviewer: Yeah, okay. Yeah maybe let's go with the first one because there is a lot of interaction with the user. So... Yeah, actually also with the other one, but...

Interviewee: Yeah, okay. So you said I should start from the lower...

Interviewer: Yeah, the exit criteria of one and two.

Interviewee: Okay, let me see again, criteria one and two. I mean, what are criteria one and two? Step one and two? *exit criteria*

Interviewer: If it's still here, you have to tell me because then I can explain it to you.

Interviewee: Yeah, exactly. Okay, let me go again *determine whether the exit criteria associated are fully fulfilled* Where do I see the exit criteria? These are the exit criteria?

Interviewer: Yes exactly.

Interviewee: All the criteria that are here? So human factor categories are actually the criteria that I have to check, if they are fulfilled okay. So if I start at TRL one, so *basic principle human factors research documented and conducted* uhm *functional structure of the work domain identified* so this is part of the human in the system. I mean, yeah, for sure this is done because this is basically the definition of the concept. So it's probably, yeah... in this stage [HRL 1], I would say that basic human factor research is actually the definition of the need from the human factor perspective. So the need of the pilots and definition of the needs of the different actors in terms of users.

Interviewer: Okay.

Interviewee 14:16
So really the pilots and the people that will be using the technology. *functional structure of the work domain identified*... Yeah, I mean, this goes with the concept probably, yes. So yes. *input output and constraints of control tasks identified and process of how tasks can be done or described* And this is procedures, roles and responsibilities. Yeah, I mean, at a high level, I would say yes, I mean, for me, this goes all in the let's say, concept definition. Maybe what I would... *environmental constraints identified* Uhm... Yeah, not sure if this tells exactly what is here to me.

Interviewer: Okay.

Interviewee: I will give you any type of feedback, okay?

Interviewer: Yeah yes. I'm actually curious, what do you see... Like, what is environmental? Here, what does it mean? According to you?

Interviewee: Yeah, exactly. I mean, for me, what it means is... let's say I can tell it in my words, that is what are the boundaries of my design space, basically, which includes physical space of work, but it includes also activities of people. So it's really... I mean, for me environmental, I don't know if it's... when I think to environmental here at this high, at this level number one, I think, really to the context and the work of people within the context, which is basically what I tried to say, at the first level, I imagine that you need really to make an appraisal of the needs that are there. And also say, okay, if needs are emerging in this type of situation. So definition of the situation, environmental its situation, so not just what is out, let's say, in the environment, but also what is in the persons and in the work that people have to do.

Interviewer: Yeah. So in that sense does the title cover the exit criteria on the right, or do you think another word...

Interviewee: It is probably possibly a bit misleading. For me. Consider also the...

Interviewer: Yeah, it's good that you say so. To give a bit of background, it actually emerged from the cognitive system engineering philosophy, that you always have to start with, like you say, the situational context, because it determines to a great extent what the human in the system does. So...

Interviewee: But it includes also people needs or not?

Interviewer: Yeah, in cognitive system engineering that comes afterwards. So you first look at the ecological or environmental constraints, and then you look at cognitive constraints. That is the idea there.

Interviewee: And so when you say basic human factors research conducted, can you make an example of what a basic human factor research at this level would be?

Interviewer: Yeah, that's a good question. Because actually, this is not my framework. I tweaked it a but it's was developed by a number of American scientists and I honestly I don't know for sure what they mean here. It's quite vague. So it's good to hear how you interpret it, because I think everyone has a different idea of this and it should become more clearly stated, and yeah, so I also don't know what exactly they mean here.

Interviewee: Okay, because I really took it as needs of people at this stage because I usually, I mean, it's a bit of also the way we usually start no? You start from the needs. So that's how I interpret it. And so that is why the word the environmental here, it sounds for me a bit limiting.

Interviewer: Yeah, okay.

Interviewee: That was the reasoning.

Interviewer: Yeah. Okay. Yeah and it's also not sure basic humans research conducted... So is it more like desk research? Or are you going to measure workload in a laboratory? It's not completely clear right?

Interviewee: No, that's why for me it was really need. So it can be also like, I speak to five pilots, two questions, three questions: What do you do when you are in a bad weather conditions? What would you need to help you? And tell me a story of near-miss or you know, a situation where you felt uncomfortable. So I would really ask five pilots these three questions and then derive, define a sort of list of basic needs that comes out of this type of documented research. No further than that for me at this stage.

Interviewer: Yeah, okay. Good to have your interpretation of that. Yeah. Okay. So...

Interviewee: Sorry, just a question, what is WDA here?

Interviewer: That is the work domain analysis. So, yeah, the one that they use in cognitive system engineering, and then the control task analysis and the... Yeah. So yeah, that's where also the environmental word comes from.

Interviewee: Yeah, okay. Okay. Yes. So this [HRL 1] is covered. Yes. For the for the head up display. Then TRL/HRL 2... *design principles established... human system performance design principle researched... human system performance on compatible systems... metrics and benchmark for successful...* Yeah. *...relevant human capability and limitation or characteristics identified... usage scenarios and task description developed...* Yeah. *organizational operators, maintainers...* I don't understand this.

Interviewer: Okay.

Interviewee: The teams and communication yeah.

Interviewer: Yeah. So basically, that's kind of saying what is the outcome of social organization and cooperation analysis. So you have defined the tasks, but who is going to execute the tasks and who will need to communicate or cooperate with each other? And also what kind of hierarchy, for example, exists in the organization, in the team?

Interviewee: Okay. Yeah, yes. Yeah, I think... Yeah, for sure we covered also, I mean, the things were... also this level [HRL 2] was covered with the heads up display. Uhm Yeah.

Interviewer: Okay.

Interviewee: *Design Principles established.* For me, this makes sense.

Interviewer: Okay.

Interviewee: Because it's really like, okay, first of all, I tell you my need and then I tell you where I'm moving within that need, which is stating principles of design, the human in the system then considering the probabilities and limitations, usage scenarios. I mean, probably usage scenario, staff description and also... at this stage [HRL 3], I don't know, maybe it's... uhm... I was thinking that at one point I would probably, early in the design process, I will probably think also to let's say bad cases. So scenarios that I don't want, that could happen or scenarios that I don't want to happen. In order to understand what could go wrong basically. Especially if I am introducing a new system like it was for the head up display case, so something that pilots never used before uhm... Yeah.

Interviewer: Okay. So I can already spoiler a bit. At HRL three, it says potential human system performance issues and risks identified. Is that what you mean?

Interviewee: Okay, then it is here, yeah. Yeah, exactly.

Interviewer: And would you say it's better to have that already in HRL two?

Interviewee: No, no, it's fine here.

Interviewer: Okay.

Interviewee: This is design principles so it makes sense that it is here.

Interviewer: And you also mentioned usage scenarios and task description... You think it's too early? or what was your thought? Because you were...

Interviewee: No no no, I think it's well placed. At this stage [HRL 2], I want to somehow start thinking, maybe generic usage scenario, maybe it's specific to that tool. So I mean, the head up display is a tool that is provided, I mean, imagine that it is provided within a operational environment like the flight deck. But of course, it's used, like the use of any other tools within the flight deck, is a within an ecosystem of other tools. So at the stage number two, I would probably just imagine usage scenarios for that specific tool. So I'm in the, in the bad weather, and I have to land. Whereas moving into system requirements, I will probably slightly enlarge the scope of my usage scenarios and place them in a in a larger, let's say, scenario in which I see also what happens before and what happens after.

Interviewer: Okay.

Interviewee: It's, I mean, maybe it's a generic suggestion, but just to... because I'm thinking that when you introduce one innovation, doesn't matter if it is a technology, a new procedure, I mean, whatever, but okay, you first of all challenge the execution of the specific tasks you are trying to support with the new tool and the new procedure. But at one point you have, not too late, you need to think to what happens before and what happens later. And if those things change, because of the new thing that you have introduced.

Interviewer: I think it's a really good point. Is there an example from a project, for example Alicia, where this changed a lot your perspective, because you looked at what happened before or after? Can you maybe remember?

Interviewee: No, it's really, I mean, maybe it comes from this theory for me. I can't make an example maybe but it's really that when you change the activity by introducing a new procedure, a new tool, I mean, the activities change, and it changes in the in complex systems, it also impacts the rest of the things that you are doing. Also I mean, this applies also in terms of... let me think... maybe [colleague] recalls it better... but I think the example could be the introduction of another tool that was maybe a stick with haptic features that would shake in specific conditions. Still in Alicia. And I think when you have to place those new tools within a flight deck that is a limited environment with a number of other tools, you really need to think through what happens before and what happens after because, for example, if you need to physically use one tool but in the same time you are supposed to do another thing where you are, for example, back turn the beat or you are operating the head up panel or whatever, you need to consider this because you might generate physical ergonomics issues, that then of course, may cause some safety problems. But maybe, I mean, if you want a bit more theoretical, then you can ask Stefano if he recalls examples for that. Because for sure he has them. Both from Alicia because he was in Alicia as well and he really did that human factor job on the field. So maybe he can provide you with an example.

Interviewer: I'll ask him yeah. I think it's really true that when you look at a complex system, it has some kind of temporal dimension, so you can't just look at two seconds and ignore all the context. So...

Interviewee: Yeah, exactly. That's the perfect synthesis of what I wanted to tell you.

Interviewer: That's good.

Interviewee: Yeah, but maybe if it is a way of slightly adapting this criterion, so the usage scenarios, I mean, if you deem then useful to slightly, let's say, include this aspect, maybe here [HRL 2], it's a good place, because it's when you... I mean, here, you're still thinking about the concept, the principles, the needs, how people can be, I mean, people, role and so on. Usage scenarios. An then here, you have to, let's say, take higher system, let's say, look, where it makes sense that you consider the task within a larger scenario, let's say.

Interviewer: Yeah.

Interviewee: Then we have potential human system... And I think, I mean, for the head up display, we considered it. So I mean, for this aspect, let's say we move to level three for sure. Then we see *potential human system performance issues and risks, identified* Yeah. *requirements to support human system...* And yeah, and it's very much connected to these actually, as you said. *preliminary design features to accommodate human capability, limitations and needs investigated* Yes. Yes. Uhm Yes. Yes. Yeah, this was done. Perfect. *manpower and personnel analysis conducted* uhm this is organization and staffing... *manpower and personnel analysis* I mean, at which level here?

Interviewer: At the level system requirements (HRL 3), so, whatever that may be.

Interviewee: Oh, okay. So it's basically which competencies and skills I need to be able to operate my concept.

Interviewer: You are the expert. So I listen to your interpretation, because to be honest, this is how they put it in the original framework and again, it's not clear what is a manpower and personnel analysis. Like does it mean how many people do we need or, it's again very open for interpretation...

Interviewee: Yeah yeah, uhm *strategies to accommodate manpower and personnel recommended...* I mean, this is the organization and staffing. So if it is organization and staffing, it's really related to... yeah it's probably also something like... I mean if you have something that cannot be used for too much time, maybe it's here that you have to consider no? So for example, if I have a screen that if you use it for more than two hours, you will lose your eyes, then, of course, I need to consider it. And this is an aspect that needs to be considered at the level of the proof of concept. No, so it's probably, yeah... makes sense and yeah, and definitely, I mean, that was not the case, though, for example... uhm... Yeah, but training is the next. Yeah.

Interviewer: So it's not completely clear what is the distinction between the manpower and personnel versus training? Or?

Interviewee: Uhm no, I mean, I would probably... *analysis conducted...* No, okay. It's, I mean, it's kind of clear the difference, though, I don't know if at this level [HRL 3], I can conduct a training analysis. Because I'm still. Probably I'm still... Maybe I can... Yeah, maybe what I can do, but then I don't know if it is here, or here, it's probably more here. What I can do is I can tell you the type of expertise I expect the personnel to cover in order to use my system.

Interviewer: Yep. Okay.

Interviewee: So what I can say, again, in terms of training analysis..., yeah, maybe training analysis conducted makes sense, actually. Because I can tell you yeah I need... I mean, for having this system working, I need this type of expertise. No no it makes sense. And actually, yeah, this...

Interviewer: But it could be more specified maybe? Instead of saying training analysis, it could say...

Interviewee: Definition of the skills and competencies needed for uhm yeah... Yes.

Interviewer: Okay.

Interviewee: Yeah, because I'm now thinking, for example, about the remote tower no? If you think to the remote tower, I mean, at one point, probably while they were designing the concept of the remote tower and then assessing potential human performance issues and risks, they said, or you know *requirements to support human system performance. Which are the metrics for a human system performance. * So at this stage [HRL 3], they probably would know the type of the profile of the user, how the user I mean, yeah, the profile, the skills, the technical skills and competencies profile of the user. So maybe I would slightly, I mean not, exactly, not maybe talk about uhm... But that's the category *training and development* So maybe here, when you say, I'm sorry I'm jumping, but manpower and personnel analysis is probably, maybe I would refer more to let's say the profile of the personnel that is expected to operate this type of systems. Then here, for example, if it is something that requires more than one person, then it's probably organization and staffing. So here, you will probably be saying, okay, to operate these systems that is done in this way that has these constraints, and so on and that has these potential human performance issues, then you need to consider the allocation of, I don't know, three operators or three pilots because they have to distribute, let's say, tasks, among them in this way, and then they have to do this type of shift during the working day, or during the working week. I mean, this is not the case of the head up display, of course, but I think I mean, it makes a lot of sense that when you say which are the issues and risks or for the human system performance then it is also when you will tell the type of allocation of responsibilities to different, possibly different users or one or more than one user, and you also state what are the level of skills and competencies that you expect them to have. So yeah...

Interviewer: Yeah. Okay, good. Yeah.

Interviewee: So to let's say, to tell you, maybe, yeah, maybe here, it's, for me, it's more like uhm... I mean, between the two, it's more like a profile that you expect to be using the system or to be operating the system. Typical profile of person. Maybe you can have, for example, more than one profile. Imagine that you have two different profiles that need to work together to do the same thing, to think about the air traffic controllers. When you have the executive and the planner air traffic controller. These are two different profiles, then they can be covered by the same person, but when they are seating, one as an executive and one as a planner, they do different things and they need to... they sort of have different masks in a way.

Interviewer: Yeah. And probably both masks require different skills...

Interviewee: Exactly. So maybe you can, maybe what I would suggest, to go a bit into the human factor pie to see the specific definitions for those two categories: organization and staffing and training and development. And see where these think of the profiles matches the best if you are convinced that it is useful. Then usage scenarios. Okay. It's what we said before, maybe here I would open it a bit to what happens before and after. And then *teams and communication... interdependence relationships that could be supported in the implemented system... directability* what is a directability requirement?

Interviewer: This actually comes from more human-automation teaming or human-machine teaming. So observe ability, whether you can observe what the automation is doing direct stability, whether you can change the commands of the automation, where you don't agree with it, predictability, whether you can predict what the system is going to do, in that sense.

Interviewee: In that sense, it is super interesting.

Interviewer: I will send the paper, it's a really interesting paper. And I think it actually applies between humans and machine, but also between team members that are both human because you want to see from your team member, what he or she is doing, you want the person to be predictable and you want to be able that if this person does something that you think is wrong, that you can change that. So...

Interviewee: Yes and it's, I mean, it's super interesting for two things, for the explainability of the artificial intelligence, which is a feature of an artificial intelligence systems for being explainable, it has to be understood by the person operating it, or I mean, cooperating with an artificial intelligence. And then it's also interesting for the concept of partnership. Partnership with automation. Where, yeah... so please, yeah, send me the paper.

Interviewer: Yeah I will send it. They call it interdependencies. Yeah. But is it clear what is... because it's now a very long statement...?

Interviewee: Yeah, I mean, it's clear. But I mean, this is a summary. You probably have a document that explains and makes examples in detail, I guess no? Because the table summarizes things and if you don't have somebody like you now helping me going through it, I might interpret it. So, I really... but I mean as it is, here *set of interdependence relationships should be supported in the implemented systems* Yeah, for me makes sense. Yeah.

Interviewer: Okay. Now that I gave you the background information it makes sense?

Interviewee: Yes, exactly. But you know, this level [HRL 3], I mean, I don't know I think what I remark between this and that for example, is that here you tell me three keywords like observability predictability and directability that I mean, if I knew that, I would exactly know what you expect as an output here, which is good. Whereas here, when I say analysis conducted, I mean, what does it mean? What shall I look for in this analysis? Here, you told me what shall I look into, at this level. So you tell me at level three you need, from teams and communication, you need to look at how relationships are observable and interdependencies are observable, predictable and directable. Or, I mean, you have to be able to define the requirements for that. Whereas here, for example, you don't tell me what analysis means. At which level? So probably what I see is a difference in the... I mean, both can be clear but here, I don't have too much to imagine what you expect to me, here I need you to explain me what you expect me to do in terms of analysis. Because here you tell me the criteria that I have to look at, the real criteria actually, if we talk about criteria. Here, you don't tell me the criteria. You tell me do the analysis. Okay. I mean, the analysis, we see it also in [other project] analysis can be also I ask you your name and how much time you're spending in the factory, but this... so it would help me having something like similar to these than something like you know generic words like analysis. Again, here, you tell me what I have to do, I have to check issues and risks. I mean, I have to identify issues and risks in human system performance. It makes sense a lot. Even if it's few few few words, but it makes a lot of sense and I know how to do it. Here I don't know how to do it. So I don't know how to answer.

Interviewer: Good point. Yeah. Yeah, definitely.

Interviewee: Maybe it applies also to others and then you will look at it if you... okay. Okay, then we'll go to four... But yeah, in the case of the head up display, this [HRL 3] was done I would say. We were probably at level four...

Interviewer: Completing the design concepts?

Interviewee: Yes, now I'm...

Interviewer: Because it also contained a simulation...

Interviewee: Yeah, exactly. So maybe we were also... relevant environment means that it's not a prototype but it's... a simulator is a relevant environment?

Interviewer: Simulated is relevant yeah.

Interviewee: Okay. Yeah. So we... and when it says demonstrated and validated what is the difference?

Interviewer: I just copied the TRL. If I would have to give my own interpretation is that demonstrated is a bit more advanced. So your concept is mature and now you have to prove that it works the way you said it's going to work and maybe validate is still a bit more like okay, we made a concept now we have to check whether... Yeah, I don't know. I think it's very close to each other. Also because the whole phase is called technology demonstration. So...

Interviewee: Yeah, sure. Okay, so TRL four, for sure we were here *design concepts completed... potential human system performance issues identified* this is the same as in HRL three?

Interviewer: Yeah. Yeah.

Interviewee: But it's the same output or it's a further refinement of what you do in three?

Interviewer: Yeah, the criteria repeat themselves as you concluded. And the only difference is that, like you say, it becomes more refined, more detailed, your concept is more advanced, so you can probably get more feedback. And you can also see on the right that the methods change. So maybe you can start NASA-TLX etc. to get a bit more specific feedback.

Interviewee: *Conformance of designs to human systems performance requirements and principles verified* okay. *support human use* what is that? Alternative strategies to support human use recommended?

Interviewer: Yeah, tell me your interpretation.

Interviewee: No, yeah, I tell you mine, while we are doing the validation in the lab, we probably have possibility of seeing the users using our system. And so we can also imagine, I mean, slight changes to the concept that we have devised. But I don't know.

Interviewer: And do you think that human use adds something here? Because it's all about human use, right?

Interviewee: Yes.

Interviewer: So what do you think this comment is adding, compared to the others that are already there?

Interviewee: I mean, for me, this is... here, it's the possibility of seeing how people use it in a lab, okay, but I see it... because I see people doing things, even if I'm doing the experiments in the lab. So I can actually see whether all the things that I had somehow envisaged before are actually met by the real user. If not, or also alternative things can emerge, while I'm observing people doing things. But I don't know if it makes sense.

Interviewer: Yeah. I mean, this is also one of the criteria that keeps repeating so the next level, you see strategies to support human use recommended. And it continues throughout the HRL.

Interviewee: *Strategies to support human use.* But yeah, yeah, so it's, yeah, it probably makes sense. I don't know if that was your interpretation as well. Or...

Interviewer: I myself was a bit confused because I thought that the other exit criteria, they're also aimed at supporting human use. So for me, it was not clear what was necessarily the added value of this comment, I feel it's more a summary statement of... but...that was my...

Interviewee: I mean, the fact that it repeats here, so alternative and then strategies makes me think that when I... okay, I divide, I think of a concept, I work on the definition of the principles then I somehow generate my prototype and then I take my prototype and I go into a lab, okay. When I go into the lab, I will actually do my first validation activities at a low TRL. But I will be able to see whether there are alternative ways of using the system. I mean, probably not alternative ways of using the system. But maybe I can think of an example for the head up display.

Interviewer: I know, for example, in Safeland, that they considered alternative... for example, the alternative is to have the ground pilot being based in the airline Control Center, or on the airport, or in a specific ground pilot center. So maybe those are alternative strategies to support...

Interviewee: Yes, yes, I think. Yeah. And probably you need to think about that when you are doing the validation in the lab, because you probably, I mean, it's a situation where you're simulating the tasks, so it's easier to consider alternative strategies to support people. Yeah.

Interviewer: But was it done in Alicia? Something similar to this?

Interviewee: No, yeah, exactly. Now I can't remember... Yeah, I mean, for sure it was part of the reasoning on some technologies, now I'm especially thinking to your example, maybe in Across when we were considering the case of the total incapacitation of the crew then... okay, we had to think about, I mean they had to think about, different solutions for the ground pilot to intervene. But that was probably already part of the concept, maybe... I mean, the fact... Okay, now maybe I'm understanding a bit. What makes me, what makes my life hard in a way is that what you said, as an example, where do I place the pilot, is he going to be in the control center, or in the examples that you did. But this is something that you think, along with the concept definition, no? Probably you don't have to go into the validation lab to think about the support of a ground pilot and where he or she should be placed. So here, for me, it's something that is somehow generated by the fact of doing the validation in a lab. So something that I'm not searching for. But it's something that pops up in a way. Because I'm putting the technologies, I'm somehow not stressing but I'm validating the proof of concept of my technology, the prototype, let's say of my technology. So it's really something that, again, if I managed to explain you the idea, then maybe again, ask [other colleague], because he might have examples, real examples. But I'm really thinking to something that you didn't consider before, not because you were not requested to consider it, because the consideration of these aspects, it's enabled by the fact of taking the technologists to the validation in the lab.

Interviewer: Okay, and does it make sense to you that there's an exit criteria that you have to consider what you didn't consider?

Interviewee: Yeah for me, yes. Yes, because for example it can... Yeah, if I am not taking a completely wrong direction, but I mean, you could also at this stage, maybe you can also discover that that specific solution that you had imagined for managing bad weather is actually super useful also when you have a wonderful sound set in front of you, and that shines into your face and you cannot see I don't know, I mean stupid examples, but so it makes sense that you also challenge a bit your technology to see whether you can have alternative strategies. But again, I mean, this is probably super personal interpretation of what they meant here by alternative strategies to support human use recommend.

Interviewer: I will ask other people as well and see if they say the same as you. Yeah. At least you...

Interviewee: Probably not. Because I'm probably taking the wrong direction.

Interviewer: No I think... Yeah, I wouldn't know how to alternatively interpret it so... I think it's very... I think one important thing is to distinguish that maybe it's not a strategy by the operator, it's really a strategy by the system developer, to support the operator. So I think that's...

Interviewee: Okay. Yeah.

Interviewer: And yeah, so yeah. It could be like you say, it's something that pops up during validation that you didn't anticipate, you didn't expect and then you have to... yeah.

Interviewee: Yes, yes. Because otherwise, I mean, it also makes sense because otherwise you uhm I mean, at which point you somehow see if there are alternatives, or if you can do things slightly different or if you can consider other aspects of what you had designed so it make sense. But yeah, please you talk to also other people. *human system performance metrics are successfully met* Yes. *strategies to address environmental constraints recommended* uhm Yeah. Yeah, I'm tempted of anticipating these aspects before. Because I think here they talk about environment, but just as the let's say, working environment, no? Physical environment, physical working environment, work domain constraints, integrated in the conceptual interface design... Yes, but yeah, maybe I'm tempted off anticipating it, because I don't think it's something, it's probably something that you, I mean, why it relates to human readiness level?

Interviewer: Because that influences the human.

Interviewee: Ah, okay. So it's environmental. I mean, not in terms of environmental sustainability, but it's again, working environment, the working space, the physical working environment, okay.

Interviewer: Yeah.

Interviewee: Yes, this is my bias. So with the Green Deal, and working on the Green Deal projects

Interviewer: It's good I think uhm a lot of people will... like you say it's misleading this term.

Interviewee: Okay, then yes. *strategies to address environmental constraints recommended* Yes, perfect exit criteria, you do the yeah, that makes sense. Whatever environmental means... *strategies accommodate...* Yeah, exactly. Here. It's like the suggestion before maybe having it a bit more detailed in terms of what's the level of the strategies that you expect at this stage.

Interviewer: Yeah.

Interviewee: *Strategies to accommodate* Yeah, okay. And it makes sense, it's probably... Yeah. *Procedures for operator and maintainer roles designed* Yes *task flow and sequencing optimized* Yes. Ah okay, *mechanism* so now you want to know the mechanisms to support observability. Okay. That's five. I mean, tell me if we need to...

Interviewer: No it's good. Yeah, you can just continue.

Interviewee: Okay. Because we have this one [Impact Wheel] as well. Do we have to look at it as well?

Interviewer: So human in the system is defined a bit more in detail here. So all the things that you could think of when you consider human in the system like workload, vigilance, acceptance, attention...

Interviewee: Okay, okay.

Interviewer: Yeah, I think for now, it's not interesting to go into that in detail. It's interesting to know whether you think that that is a good thing to have there, or whether you think that is unnecessary?

Interviewee: Um, I mean, for me, it's, at this level, so at the human in the system level, yeah, for me, it's super interesting. I will do a similar thing for the others. Okay, because it helps me a lot in then, and you probably, it helps you a lot too I think, in further defining the exit criteria of these items that are a bit unclear, still unclear because it's generic no? But if you exploit them in a way. I mean, if you say basic human factor research documented and conducted, of course, at this level, we are not talking of, I don't know, stress, or trust or workload or attention. What we can do here in one, it's stage one is that we say, human factor research is probably at the level of the need of what people need to do. So here the human in the system, it's really what are the needs of the of the operators that I'm trying to address with the new solution that I'm offering. But yeah, this is useful. It's useful, although I think it's super useful at higher level, at the higher stages.

Interviewer: So maybe at the beginning, you can only say which ones you think are going to be important later on or?

Interviewee: Yeah. So here no? When you say, when you start thinking about metrics for successful human performance, you would probably... *metrics and benchmark for successful* Yeah maybe it's there... Two and three is when you start going into these items and you start selecting them or rating them in terms of metrics that you expect, expectations that you expect to meet with your solution.

Interviewer: Okay.

Interviewee: Okay. So five *potential human system performance issues identified* Okay, that's like, okay, we are in technology validation... Okay, we are in the relevant environment. We are in the simulator. *Conformance of prototypes to human system performance requirements and principles verified* Yes. *Strategies to support human use* okay this is what we saw before. Yeah. Yeah, this is, I mean, again, how this is different from that? It's probably different metrics or different human factors. Let's say, performance items that I look at no?

Interviewer: Yeah, the environment is different.

Interviewee: Yeah, exactly. Because the environment is different. So I don't know if this is... Again that's why I think this has to be somehow... You need another document going into the detail for that. But yeah, this is what happens when you do a validation so it makes sense. Okay. Okay but this is very similar to the previous one. Okay. Okay. Okay... *practice...* Can I go to six?

Interviewer: If you feel that five was also completed?

Interviewee: No I mean, yeah, I think we were here with the...

Interviewer: You were on your way to complete HRL 5?

Interviewee: Yes. I think HRL 5 was completed actually with the head up display. But we were probably not here yet.

Interviewer: Okay.

Interviewee: *...demonstrated ... predict* Let me read it, the exit criteria *practice to track and resolve human system performance issues evaluated... conformance human system ...verified... strategies...* Okay, here... *operator and maintainer...* We didn't do all of them in..., for the head up display, this part, let's say the organization and staffing and training was not that much considered. Same with the teams and communication but for the others so the human in the system, the working environment and the procedures, roles and responsibilities, yeah, I mean, reading HRL six, we were probably towards six, let's say.

Interviewer: Okay.

Interviewee: Because we had also a phase of, let's say uhm solution of issues and risks identified. So we were a bit like in the iteration, we were probably more towards six than five actually.

Interviewer: Okay. So what would be the next activity you would plan?

Interviewee: Uhm... I mean for sure, I would work on the organization, training and staffing because that was left a bit behind. So I will try to, let's say, raise the levels of those three items. So basically on organization and staffing, training and development and teams and communication up to level six, because I think the work was very much focused on the technology and on the human in the system. But it was super focused on the technology of course so the human the system was a way to demonstrate that the technology was useful and was well, let's say, designed. But of course, when you, I mean doing it in a project, we had the possibility of somehow influencing the full technology-driven process by actually working a lot on the human in the system. So this was quite developed. We had I mean... so yeah, the next steps for that would be probably a refinement of the human in the system. But the organization, training and teams and communication were probably the areas that were not looked at all. So yeah which makes me think to one thing that is; are those exit criteria, I mean, do I have to cover all of them to move to the next step? Because this is interesting, actually. Though, I don't know how much it is feasible when we...

Interviewer: In the whole program structure?

Interviewee: Yeah.

Interviewer: Yeah, so officially, yes. The answers is yes. So that would mean that if you focus on the human in the system, and you don't consider organization and training for as far as it's relevant, of course, that you will always stick with, let's say, HRL one, even though maybe you have done simulate simulation exercises. Yeah. So that would be the official answer. But do you think that makes it a bit unfeasible to use it or?

Interviewee: Uhm no, but you know, to, I'm fully convinced that this is the way to do this thing. But for me, what will be really useful is to be able to demonstrate actually, that having this combined approach, this integrated approach that makes all the categories advance together, is actually a good way of doing thing, because it's helpful to the technology development. So it's not just because we ask for doing it, but because it brings an added value, benefits in a way the technology development.

Interviewer:Okay, so...

Interviewee: Yeah, please...

Interviewer: No, go ahead.

Interviewee: That it's not just, you know, a box to tick, but it's something that's actually gives me... I mean, the strategies to accommodate training are updated as an exit criterion here, it's super crucial for me to go to the next level, because it produced me benefits in the whole system design. But I know this is actually quite dreamy.

Interviewer: Yeah the thing I can think of right now is that, for example, I think it was [interviewee #1] that said it, that's there's always this triangle of safety, costs and training. So you can, when there's a safety problem, you can solve it by putting more money and making, like, I don't know, giving everyone in the plane a parachute in case the plane crashes. So it is going to cost a lot of money. Or you can train people, of course that also costs a lot of money. So there's like this trade off. So I think training is very important to consider early on, because maybe you can fix an issue not by changing the system, but by changing the training, for example, or the manpower.

Interviewee: Absolutely.

Interviewer: But that's the only thing I can think of. And maybe it also has to... because if you have cases in which it was done this way, you can also start to prove that it is a benefits. So I think it also emerges from applications. Because I think...

Interviewee: You know what?

Interviewer: Yeah, tell me.

Interviewee 1:18:48
I don't know if it can be useful, but if you try to have a look at a couple of those videos, that tells you about disasters, like the aircraft in the East Midlands crashing like few meters before the landing, the runway, because they switched off the wrong engine. I don't know if you're familiar to that. But we have a number of videos that tell you, I mean uhm...

Interviewer: Like air crash investigation?

Interviewee: Yeah, sort of air crashes investigations. And it's uhm yeah, I think I mean, maybe you can try just with one. This one of the wrong engine, basically they had a bumping of the aircraft at one point and they analyzed the situation. And based on the checklist, analysis and so on, they came to the conclusion that they had a problem to one the two engines, and that was the right hand engine, so they switched it off. And the problem appeared to be solved because the aircraft stabilized. And it could continue, although they had, of course planned for an emergency landing, but in a actual airport. By one point, while they were very, very close to the airport, they had this bad thing again. And actually, they realized that they were having problems with the second engine. And of course, an aircraft can fly with one engine broken, but it cannot fly with two engines broken. But actually, I mean, of course, the aircraft crashed, the two pilots survived together with a number of passengers. But then during the investigation, they saw that actually, they had switched off the wrong engine. And the reason for doing that was a number of I mean, of course, it's a chain of events. But one of the few things that I remember was a bad interface that were recently changed. So from the interface, they had this confusing information that didn't help them in, let's say, identifying one number and one information and putting it in the right place of the checklist. Plus training, they had a super stupid training. So I don't know if, when you go, you analyze using this HRL system, if you take an example of a disaster, and you try to place the things that went wrong. Just to see whether actually having... just to demonstrate that there is actually a benefit in approaching the thing in an integrated way.

Interviewer: Yeah, you also made me think of the Boeing 737 MAX. Yeah, they will never pass officially HRL nine, because they didn't change the training of the pilots.

Interviewee: Yeah exactly. And they are always left aside, especially training, teams and communications, unless it's a specific objective of the project, and also the staffing and organizations. So yeah everything is super focused on the human in the system, which in most of the cases is reduced to the HMI. But this is not the whole story. I mean, the whole story is much wider, and it's much bigger.

Interviewer: Yeah. Yeah, good points. Okay. Well, I got some good feedbacks on the framework. Is there anything you have to say about it? That is not related to anything we talked about so far?

Interviewee: No, I just wanted to ask you: assumptions and affordances. Can you explain these two things?

Interviewer: Yeah, definitely. That would also need some explanation in an additional document, I guess. What is meant here is that, during the process, I saw a lot in the project, Alicia and Across, a lot of assumptions are being made. For example, in Safeland, we also make assumptions about the data link connection. We assume that we have the perfect Data Link connection that has no delays, stuff like this. So these assumptions need to be feed back into the technology development, because if we make these assumptions and technology department doesn't get to this perfect Data Link connection, then we can assume that this is the case, but it's not. So it's very important that these assumptions that we make on the human factor side that they also come true on the technology side, so to say, and then the affordances is meant, like the technical constraints. Affordances is the positive version, the opposite, of constraints, so it's the technical technological constraints that are put on the system and that we have to take into account as human factor specialists. So if the data link is what it is, and it can't improve, this is how it works, then that is a constraint that we have to work with. So in that sense, it's kind of a relation between the technology and the human, or the Human Factors part of the system. Does that make sense?

Interviewee: Yes, yes. Makes sense. Absolutely.

Interviewer: Okay. Anything else?

Interviewee: No, not really.

Interviewer: Would you use it? Do you think it could be useful in projects?

Interviewee: I would use it, if it is more detailed, a bit like, think about the framework we are trying to use in [other project] no? And the framework actually selected some of those categories and specified the items that we have to look for, than if I could have something like a matrix saying that those items are essential for, let's say, validate the human readiness, that specific human readiness level. Yeah, that will be super useful. With a very practical support, but I'm really talking about practical use. I think the usefulness of these type of things is also what we were saying before, in terms of like, suggesting a different way of designing systems. So for that, I don't think you need to go into the super details of elements or the details of the framework that we have in [project] about, but it's really like showing that there is a parallel path, that the two things are in connection. And it's I mean, it's a sort of like crossing thing where you have, when you go from seven to eight it's actually because this feeds that, and this feeds that and they cross-feed each other. And this is for me... So the two level, I mean, this, I don't know which of the two levels is more important for us. In practical terms, that's a bit too high level for me to use it. But in conceptually and methodologically it's super relevant. That's why I was telling you, I mean, that would be super good for me to stress on how those categories are crucial. And they are important to be considered, even at lower levels of maturity, because they actually give a benefit to the next level of technology readiness.

Interviewer: Yeah, okay. It's good, it's very useful. Thank you very much.

Interviewee: Thank you. That's interesting. Also from time to time to be brought into the, you know, this type of reasoning. So I really liked it.

Interviewer: That is good. Yeah, for me it was useful, because I read about the human factor pie and all these things that we are now also using another project. So that was a good background information for me.

Interviewee: Nice. Okay.

Interviewer: Nice. All right. Thank you and happy Easter, enjoy the weekend.

Interview with maritime expert from TNO (Dutch)

Interviewer: Oké. We kunnen nu de HRL openen en ik denk dat het het handigst is al je hem zelf opent en het scherm deelt met mij dan kan je makkelijk navigeren.

Interviewee: Ja dat is goed. Want ik heb hem inderdaad open staan...

Interviewer: Yes. Oké.

Interviewee: Kan jij het ook goed zien?

Interviewer: Ja, ik ken hem onderhand al een beetje uit mijn hoofd.

Interviewee: Ik heb jou camerabeeld er naast staan dus ik kan niet het hele scherm gebruiken maar ik kan het goed lezen.

Interviewer: Oké, laten we beginnen bij het begin, pagina 1. Ik weet niet of je al tijd had om het een beetje door te nemen?

Interviewee: Ik had hem iets meer door willen nemen. Ik heb het niet helemaal goed kunnen lezen. Ik heb alleen even snel doorheen gescand.

Interviewer: Ja prima, dat is voldoende hoor.

Interviewee: Wat wil je precies vandaag van mij allemaal weten? Dat vroeg ik mij nog wel af.

Interviewer: Ja, oké oké. Ja is goed. Wat ik het eigenlijk liefst wil doen, eigenlijk een project nemen waar je in betrokken bent of betrokken bij bent geweest als voorbeeld. En dan daar de HRL op toepassen en dan eigenlijk een soort van think-aloud doen waarbij je hardop nadenkt van oké ik kijk nu naar HRL 1 en deze activiteit hebben we gedaan dus ik denk dat HRL 1 compleet is. Een beetje op die manier. Zodat ik kan zien hoe het is om dit te gebruiken voor iemand die tot die target group behoord.

Interviewee: Ja oké. En heb je dit al vaker met iemand doorlopen? En ga je dit met meerdere mensen doen?

Interviewer: Ja met wat collega's hier heb ik het gedaan en ik heb wat mensen gecontracteerd, mensen bij Philips... want het is voor mij ook wel interessant om meer uhh ja het bedrijf waar ik werk zit heel erg aan de onderzoeks-kant, doen heel veel Europese onderzoeksprojecten waar je eigenlijk nog heel ver af staat van een eindproduct, dus ik was ook wel geïnteresserd in industriele organisatie dus vandaar dat we ok wel bedrijven zoals Philips en Pro-Rail hebben benaderd.

Interviewee: Ja, leuk!

Interviewer: Ja en dan hoop ik dat het ook nuttig is voor degenen die ik spreek. Ze zijn nu bezig met het publiceren van de HRL en ook het opstellen van een standaard, dus dit zou standard practice moeten worden. Dus het is wel belangrijk dat het een beetje aansluit bij wat experts op dit moment doen.

Interviewee: Ja. Ja.

Interviewer: Dus ja, laten we er gewoon doorheen gaan, laten we het een keer toepassen. En als er dan nog tijd over is, heb ik nog een aantal specifieke usability vragen.

Interviewee: Ja.

Interviewer: Dus ja, heel basic, het is een schaal waarmee je de volwassenheid van een systeem kan weergeven en daarbij focust het heel erg op hoe ready een systeem is om gebruikt te worden door mensen. Het is dus een aanvulling op de TRL, die eigenlijk aangeeft hoe ready een technologie op zich is. En ze kunnen dus ook naast elkaar worden gebruikt, het is niet dat het een het ander vervangt. En het is bedoelt voor managers van programma's, van projecten die dit soort systemen ontwerpen en ook de systeem ontwikkelaars zelf en mensen die zich bezig houden met validatie. Het is best wel een heavy tool dus het is echt bedoeld voor complex, grote, socio-technische systemen die ook een bepaald risico met zich meebrengen.

Interviewee: Ja.

Interviewer: Verder is het interessant om te melden dat het eigenlijk net zoals TRL toegepast moet worden door het hele proces van de ontwikkeling heen, vanaf het begin, het liefst zo vroeg mogelijk. En ook in lijn met de TRL. Dus als je met TRL 3 bezig bent, zou je eigenlijk ook met HRL 3 bezig moeten zijn. Dus dat is het idee. En dat is natuurlijk belangrijk omdat... nou we weten allemaal dat hoe vroeger je achter bepaalde issues komt, hoe goedkoper en makkelijker het is om het op te lossen. Dus dat is ook 1 van de voordelen. Je kunt voorkomen dat je later met hoge kosten en tijdsuitloop te maken krijgt. En verder kun je natuurlijk het risico HF issues omlaag brengen en het is ook een handige tool om projecten te plannen en bepaalde keuzes te maken. Dus dat is heel beknopt wat het is.

Interviewee: Ja het komt weer terug. Fijn dat je het even verteld.

Interviewer: Mooi. Dan kunnen we naar de tweede pagina en daar staat beschreven hoe de HRL gebruikt zou moeten worden. Dus het begint met stap 1 - het definiëren van het systeem en de grenzen van het systeem. Dat laatste is ook altijd belangrijk, wat neem je mee en wat valt buiten het systeem. Dan is stap 2 om te kijken naar HRL 1 en 2 om te checken of de criteria die bij deze twee levels horen volbracht zijn. Dat is belangrijk omdat HRL 1 en 2 gaan over of je het juiste systeem aan het bouwen bent, terwijl de andere levels meer gaan over ben je het systeem juist aan het bouwen. Dus dat zijn twee verschillende dingen en bij die eerste twee ligt de focus echt op de definitie van het concept. Vandaar dat die twee echt moeten worden volbracht en vervolgens kan het zijn dat je van 2 naar 4 springt, vanwege tijdsdruk, financiële druk,... Officieel moet je ze natuurlijk allemaal volgen maar ik heb al gezien dat het in veel projecten niet zo gaat. Dus dat is stap 2. Stap 3 is om te bepalen welk cluster het invalt. Er zijn 3 clusters - onderzoek & ontwikkeling, technologie demonstratie en productie en implementatie. En bij elke van deze 3 categorieën horen dan 3 van de HRLs. Dus research & development is HRL 1,2 en 3. Dus dan kijk je eerst naar het cluster en vervolgens moet je 1 van die drie selecteren, een van die drie HRLs. En dan ga je bij stap 4 kijken naar alle exit criteria die bij dat HRL horen.

Interviewee: Je zegt je gaat van HRL... Ik ga heel even naar de volgende pagina... Want je zegt dan moet je 1 van de drie kiezen.

Interviewer: Ja dus aan de linkerkant in blauw staat research & development etc. Dus daar moet je 1 van kiezen, waarvan je denkt het valt momenteel onder R&D en daar horen vervolgens drie HRLs bij. Dus je selecteert eerst de categorie, de onderzoeksfase, en dan het HRL.

Interviewee: Ja. Maar je zegt HRL 1 en 2 die je moet in ieder geval gedaan hebben.

Interviewer: Die moet je eerst checken ja. En als je dat niet logisch vindt, mag je dat ook zeggen want dat is ook feedback.

Interviewee: Nou ik snap natuurlijk dat je begint bij onderzoek, R&D uhh maar daarna uhh ik begrijp niet helemaal waarom je daartussen moet kiezen want dat is eigenlijk toch gewoon de logische...

Interviewer: Ah ja oké. Ik bedoel omdat we nu de HRL aan het vaststellen zijn, welke HRL het systeem heeft. Dus we zijn nu even aan het kijken waar het zit op dit moment maar natuurlijk willen we ze allemaal doorlopen.

Interviewee: Oh ja, ik snap het.

Interviewer: Ik ga er nu even van uit dat het systeem al een bepaald HRL heeft en we gaan nu vaststellen wat de HRL is en plannen wat er vervolgens moet worden gedaan.

Interviewee: Ja, oké.

Interviewer: En als je dan het HRL hebt ingeschat, ga je kijken naar al die exit criteria die erbij horen om te kijken of die allemaal volbracht zijn. Zo niet, dan zit je eigenlijk nog in een lager HRL. Dus dan ga je weer eentje naar beneden totdat je bij een HRL komt waarvan je alle exit criteria hebt volbracht. Dat is het huidige HRL. Dan kun je dus gaan kijken van: oké, hoe kan ik dan naar het volgende HRL? Wat moet ik daarvoor doen? Dus dat is het idee.

Interviewee: Ja, oké.

Interviewer: Dus ik zou je willen vragen om mij eerst even kort te vertellen op wat voor project je dit zou toepassen?

Interviewee: Ja ik zit te denken aan een project, dat komt nu meteen in me op, dat is een tool die we aan het ontwikkelen zijn voor de marine. Dat is een software-tool die uhh en die is nu nog in de onderzoeksfase, wij zijn hem aan het ontwikkelen binnen TNO maar dat is altijd tot een concept niveau. En het is de bedoeling dat de industrie dat nu gaat bouwen op basis van onze requirements. Dus dat past volgens mij wel heel goed hierin.

Interviewer: Ja.

Interviewee: Dus daar dacht ik meteen als eerste aan. Wat ook wel interessant is, we doen dus voor de marine heel veel inrichtingswerk, voor inrichten van ruimtes. We hebben nu bijvoorbeeld ook net een concept ontwikkelt en specificatie opgesteld en dat wordt nu ook echt overgedragen aan de industrie om te gaan bouwen. Of om verder eigenlijk productie rijp te maken, dus verder te detailleren. Ik weet niet zo goed... het is een beetje vergelijkbaar niveau uhh maar ik weet niet zo goed welke van de twee nou het interessants is om nu te bekijken. Voor de eerste stap, van waar zitten ze, vind ik het wel interessant om ze even allebei te vergelijken.

Interviewer: Ja oké , is goed.

Interviewee: En dan even kijken waar de meeste uitdaging in zit.

Interviewer: Ja oké. Dus die eerste is een software-tool en die andere is een inrichting?

Interviewee: Ja, een 'inrichtingsconcept' kun je het noemen.

Interviewer: Oké, ja laten we dat doen. De laatste slide definieert specifieker wat Human in the system is, omdat je het heel vaag over human in the system kan hebben maar ook specifiek over workload, situational awareness, error, mental model,.. dus het is eigenlijk een geheugensteuntje op het moment dat je moet vaststellen wat voor risico's of issues er misschien plaats kunnen vinden. Dit helpt dan om daar over na te denken.

Interviewee: Ja, ik vind het een heel mooi schema want ik ken eigenlijk alle onderwerpen wel maar het is leuk om ze nu een keer bij elkaar te zien. Komt het uit de literatuur?

Interviewer: Hij is gebaseerd op de impact wheel. Ik weet niet of je die kent? Die was ik tegengekomen bij een van de projecten hier. Maar ik heb hem een beetje aangepast omdat er wat dingen mistten... dus hij is een beetje aangepast maar hij is daarop gebaseerd.

Interviewee: Ja, leuk.

Interviewer: Oké, dus stap 1 hebben we gedaan. We hebben gezegd aan welk concept we een HRL willen toewijzen. Dus dan kunnen we nu kijken naar HRL 1 en 2 om te checken of deze volbracht zijn.

Interviewee: Ja. Ja. 'Environmental constraints identified' dat is altijd waar we zo'n project mee beginnen inderdaad. Van wat zijn de uitgangspunten, de operationele uitgangspunten, technische uitgangspunten, uhh in wat voor soort omgeving moet die tool of inrichting gaan functioneren. Dus dat is altijd waar we mee beginnen. Nou deze projecten hebben die stap al gezet.

Interviewer: Oké ja.

Interviewee: Het is wel iets wat je natuurlijk altijd weer... je komt tijdens het project toch altijd weer dingen tegen waarvan je denkt ah die moeten we eigenlijk daar nog in vast leggen. Maar dat is meer de feedback loop die je altijd hebt. Dus ja, volgens mij kunnen we zeggen dat HRL 1 volbracht is.

Interviewer: Oké. En ik vind het ook wel interessant om even te vragen waar je aandenkt bij environment? Want daar verschillende de meningen nog wel een over. Dus ik ben wel benieuwd naar je interpretatie daarvan.

Interviewee: Nou ik vergelijk het wel... ken je de SCE methode? Dat is een tool die wij... nou nee, er is geprobeerd een tool te maken, maar de SCE methode is de situated cognitive engineering. Is van Mark Neerincx, een deeltijd hoogleraar bij ons. En uhh daarin heb je foundation fase waarin je eigenlijk je analyse fase doet van elk ontwerp, traject. Maar daar heb je dus die HF uitgangspunten, operationele demands en je technical principles. Het gaat over... en dat zie ik eigenlijk allemaal als in welke omgeving moet het gebruikt gaan worden, dus dat is je doelgroep en wat zijn de kenmerken van de mensen die ermee moeten gaan werken.

Interviewer: Oké, dat ook al?

Interviewee: Ja dus eigenlijk de basic HF. Maar ook dus de operationele eisen, in wat voor situaties en werkdomein moet het gaan plaatsvinden en wat zijn de technische uitgangspunten die je daar in mee neemt, in wat voor technische omgeving, welke... dus eigenlijk basic principles van de TRL misschien wel pak je daar al in mee. De stand van de techniek waar je het mee wilt gaan opbouwen.

Interviewer: Oké interessant. Ik zal het ook eens opzoeken. Oké uhh..

Interviewee: Ja en ik weet niet of ik dan ook het punt van HRL 2 te pakken heb. Dat zijn meer de design principles uhh 'human system performance...' uhh ja kan je dat iets meer uitleggen, wat je hier als design principles ziet?

Interviewer: Ja het interessant is dat ik hem zelf niet ontworpen heb, alleen visueel gemaakt. Dus heel veel dingen zijn voor mij ook niet helemaal duidelijk. Dus voor mij is het interessant wat je interpretatie ervan is om te kijken hoe dat dan verschilt bij verschillende experts en wat het dan zou moeten kunnen zijn.

Interviewee: Ja oké.

Interviewer: Heb je absoluut geen idee van wat ze bedoelen of heb je een idee van misschien bedoelen ze dit, of misschien dat?

Interviewee: Ja het gaat over de uitgangspunten, over hoe je met het systeem wilt gaan functioneren of wilt gaan werken denk ik. Ook hier maken we ook meteen een, moet ik even nadenken hoor, hoe noemen we dat uhh interaction pattern.

Interviewer: Ah ja.

Interviewee: Waarin je eigenlijk wat voor generieke elementen kan je gebruiken in die interactie waar mensen al mee vertrouwd zijn. Hoe snel, efficiënt moet iets gaan werken, hoeveel muisklikken wil je bijvoorbeeld, wil je touch bediening gaan gebruiken of uhh ja voor de inrichtingsvraagstukken vind ik hem... wat nemen we daar als generieke uitgangspunten? Het zijn eigenlijk een beetje de bouwblokken en ik zit nog even te zoeken hoe ik het in een inrichtingsvraagstuk zou omschrijven. Uhh wat wij bij de inrichtingsvraagstukken altijd als uitgangspunt nemen is de samenwerking, zijn de samenwerkingsrelaties, dus de mensen, weten welke mensen in die ruimte moeten werken en het gaat er heel erg om wat de samenwerkingsrelaties tussen die mensen binnen dat team zijn en die moeten ondersteund worden. Dus we maken een link-analyse over wie met wie moet communiceren, wie heel nauw moet kunnen samenwerken, die mensen wil je het liefst naast elkaar hebben, andere mensen moeten oogcontact hebben of eenvoudig met de koppen bij elkaar om een beslissing te nemen. Dat zou ik dan ook in deze stap een plek geven.

Interviewer: Ja, dat is een van de design principles?

Interviewee: Ja. Ja. En 'comparable systems...' nou voor marine schepen wil je natuurlijk dat er met andere schepen ook een bepaalde mate van overeenkomst is zodat mensen die van het ene schip naar het andere schip overstappen niet totaal nieuwe proces hoeven te leren, kennen. Er zit ook een stukje uhh standaardisatie wil je eigenlijk bereiken uhh flexibiliteit is misschien ook een van die uitgangspunten die hierin past. Dat elke werkplek zoveel mogelijk generiek is, zodat je flexibel bent in waar je gaat zitten. Of als er iets uitvalt, dat je dan verder kan blijven uitvoeren vanaf een andere plek. Dat soort elementen zou ik hier onder verstaan. En dat is wel iets wat we ook in onze projecten, in allebei de projecten, uhh is dat de basis waarop je gaat bouwen. En de fase waarin die twee beide projecten zitten, is dat wel iets wat inmiddels wel helemaal duidelijk is. Die stap hebben we wel uhh ik zou denken wel afgerond. Maar dat is ook altijd iets waarmee je door gaat bouwen en dat je erachter komt hier moeten we een design principle over vastleggen.

Interviewer: Dat is goed want op dit moment is het niet echt visueel gepresenteerd dat er ook een feedback loop in zit.

Interviewee: Nou ja, dat weet ik ook niet of dat in een schema als dit ook echt een plek zou moeten hebben. Maar.. ja , ik denk het wel eigenlijk. Zodra je begint met de volgende fase kom je altijd nog dingen tegen die je nog even moet aanscherpen. Anders zou je ook niet efficiënt werken. Als je alles 100% moet afhebben voordat je doorkan naar de volgende fase.
Interviewer: Ja dan ben je nooit klaar.
Interviewee: Nee.

Interviewer: Ja het idee is natuurlijk ook dat je verdergaat en het dan in een specifiekere omgeving, en meer gedetailleerd gaat testen en dan update je natuurlijk altijd dezelfde elementen weer want je blijft kijken naar samenwerking of uhh ...

Interviewee: Ja.

Interviewer: Oké en zou je zeggen dat het nog in onderzoek & ontwikkeling zit of denk je dat het wellicht al de demonstratie van de technologie zit?

Interviewee: Uhh. Even kijken. Demonstratie van de technologie? Ja we hebben wel.. we hebben inderdaad wel demonstrators al gebouwd van de generieke werking, van die tool is inderdaad een demonstrator gebouwd. Zijn we ook vrij snel mee begonnen om ook meteen de intentie van het product, van de tool te kunnen afstemmen van is dat inderdaad wat jullie zoeken binnen de doelgroep, binnen de marine meteen toetsen: we denken in deze richting, zou je daarmee geholpen zijn? En in de inrichtingsvraagstukken bouwen we gewoon die ruimte in 3D, 3D ontwerp en dat kunnen we bij ons in een VR omgeving meteen laten ervaren. Dus mensen kunnen ook rondlopen in de omgeving, met elkaar, en meteen de interactie vanaf hun werkplek eigenlijk beoordelen. De zichtlijnen en de afstanden. Dus dat is de demonstration.

Interviewer: Oké dus het zal wel in HRL 4, 5 of 6 zitten?

Interviewee: Ja.

Interviewer: Oké.

Interviewee: 'design fully matured' nee HRL 6 uhh HRL 6 is het niet, denk ik. Want ja wanneer is het design fully matured? Dat is inderdaad dan wel even... en de volgende stap, 7 'evaluation of established design in operational environment' dan ga je het echt in de operationele omgeving evalueren... uhh nou die BDR die hebben we wel echt op een gegeven moment uitgerold op het schip met echt alle actoren die mee moesten doen hebben we ook echt een interface gegeven en zo realistisch mogelijk geprobeerd te evalueren. Dus ja het was wel de echte omgeving maar wel een gesimuleerde situatie. Dus je geeft de mensen wel een scenario van ga dit nu met elkaar doorlopen. Wel levensecht geprobeerd te maken maar ik denk dat dat dan toch nog wel in HRL 5 zet.

Interviewer: Oké. Ja dat is wel een interessante, dus de tool zelf was nog niet heel volwassen maar de omgeving waarin je het test is eigenlijk al heel erg nauw aan de werkelijkheid.

Interviewee: Ja. En als we dan kijken naar HRL 7, daar staat 'system prototype demonstrated' het was wel een prototype van het systeem dus uhh ... en TRL 9 is het actual system. Nou ik... ik denk dat we dat toch nog wel onder TRL 5 moeten blijven zien. Want kijk, het idee is dat nu de industrie het gaat bouwen en dan werkt het pas echt zoals het bedoelt is. En daarvan moet je een eerste prototype... ik denk dat dat dan TRL 7 zou zijn. Ik zie daar wel een soort knik tussen. Tussen wat wij hebben gedaan en dat de industrie de officiële tool gaat bouwen. Daar tussen 6 en 7 zit ook een mooie witregel. Daar zie ik ook echt wel die knip.

Interviewer: Oké. Dus misschien betekent die witregel wel dat het naar de industrie wordt overgedragen of?

Interviewee: Ja.

Interviewer: Oké

Interviewee: Zo zou ik het voor nu allebei de projecten wel interpreteren.

Interviewer: Oké dan kunnen we eens kijken naar HRL 5. Welke exit criteria daarbij horen. En kijken of dat inderdaad allemaal behandeld is.

Interviewee: Ja. Ik moet even kijken hoor. Vooral de rode kolom of wil je...?

Interviewer: Laten we ze allemaal doorlopen. Is er een reden dat je zou zeggen: ik focus op de rode kolom?

Interviewee: Nou nee, maar ik merkte dat de titels van de kolommen buiten beeld vallen dus ik was even benieuwd naar wat is het verschil tussen de kolommen. Dus ik moet even goed doorlezen nu. Zodra ik naar 5 ga ben ik ze kwijt. Wel handig om even die verschillende kolommen erbij te hebben. Dus HRL 5 uhh daarvan willen we nu even kijken of we die allemaal al doorgelopen hebben. Ja dit is juist iets waarvan ik denk uhh oh hier staat, even kijken... ja dit [Potential human-system performance issues and risks identified] is iets waar je wel heel intensief over na aan het denken bent maar ik moet zeggen, zodat je dus, want we zijn dus nu, bijvoorbeeld die inrichtingsvraagstukken zijn we nu ook samen met de industrie een stapje verder aan het brengen. Dus de industrie gaat verder engineeren, de werkposities en de inrichting en uuh daar heb je eigenlijk elke keer weer, maak je daar de afweging nou als het net iets anders uitpakt wat zijn daar de risico's van voor de gebruiker. Dus ik vind dat identificeren van issues en risks... die zit dus op meerdere niveaus, die zit eigenlijk op alle niveaus denk ik. Ja. 'conformance of prototypes to human systems performance requirements and principles' uhh ja daar ben je eigenlijk steeds mee bezig.
00:40:30

Interviewee: 'strategies to support human use recommended' uhh ja ik vind het moeilijk om te zeggen dit is afgerond eigenlijk.

Interviewer: Ja? En waarom?

Interviewee: Omdat het product nog niet helemaal uitontwikkeld is. Er zijn nog allerlei keuzes die je kunt maken op detail niveau die hier ook invloed op hebben. Dus...

Interviewer: Maar zoals je zelf zei, het komt ook weer terug in het volgende level.

Interviewee: Ja precies.

Interviewer: Dus het gaat er meer om of je het afgerond hebt op het niveau van een prototype in een gesimuleerde omgeving.

Interviewee: Op dat niveau heb je dat dan afgerond. Ja. Ja. Nou even naar de volgende, de working environment uhh 'environmental constraints' ja dat is ook iets... ja die loopt ook gewoon door natuurlijk in de verschillende niveaus eigenlijk uhh voor zover we die kennen, zijn ze meegenomen. Het hele schip is nog in ontwikkeling dus er kunnen steeds weer wat wijzigingen komen dus dat is ook iets... maar dat loopt ook gewoon... dus dat is niet echt iets waar uhh ik loop ze zeg maar even langs om te kijken waar je dan echt kan zeggen dit is gedaan uhh. 'strategies to accomodate manpower and personnel updated' uhh ja, is ook meegenomen 'strategie to accomodate training updated' daar wordt over nagedacht 'procedure for operator and maintainer roles...' daar wordt ook over nagedacht 'task flow and sequencing optimised' uhh ja dat wordt, dat zijn allemaal aspecten waar je op dat niveau gewoon over na denkt. Ja. 'cognitive walkthrough' das is echt iets wat we in zo'n evaluatieomgeving doen.

Interviewer: Oké. Dat is ook iets wat jullie gedaan hebben met die BDR tool?

Interviewee: Ja. Dan ga je langs van wat wil je hier dan in kwijt, wat mis je nu, wanneer wil je extra informatie, dus dat hele proces over hoe ze door die tool heenlopen die wordt dan echt met de mensen doorlopen.

Interviewer: Ja. Met behulp van die simulatie of wordt dat eigenlijk gedaan voordat de simulatie plaatsvindt?

Interviewee: Uhh allebei. Je probeert dat wel bij het bouwen van het concept al die input zoveel mogelijk mee te nemen. Het proces wat ze nu doen, dat ze nu gebruiken, nu doorlopen, zo goed mogelijk kennen om te weten wat je kan digitaliseren. En ze dan als je dat heb gebouwd nog een keertje toetsen op papier en dan in zo'n prototype ook echt laten ervaren of in een gesimuleerde omgeving ook echt laten doen. Dan komen er nog weer nieuwe dingen uit.

Interviewer: Ja.

Interviewee: Dus dan komen we bij 6. Dus 5 hebben we wel... die is echt wel afgerond.

Interviewer: Ja. Oké.

Interviewee: Ja. Ik zit even te zoeken waar de grote verschillen zitten, tussen 5 en 6. Het een is, ja, design fully matured uhh...

Interviewer: Kan het zijn dat die misschien niet helemaal duidelijk zijn, de verschillen tussen 5 en 6?

Interviewee: Ja 5 is echt het evalueren van het prototype uhh het is dus eigenlijk, het is echt wel een bepaalde fase, een bepaald moment van nu gaan we het evalueren en uhh en daarna maak je dan nog een ontwerp... een soort optimalisatieslag om een uhh fully matured design te hebben. Maar fully matured klinkt nogal definitief en eigenlijk, zeker die tool, het is ook iets... je bouwt de eerste versie en je weet dat je... we zijn een ontwikkeling gestat en hopelijk loopt het nog 10 jaar door, of nog langer. Dus fully matured klinkt te definitief. Dus dat is dan die versie maar ook uhh het is echt wel een heel complex uhh zeker die BDR tool, dat grijpt in op heel veel verschillende systemen, moet eigenlijk geïntegreerd worden met heel veel verschillende systemen. Het idee is dat het eerst maar even stand-alone moet kunnen gaan werken. Dus dan kun je zeggen nou de stand-alone versie, die moet fully matured zijn maar uiteindelijk zie je heel veel ontwikkelingen in de toekomst.

Interviewer: Oké. Ja ik ben wel benieuwd wat... ja ik kan me daar wel iets bij voorstellen dat design fully matured... dat is eigenlijk iets te definitief. Maar wat zou er dan tussen HRL 5 en 7 kunnen zitten? Zit daar nog iets tussen? Of ga je gelijk van het evalueren van een prototype in een gesimuleerde omgeving naar het valideren van het uiteindelijke systeem er rekening mee houdende dat daar nog dingen aan kunnen veranderen maar in ieder geval het uiteindelijke concept in een operationele omgeving. Wat zou daar tussen kunnen zitten?

Interviewee: Ja daar zit zeker wat tussen. Uhh ik zit ook even te kijken naar de TRL...

Interviewer: Ja?

Interviewee: TRL 5 is technology validated in relevant environment dus dat komt inderdaad wel overeen met dat prototype uhh en bij TRL 6 technology demonstrated in a relevant environment uhh en 7, system prototype demonstrated uhh ik vind TRL 6 ook niet zo'n hele logische stap eigenlijk. Als we kijken naar de HRL, je hebt een evaluatie... dus laten we even het inrichtingsontwerp nemen uhh dan... ik denk dat die even iets makkelijker is uhh 'evaluation of prototypes' nou daar kunnen we inderdaad prototypes evalueren, dan geven we er wel een klap op, dan zeggen we nou dit is tot waar we het uitwerken tot dit detailniveau, we hebben de input van de evaluatie meegenomen en dat komt in een rapport te staan en dat gaan ook met specificaties naar de industrie. Dus dan kan je zeggen... maar matured vind ik niet helemaal... maar het is wel een...

Interviewer: Wel een soort van beslissing, wat je zegt, klap erop geven klinkt wel als HRL 6?

Interviewee: Ja ik denk dat je dat wel als een soort afgebakend detailniveau eigenlijk, zo zie ik dit dan vooral, kan zien. dan heb je het concept vastgelegd eigenlijk en uhh ja zo noemen wij het dan. En dan gaat het inderdaad met de specificaties naar de industrie. Die gaat dan echt een system prototype bouwen. Maar er zijn altijd wel dingetjes die nog even anders moeten en uhh dus je grijpt daar wel af en toe terug op HRL 6 om toch nog wat aanpassingen te doen of toch dingen die nog niet helemaal in detail waren uitgewerkt, die moeten dan in HRL 7 uhh uitgewerkt worden en dat ga je ook weer evalueren.

Interviewer: Wellicht weer in een simulatie? Dus dan zijn we weer in HRL 5 misschien?

Interviewee: Ja want in operational environment, dan moet je het aan boord van het schip zetten. En dat wordt, dat is dan nog niet denk ik uhh ik zit even inderdaad... 'production and deployment' waar zit dan het echte deployment? Het wordt aan boord gezet, dan wordt het nog een keer getest van doet het het allemaal. Maar dat is meer, dat is echt de technische check dan. Of het allemaal werkt. Want die HRL, ja wat ga je dan op dat moment nog... ja, wat je ook wel merkt is als we een nieuw inrichtingsconcept hebben bedacht, zeg je nou die mensen moeten dan straks ook iets anders gaan werken dan dat ze nu gewend zijn. En dat ga je natuurlijk pas aan boord echt ervaren want ja, dan pas zit je in de totale operationele omgeving. Dus eigenlijk is dat HRL 7 dan. Ik weet alleen dus niet of ze dan nog helemaal synchroon lopen met de TRL.

Interviewer: Ja oké. Omdat HRL 7 bij TRL 9 zou zitten of?

Interviewee: Ja. Dat gevoel krijg ik nu een beetje. Maar goed, misschien is dat een hele grote stap. Want je zet het dan aan boord, en het is zo goed mogelijk uitgedacht hoe het allemaal technisch moet werken en dan kan je er pas operationeel mee aan de slag. En dan kan het zijn dat je dat gaat van ja maar die mensen gaan het werk op die manier verdelen. Maar dat werkt dan toch niet helemaal handig dus gaan ze zelf daar een iets andere operationele procedure bij bedenken om dat goed te laten werken. En dan pas heb je je human readiness... kan je daar pas op afstemmen.

Interviewer: Ja precies. Er staat wel bij procedures, roles and responsibilities bij HRL 9 'procedure for operator and maintainer updated if required' dus er is daar wel een soort van, ja wat je zegt, het kan zijn dat uiteindelijk de rolverdeling anders uitkomt dan je van te voren had bedacht.

Interviewee: Ja. Inderdaad. Dat zit in HRL 9. Ja. Ja precies. Want daar, eigenlijk zit daar in die lichtblauwe kolom uhh daar zit een beetje de uitdaging. Je hebt wel bedacht dat als je mensen iets anders gaat ondersteunen dat ze dan ook iets anders het werk gaan uitvoeren. En in HRL 9 heb je pas de definitive nieuwe procedure.

Interviewer: Ja. Die misschien ook blijft veranderen...

Interviewee: Ja dat kan ook nog een tijdje, ja dat kan ook nog langere tijd uhh zoeken naar een optimum. Ja. Want in TRL 7 dan 'procedures for operator and maintainer roles tested... operator and maintainer scenarios and tasks tested' die testfase die wordt niet altijd, die kan niet altijd helemaal officieel worden uitgeprobeerd, bij een inrichting dan. We hebben wel eens een inrichting helemaal na-gesimuleerd uhh maar dan zit, dat is dan TRL 6 meer, daar test je dat ook al.

Interviewer: En waarom kan dat niet helemaal worden getest in een operationele omgeving? Is daar een achterliggende reden voor?

Interviewee: Uhh nou de inrichting is nog al definitief. Dus als je alles inricht, staat het allemaal vastgeschroefd aan de vloerplaten. Dus je kan niet zomaar zeggen ik ga die ruimte nog eventjes wat anders rangschikken. Dat is op een schip natuurlijk misschien anders dan in een kantooromgeving waar je kan zeggen nou het werkt toch niet dat al die tafels in een ronde opstelling staan, ik ga ze toch in rechte rijen zetten. Ja, aan boord van een schip is dat uhh kan dat dus niet. Dus daar zit misschien uhh het is dus de vraag welke vrijheid er nog zit in het ontwerp om die procedures dan nog aan te passen.

Interviewer: Op zich zitten procedures niet vastgeschroefd toch? Of zitten die vastgeschroefd samen met de techniek?

Interviewee: Nee, nee, nou het is wel zo dat je... je probeert die procedures zo goed mogelijk te ondersteunen met je inrichting. En als je merkt dat die inrichting... dat heb je dan uitgedacht om dat op elkaar af te stellen, dus als dan blijkt dat die procedures toch niet zo lekker lopen, dan kan je de inrichting niet meer op die nieuwe procedures afstellen. Daar zit eigenlijk een stukje gebrek aan flexibiliteit denk ik.

Interviewer: Ja dus het goede is natuurlijk dat je hier al naar kijkt vanaf HRL 1/2.

Interviewee: Ja.

Interviewer: Maar zou je dan zeggen van dit stopt eigenlijk vanaf HRL 7, of is het wel goed om het erin te houden voor het geval dat er bepaalde dingen wel nog aan te passen zijn? Wat zou zeg maar de oplossing hiervoor zijn?

Interviewee: Uhh dan moet ik even goed nadenken. We zijn inderdaad vanaf begin af aan aan het nadenken... wat bijvoorbeeld een onderwerp is, is om met minder mensen te gaan varen. Ze proberen het werk zodanig te verdelen zodat je met minder mensen het werk kan doen. En daar pas je dus de procedures op aan. En dan moet er ook met automatisering ondersteunt worden. En uiteindelijk wordt er natuurlijk altijd de oplossing gevonden. Dus ik zit even van, natuurlijk worden er oplossingen gevonden maar hoe ga je dat hier nou in uhh wat beter in vertegenwoordigen, zichtbaar maken.

Interviewer: En wat voor oplossingen worden er bijvoorbeeld gevonden? Zijn dat wel procedures of gaat het misschien meer over training,... wat voor oplossingen zijn het?

Interviewee: Nou wat je eigenlijk vaak ziet, is dat ze uiteindelijk dan toch, als het niet werkt, dan toch een extra persoon erbij plaatsen. Maar die heeft dan eigenlijk geen stoel. Daar zit ik een beetje mee, van hoe ga je dat dan uhh... Ja dat probeer... het punt is dat we weten natuurlijk dat het spannend is of iets wel of niet gaat werken, want je kan heel veel uitdenken van tevoren maar als het pas in de praktijk gebruikt wordt, dan pas ervaar je of er knelpunten zijn en hoe je di het beste kan oplossen. Dus houden er altijd wel rekening van nou misschien is die persoon toch wel nodig. Dus er zit altijd wel een fall-back scenario al in het ontwerp. Dus misschien moet ik ook niet te kritisch zijn want dat is wel iets wat eigenlijk al altijd genomen wordt, van nou laten we er maar van uitgaan dat we daar nog wel een flexplek beschikbaar hebben. Die kunnen we óf gebruiken voor die ene persoon die we toch nodig hebben of anders komen er in de toekomst wel andere behoeftes om een extra mee te nemen. Dus misschien is het ook.. moet ik daar niet te kritisch in zijn. Dus HRL 7 is evaluatie... 'established design' uhh 'validation of final system' uhh dus die validatie is eigenlijk de eerste keer dat je het echt gaat uhh echt gaat inzetten... zo kan ik het wel...? Dan ga je het echt valideren? Van werkt het nou goed? En HRL 9, dan wordt het gewoon langdurig gebruikt en dan ben je erachter nou dit is echt een goede oplossing, dit kunnen we ook op andere schepen uitrollen. Zo zou je het kunnen... En dat doen we ook inderdaad, die monitoring of fielded system, dat is ook de check achteraf van zijn mensen nou tevreden met ons ontwerp, wat kunnen we hier nog weer van leren voor nieuwe inrichtingen, wat werkt heel goed, wat zijn nog kritische punten uhh dat zou ik echt als een uhh ook weer een ja, dat geeft ook weer input voor nieuwe ontwikkelingen.

Interviewer: Ja. Dat linkt eigenlijk weer terug naar een HRL 2 op het moment dat je gaat kijken naar comparable systems.

Interviewee: Ja. Precies, en de principles die je wilt vasthouden in nieuwe ontwerpen, ... ja. Ja. Uhh 'interaction design patterns' was trouwens het woord dat ik zocht, wat we gebruiken bij die design principles.

Interviewer: Ik heb een paar artikelen gelezen over design patterns met automatisering, tussen de automatisering en de mens en hoe je er dan voor kunt zorgen dat ze goede teamplayers zijn.

Interviewee: Ja we zijn inderdaad nu allerlei team design patterns aan het ontwikkelen, waarbij het team kan bestaan uit mensen of automatisering of een combinatie met automatisering. Dus dat zit ook wel echt op het HRL 2 niveau. Dat is echt verkennend onderzoek, wat we aan het doen zijn om ook weer toe te kunnen passen in dit soort toegepaste projecten.

Interviewer: Oké. Nou we hebben er volgens mij wel goed doorheen gelopen. Ik was wel benieuwd wat je dan nu als, dat project hebben jullie nu overgedragen naar de industrie, maar wat zouden dan volgens de HRL de volgende stappen zijn. We stellen dus vast dat het in HRL 5, of wellicht 6, zit...

Interviewee: Ja, het zit inderdaad uhh ja ik zou op uhh eigenlijk na 5, geven we hem over, denk ik.

Interviewer: Na 5 al?

Interviewee: Ja want die laatste ontwikkelstap uhh ja dat vind ik nog wel een lastige, of je ... ja ik vind HRL 6 een beetje te definitief.

Interviewer: Maar misschien moet daar dan wel een andere naam voor komen?

Interviewee: Ja precies. Nee maar ja dan is het wel... je levert wel een soort af product, een afgebakende versie, die lever je dus eigenlijk aan de industrie en die gaat dan een echt prototype bouwen. Dus wij zijn nog van de technology demonstration uhh dus TRL 6 inderdaad en HRL 6, dat is wel waar wij denk ik, tot waar wij gaan. En wat we nu dus nog doen met de industrie is dat we de industrie begeleiden en eigenlijk een review rol hebben om alles wat wij in al die onderliggende HRL niveaus hebben ontwikkeld en kennis hebben, om dat zo goed mogelijk te blijven waarborgen in die volgende... Want het punt is, in die industrie ligt de focust op de TRL, en minder op die HRL. Het zijn echte techneuten die er dan mee aan de slag gaan en daar moet je dus, dat is misschien ook wel een interessant... die HF kennis, die moet wel echt heel goed vertegenwoordigd blijven zodra je van het systeem een echt prototype gaan bouwen. Dus daar zien we dan nog echt een rol als, voor ons, om die HF kennis te blijven borgen.

Interviewer: En zijn jullie dan wel betrokken op het moment dat die technologie dan geproduceerd is? Doen jullie dan weer een simulatie daarmee? Of ja, simulatie is het dan niet meer, doen jullie daar dan nog een test mee? Of is het dan helemaal uit jullie handen?

Interviewee: Nee dat ligt een beetje aan... meestal krijgen we wel nog een review moment waarbij we het kunnen toetsen, of het inderdaad nog steeds voldoet aan alle ergonomisch uitgangspunten en operationele uitgangspunten en uhh dus uhh dus in principe hebben we die rol altijd. Maar goed, soms lopen dingen anders. Maar die, ja, dus dat is dan wel die eigenlijk die uhh in HRL 7, is dat de evalution of the established design uhh wij beoordelen dan een prototype, en daarna gaat, op basis daarvan, krijgt de industrie ook de opdracht van nou dit mag je gaan bouwen en neer gaan zetten. En dat zou dan inderdaad de grens zijn van 7 naar 8.

Interviewer: Ja precies. De go/no-go voor de industrie?

Interviewee: Ja.

Interviewer: En die review is HRL 7 of doen jullie... jullie doen een review van het prototype, maar wordt er daarna ook een review gedaan van het uiteindelijk systeem of?

Interviewee: Nou dat heet de... je hebt dan verschillende acceptancy tests.

Interviewer: Ja?

Interviewee: Die worden door de marine zelf gedaan. Dus de factory acceptancy test, dus dat is denk ik dat prototype, in de fabriek, een laatste evaluatie van is dit prototype geschikt om in productie te nemen. Dan heb je nog een harbour acceptancy test dus dan wordt het in de haven... volgens mij is het dan al aan boord ingebouwd. En dan heb je nog een ship acceptancy test dacht ik uhh ja dat moet ik eventjes... maar eigenlijk moet je dus ook nog, het moet dus aan boord ook nog blijven werken, wanneer het schip gaat varen. Als het niet zeewaardig blijkt, dan...

Interviewer: Dus die factory acceptancy test dat is meer die HRL of TRL 7, dan harbour acceptancy is meer TRL of HRL 8 en dan ship...

Interviewee: Ja. Ja.

Interviewer: Oké. En jullie zijn betrokken bij die factory acceptancy test?

Interviewee: Ja. En daarna gaan we aan boord om nog eens na te vragen aan de operationele mensen van hoe heeft het gewerkt. Dat is echt die HRL 9.

Interviewer: Oké. Interessant. Dus ja het loopt wel redelijk in lijn met de werkelijkheid?

Interviewee: Ja.

Interviewer: Oké. En zou je zoiets als dit overzicht, deze tool, kunnen gebruiken in toekomstige projecten? Of heb je zoiets van, nou dit maakt het alleen maar complexer?

Interviewee: Nou uhh ik vind het uhh het is natuurlijk best wel complex. Het dwingt je om echt in hele kleine stapje te gaan nadenken. Dus dat kan het aan de ene kant wel complex maken. Maar ik vind het dus heel krachtig dat het tegen de TRL niveaus aangezet is. Want die zijn algemeen geaccepteerd. En daarmee kan je wel ook juist de techneuten, de industrie af en toe weer even bij de les roepen van wacht even, jullie kijken nu naar de technologie maar die HF kant is net zo belangrijk. Want als dat niet werkt, dan werkt de technologie ook niet. En soms is dat wel nodig. Want we lopen er toch wel vaak tegen aan dat de technologie toch wel belangrijker wordt geacht dan de HF aspecten. En dat is dus een beetje een eeuwenoude strijd volgens mij. En dan zou dit daar dus wel in kunnen helpen, om het ook een formele status te kunnen geven. Dus daar zie ik het eigenlijk vooral in, als communicatie middel.

Interviewer: Oké. Ja, goed punt. Dus dat is de kracht ervan. En wat zou je dan zeggen wat de zwakte is, of wat eventueel puntjes voor verbetering zijn.

1:10:27

Interviewee: Nou wat ik merk is dat de rijen uhh niet zo veel uhh daar zit heel veel overlap in de uhh in de tekst ook gewoon dus dat maakt het, vind ik, toch wel lastig om die verschillende niveaus echt te kunnen onderscheiden. En ja, als er zoveel dubbele tekst in staat, dat maakt het model ook minder lekker leesbaar. Dus dat zou denk ik nog iets beter kunnen.

Interviewer: Efficiënter ook...

Interviewee: Ja.

Interviewer: Ja. Oké.

Interviewee: Ik weet niet meteen hoe dan, maar...

Interviewer: Oké. Even kijken of ik verder nog brandende vragen had... Je zei eerder het is best complex uhh is het overbodige complexiteit of is het ook gewoon een complex proces? Denk je dat het versimpeld zou kunnen worden? Afgezien dan van redundancy in de tekst.

Interviewee: Ik denk dat als het wat versimpeld is, dat het wel weer beter communiceerbaar wordt. Want als je dit aan mensen met een technische achtergrond geeft van kijk dit is de HRL tegen de TRL, dan... ik vind de kolommen wel heel krachtig, want dat zijn wel alle elementen waar je op moet letten voor de HF categories, dat vind ik een hele krachtige uhh maar het verschil binnen die kolom, tussen de verschillende niveaus, daarvan denk ik dat is niet zo lekker communiceerbaar.

Interviewer: Ja. Ja oké.

Interviewee: Dat vond ik zelf ook al lastig, om daar echt onderscheid in te maken.

Interviewer: Snap ik ja.

Interviewee: En ik zie nu wel per uhh 7, 8 en 9 hebben eigenlijk heel veel overlap. Dus ja daar, die overlap, denk ik dat daar nog een stapje in gezet kan worden.

Interviewer: Ja precies. En de originele HRL was een Excel file en daar kwamen dus elke keer dezelfde vragen terug maar je zag het verband daar niet tussen dus het waren dus elke keer dezelfde vragen maar het duurde even voordat daarachter kwam. Dus ik heb het in categorieën gegoten om wel te laten zien van kijk dit zijn dingen die constant terugkomen maar dan dus op een ander HRL, in een andere testomgeving. Dus dat is de weg hier naartoe maar ik snap wel je feedback en ik denk inderdaad dat de volgende stap dan is om te kijken hoe dat dan logisch eruit kan zien uhh ja, nu ga je eigenlijk alles heel gedetailleerd doorlezen en dan kom je erachter van: oh maar het is hetzelfde, oké wat is dan het verschil.

Interviewee: Ja precies. Ja komt ook natuurlijk omdat ik er nog niet echt van tevoren doorheen had kunnen gaan. Want dan had ik meteen al wat duidelijker gezien van: oh wacht dit is de overlap en uhh..

Interviewer: Maar dan nog, dan is het alsnog lastig te zeggen wat dan het verschil is tussen de levels?

Interviewee: Ja. Ja dat vind ik nog wel een lastige. Maar ik denk wel, doordat het zo in een overzicht allemaal geïntegreerd is, maakt het al wel weer een stuk overzichtelijker. Als je zegt dat er allemaal vragen langs komen waarvan je eigenlijk niet goed weet waar je antwoord op geeft, waar het samenhangt, dan is dit denk ik al wel een stuk, echt wel weer een stap in de goede richting.

Interviewer: Ja het gaat de goede richting uit, maar er is natuurlijk nog voldoende ruimte voor verbetering. Oké. Nou ik heb goede feedback ontvangen dus hier kan ik weer wat mee, dat is mooi.

Interviewee: Ja. En de laatste kolom hebben we helemaal... ja alleen die cognitive walkthrough hebben we even...

Interviewer: Ja die toolbox, daarbij is eigenlijk mijn vraag, zou je die toolbox kunnen gebruiken, zou je die wíllen gebruiken, is het handig? Of heb je zoiets van nou wij hebben daar al bepaalde frameworks voor die we gebruiken?

Interviewee: Ja ik ken niet alles wat hierin staat.

Interviewer: Ja het is luchtvaart-gericht dus het kan zijn dat...

Interviewee: Ja hier herken ik maar heel weinig van dus uhh als ik er zo doorheen ga, dan denk ik dat dit best wel domein-afhankelijk is en uhh wat dan bedrijven of de toepasser eigenlijk zelf zou kunnen... want ik denk dat wij hier wel heel veel dingen zelf in zouden kunnen zetten maar dat is dan weer heel specifiek voor het domein waarin wij werken.

Interviewer: Ja oké. Maar zou dat nuttig zijn? Als je bijvoorbeeld een maritieme toolbox hebt, of een luchtvaart toolbox, ...

Interviewee: Ik denk dat wij hier best wel onze eigen toolbox weer aan kunnen hangen. En dat geeft voor ons weer een heel mooi communicatiemiddel, naar de opdrachtgever van waarom willen we dit doen - nou om deze HRL niveaus te kunnen invullen en afhandelen. Dus ik vind de link naar die tools wel heel nuttig. Maar ik zou dan nu denken van dan wil deze tool, dit schema wil ik hebben, maar nu wil ik daar zelf mijn eigen dingen in kunnen zetten. Als een soort van open ruimte.

Interviewer: En heb je dat dan alleen bij de tools? Of geldt dat misschien ook voor andere kolommen? Of heb je zoiets van nou dat moet vast staan voor elke industrie?

Interviewee: Nou ik vind dat het gewoon een algemeen model is, over alle industrieën heen. Dat is denk ik ook de kracht van de TRL. Ik zou het toch echt op hetzelfde niveau houden als de TRL want dat is een wereldwijd bekend model en dat vind ik heel krachtig, dat we dit er tegen aan kunnen zetten. Dus dat mag wel algemeen blijven. Maar die link, die stap naar de toolbox die, die denk ik dat die veel domein-specifieker is. Of daar moeten wat... ja, in ieder geval veel dingen die hierin staan... ja kijk interviews, debriefing, over-the-shoulder observation... dat ken ik natuurlijk wel. Maar er staan ook veel tools in die wij in ieder geval helemaal niet gebruiken. Dus ik kan me wel voorstellen dat je dus een hele algemene variant hebt en eentje die domein-specifiek is... dat je dat misschien als twee versies kan beschouwen.

Interviewer: Ja oké, een soort classificatie, dus interviews en dat soort dingen en dan kun je er zelf nog domein-specifieke dingen aan toe voegen.

Interviewee: Ja. Ja.

Interviewer: Ja oké, interessant.

Interviewee: Ja want ik denk dat als je hem generiek houdt, dan zou die kolom eigenlijk ook generiek moeten blijven.

Interviewer: Oké. En heb je verder nog dingen waar je aan denkt die we nog niet besproken hebben?

Interviewee: Nou op dit moment niet.

Interviewer: Nou je mag me altijd nog mailen mocht er iets te binnen schieten.

Interviewee: Ja mocht ik nog een nabrander hebben, dan stuur ik hem naar je toe.

Interviewer: Is goed. En ja ik zal na deze ronde van evaluaties weer wat dingen veranderen en dan kan ik het doorsturen met een stukje uitleg. Is misschien ook interessant om aan collega's te laten zien en ja het zou natuurlijk mooi zijn als het op een keer als communicatie tool kan worden gebruikt, daar zou ik alleen maar blij mee zijn.

Interviewee: Ja daar zie ik echt het nut van in, dus dat is leuk.

Interviewer: Nou heel veel succes dan nog.

Interviewee: Ja bedankt, en heel erg bedankt voor het vrijmaken van je tijd en voor het meedenken, erg nuttige feedback. Dus ja, heel erg bedankt!

Interview with railway experts from ProRail (Dutch)

Interviewer: Nou dit is gewoon even een stukje introductie... Een schaal om aan te geven hoe volwassen een technologie is, maar dan niet of de technologie zelf per se werkt maar vooral ook wat er gebeurd als je er een operator in zet. En het is dus een tool die bedoeld is voor aan de ene kant mensen die het systemen ontwerpen, die de validatie uitvoeren en aan de andere kant ook managers, die het project of programma managen. Het is best wel een heavy tool dus het is vooral bedoeld voor hele complexe system waarbij er risico is gemoeid. En het is geen vervanging van de TRL, het is een aanvulling. Het liefst wil je ook dat ze alle twee tegelijk oplopen, dus als je in HRL 5 bent, zou je ongeveer ook in TRL 5 moeten zijn. En verder natuurlijk hoe eerder je het toepast, hoe meer nut het heeft, hoe meer kosten je eventueel kan besparen omdat je tijdig bepaalde issues vindt. Dus dat is heel kort de HRL. En dan was ik benieuwd of jullie dus al een project in gedachten hadden wat hier in past, waar dit op toegepast zou kunnen worden? En of jullie daar, jullie kunnen natuurlijk niet alles prijsgeven, maar in ieder geval mij een idee kunnen geven over wat voor systeem het gaat.

Interviewee 1: In principe hebben we geen project wat helemaal tot TRL 9 gaat. Dat is misschien de eerste opmerking. Als je naar beneden scrolt [Interviewee 2], ...

Interviewer: Het hoeft ook geen uhh we gaan nu ad hoc deze tool toepassen dus het is niet zo dat we per se een project moeten hebben die al ergens zit qua niveau uhh ja het is meer om te kijken naar hoe is het proces in het echt gegaan en hoe zouden we het doen met de HRL. Op die manier. Dus voel je vrij om elk project te kiezen. Zolang het om een complex, socio-technisch systeem gaat dan zitten we goed.

Interviewee 1: Nou in principe zou je kunnen zeggen als die toekomstige veranderingen die we bij het spoor uhh wat we willen gaan implementeren dus bijvoorbeeld nieuwe veiligheidssystemen, het ERTMS, de zelfrijdende treinen die gaan we hier doorheen laten lopen of die gaan door deze fases heen dus wat ons uhh we hebben een ERTMS in gedachte als casus uhh en dan kunnen we denk ik ook wel zeggen welke fases we tot nu toe hebben gedaan. Even kijken. De eerste fase was eigenlijk dat we een impact analyse gingen doen met uhh dus de vraag was daarachter wat voor impact heeft uhh dus ERTMS als nieuwe systeemwijziging, beveiligingssysteem voor impact op het werk van de operator, in dit geval de treinverkeersleider. En ik denk dat je, als je dat wilt plotten, dat je zegt dat zit dan eigenlijk al in level 2.

Interviewer: Oké. En op welke informatie baseer je die keuze? Is dat algemeen over waar HRL 2 over gaat of kijk je dan wat er rechts staat? Hoe schat je dat in?

Interviewee 1: Met name word ik getriggerd door wat er staat in die 4e kolom: 'basic HF research documented and/or conducted' en uhh wat ik daaronder versta is best wel fundamenteel onderzoek. En vaak kijk je dan naar de huidige situatie, of je kijkt naar bepaalde context waarin je denk ik met name eigenlijk in uhh je zit in huidige beperkingen denk ik, of limitaties waar je dus eigenlijk onderzoek doet. Dus bijvoorbeeld, stel dat je kijkt naar uhh ja werkbelasting is ook een steeds terugkerend onderwerp. We willen graag een tool ontwikkelen waarin je werkbelasting kan identificeren bij een operator en dat je aan de hand daarvan, dat een teamleider kan ingrijpen en als een soort van een decision support tool een meter heeft en kan ingrijpen als een treinverkeersleider hulp nodig heeft. En dan zou ik kijken van wat is werkbelasting nou precies, hoe ziet de huidige situatie eruit. Ja dan zit je denk ik echt heel erg op dat niveau nog, op HRL 1.

Interviewer: Oké. En zijn dan al die statements bij HRL 1 afgerond of moet er daar nog wat gebeuren om HRL 1 af te ronden?

Interviewee 1: Wat bedoel je?

Interviewer: Dus nu hebben we het over de ERTMS. En dat zit nu ongeveer in HRL 1/2?

Interviewee 1: Nou ja, als het hebt over het begin van waar zitten we nu en waar willen we naartoe gaan. Als je het hebt over nu, dat je in 1 zit. Maar ik kan me ook voorstellen, als ik kijk naar de overige punten die er in staan, bijvoorbeeld dat er staat ook dat uhh ik zie er elementen uit de CWA staan, zit je met HRL 1 eigenlijk al in het begin van de eerste CWA fase. Dus in principe zie je ook wel hier uhh van misschien is dat al een clash. Misschien zou er ook een HRL 0 moeten zijn. Want als je het dan hebt over de huidige situatie en hoe je van huidig naar toekomst gaat, dat wat ik net beschreef eigenlijk HRL 0 zou moeten zijn en HRL 1 als ik kijk naar de CWA componenten dat het al een eerste schets is van de toekomst. Dus ik denk dat wat er in de eerste kolom staat, of in de vierde kolom die 'basic HF research documented and/or conducted' dat vind ik uhh als ik kijk naar de CWA uhh dan weet ik eigenlijk niet zeker uhh kan ik dit op meerdere manieren interpreteren. Dan kan het zijn dat je de eerst fase hebt van je CWA, van een toekomstige verandering uhh ja dat dan met name. Je hebt het al over toekomstige verandering want als we kijken naar die andere niveaus, die bouwen voort op de eerste CWA componenten dus met andere woorden je hebt al een basis gevormd van een toekomstige verandering. Dus HRL 1 gaat al over de toekomstige verandering en dan uhh dus met andere woorden, nog even concreet maken, zou ik denk ik een HRL 0 toevoegen waar ik dan een soort van basic HF research erin zou zetten, of iets met fundamental research waarin je dan naar de huidige situatie kijkt.

Interviewee 2: En waarom zou je dat doen dan?

Interviewee 1: Omdat je als je het hebt over toekomstige verandering heb je altijd een bepaalde nulmeting nodig. En om te beginnen van we gaan iets veranderen zonder bijvoorbeeld te onderzoeken van hoe gebeurt het nu, lijkt mij ... ja is een mogelijkheid maar je wilt ook natuurlijk weten hoe gaat het nu, wat werkt wel, wat werkt niet, en van daaruit gaan naar die uhh die verandering.

Interviewer: En hoe onderscheidt dat zich dan van een work domain analysis? Waarin je dus eigenlijk kijkt naar wat is het werk wat gedaan moet worden, wat kenmerkt dit werkdomein, los van of het nu het werkdomein is of het toekomstige wekdomein. Of interpreteer ik dan de WDA anders?

Interviewee 1: Nou ik denk meer dat het is van je kan die WDA natuurlijk ook nu doen maar mijn punt is met name uhh volgens mij moet je onderscheid maken of je de uhh ik denk met name als je kijkt naar de toekomstige stappen, stappen in HRL 2 en 3, daar heb je ook andere cognitive work analysis componenten staan. Dus stel dat je CWA doet en je wilt alle drie doen in de huidige situatie dan zit je eigenlijk niet in HRL 1, 2 of 3 want je heb het over de huidige situatie.

Interviewer: Oké dus je zegt óf je moet al die methodes toepassen op de toekomstige situatie óf allemaal op de huidige situatie.

Interviewee 1: Ik zou zeggen van wel omdat het idee van die TRL dat je iets nieuws ontwikkelt. Dus die fases zijn, die gaan naar een bepaald prototype, die gaan naar iets nieuws, niet de huidige situatie.

Interviewer: Ja, oké. En als je het dan hebt over een HRL 0 waarbij je gaat kijken naar de huidige situatie, waar kijk je dan precies naar? Wat voor methodes gebruik je dan om daar naar te kijken?

Interviewee 1: Nou ik kan mij ook voorstellen dat je ook bijvoorbeeld een CWA doet, van de huidige situatie en dat je als volgende stap een CWA maakt van de toekomstige situatie. En even kijken, wij hebben dat gedaan bijvoorbeeld voor een taak-analyses van de huidige situatie en wat we ook hebben gedaan voor ERTMS uhh nou het is geen cognitieve work analysis maar een hierarchical task analysis, die hebben we dus ook gemapt voor wat dat dan doet met ERTMS. Dus stel dat je langs al die doelen en activiteiten gaat waar zit het verschil in tussen de huidige situatie en in de situatie met een ERTMS beveiliging.

Interviewer: Ja precies. En hoe specifiek vergelijk je dat dan al? Zijn dat al simulaties of is het meer een gedachte experiment?

Interviewee 1: Uhh ja dat laatste, dus het is een desktop analyse waarin je wel een aantal scenario's bedenkt en dan doorloop je dat met een taak-analyse, met de HTA.

Interviewee 2: Ja want meestal, onze projecten zijn meestal zo dat we niet iets heel nieuws maken maar in een bepaald werkdomein introduceren we een nieuwe techniek.

Interviewer: Ja dat zijn hele goede punten. Ik heb ook wat andere projecten geanalyseerd en daar zag ik inderdaad ook dat het vaak om een verandering gaat en dat ze het altijd hebben over een reference scenario waar je het dan tegenover zet.

Interviewee 2: Dat is eigenlijk wat jij ook zegt, hè [Interviewee 1]?

Interviewee 1: Ja.

Interviewer: Oké. En als we dat ERTMS systeem nemen, wat zouden dan, gebaseerd op de HRL, de volgende stappen zijn die je zou nemen, de volgende activiteiten die je zou plannen?

Interviewee 1: Uhh nou ja wat we nu gedaan hebben, vanuit ProRail, is uhh er liggen nu ook principes rondom hoe er samengewerkt moet worden dus wat is er vanuit beveiliging nodig, ook naar operators toe, maar ook naar binnen het systeem toe, dat noemen we gebruikersprocessen en er is een eerste user interface bedacht om deze gebruikersprocessen toe te kunnen passen maar ook om toch wel bepaalde of nieuwe functionaliteiten aan te reiken aan operators of een betere situational awareness te bereiken voor treinverkeersleiders. Dus ik denk wat we hebben gedaan dus die tweede stap is uhh zou je user interface principes uhh zijn er dus bedacht en new way of working. Ik zit even te kijken in de kolommen of dat ook zo genoemd wordt. Ja het zijn niet, ja het zijn meer dan principes eigenlijk. Ik zit te kijken naar die kolommen 'human-system performance design principles researched' uhh ja in principe...

Interviewee 2: Zit het niet al wat hoger?

Interviewee 1: Nee ik zit te kijken volgens mij komen we nog niet tot en met 3 met het project want bij 4 zit je al op validated in lab en volgens mij zitten we dus bij 3 nog... oké dus de volgende stap is geweest, nou er is dus dat idee bedacht van zo zou het nieuwe user interface kunnen uit komen te zien dus bijvoorbeeld met ERTMS zal de trein als het ware slimmer worden dus je zou als treinverkeersleider zou je meer informatie over de trein kunnen verkrijgen, of dat zou mooi zijn want dan krijg je die shared situation awareness als het ware tussen de machinist en treinverkeersleider. Dat is bijvoorbeeld een van de voorbeelden van een vernieuwde user interface. Ze kunnen zien bijvoorbeeld wat de status is van de trein, in welke modus - zo noemen ze dat dan - staat die in. En dat hebben we dan nog een keer getest in een simulatieomgeving wat dan ja uhh op bepaalde aspecten wel representatief is aan de huidige omgeving maar niet volledige overeenkomt of toch wat abstracter is in vergelijking met bijvoorbeeld een opleidingssimulatie. En dat is denk ik die level 3. Dus daar heb je het al over 'preliminary design features to accommodate human capabilities, limitations and needs investigated' uhh en in principe, even kijken hoor, misschien heb je het, misschien zit er eigenlijk ook wel een overlap in met, toch wel met level 4, als ik zo kijk.

Interviewee 2: Ja want wordt er, bij level 3 staat niet hoe je dat dan onderzoekt? Of je met... prototypes kwamen wat hoger voor ja...

Interviewee 1: Ja precies. En als je dan kijkt naar die 4e, dan zie je wel dat daar al die concepten in zitten waarmee je het eigenlijk test.

Interviewer: Ja want ik hoor je al simulatie zeggen? Dat is in principe iets wat in HRL 3 nog niet voorkomt.

Interviewee 1: Nee.

Interviewer: Dus klopt het dat jullie die simulatie misschien ook al gebruiken om die eerste system requirements op te stellen en dat jullie dan weer een simulatie gebruiken om te testen of dat ook echt werkt of?

Interviewee 1: Ja ik denk met name die system requirements zijn in principe zelfs al meegenomen maar ik denk, waar ik ook altijd een beetje mee zit, je kan het hebben over de validiteit van je simulatieomgeving. Dus bijvoorbeeld omdat er al staat 'validated in lab' dan uhh ik vind validate altijd een groot woord al en in principe de fase waar we in zitten is ehh eerst gewoon verkenning waarin je, naar mijn idee, met name nog daarom in level 3 zit, het is een eerste keer dat we daarmee beproeft hebben, wel met onderdelen van level 4 al maar... ja, oké je kan ook beargumenteren het is een eerste simulatie of een eerste onderzoek als onderdeel van niveau 4 en daarin zullen nog meerdere iteraties volgen.

Interviewer: Ja.

Interviewee 1: Ja oké. Dus als je het zo ziet kan je zeggen uiteindelijk zullen daar veel iteraties in plaatsvinden, ik denk in 4. Wat ik ook hier zie is dat de stappen best wel groot zijn eigenlijk. Met name level 4 zal echt een langer traject zijn. Het zal een traject zijn waar je echt meerder concepten gaat beproeven waardoor uhh ja ik denk dat je uhh in deze uhh hier zijn alle vakjes even groot en dat is natuurlijk ook geen issue. Maar als je bijvoorbeeld kijkt over tijd en hoe een project loopt, zou je 4 echt heel erg willen stretchen.

Interviewer: Oké. Ja het is natuurlijk wel heel belangrijk dat het visuele aansluit op de inhoud. Dus ik denk dat het op zich wel een goed punt is.

Interviewee 2: Ja maar misschien is dat wel anders voor...

Interviewee 1: Ja ik kan mij voorstellen, dat hangt een beetje van je doeleinde af, ik kan mij voorstellen als je zo'n overzicht hebt als deze, hoef je het denk ik niet te visualiseren maar als je het hebt over bijvoorbeeld de systeem verandering op zich of je hebt het over systems engineering ofzo, dat je het dan wel zo visualiseert, of project management of zo, dat je visualiseert hoeveel tijd waar in gaat zitten.

Interviewer: Ja precies. En in principe staat er dus bij HRL 5 'evaluatie van prototype in gesimuleerde omgeving'. Kun je dan uitleggen waarom je het dan niet daarin plaatst? Aangezien jullie wel een simulatie hebben uitgevoerd? Waar twijfel je dan bij?

K: Dat is ook wel iets wat gewoon binnen het bedrijf speelt. Wij hebben verschillende soorten simulatie systemen en dat gaat eigenlijk over de representativiteit van de simulatie waar je in simuleert, dus bij 5 zou je bijvoorbeeld uhh even kijken wat daar staat uhh zou je in principe een omgeving willen hebben waarin bijvoorbeeld het systeem wel iets representatiever is voor de operator ook met name omdat je daarin neuropsychological measures erin hebt bijvoorbeeld of ik kan me voorstelen dat zoiets als werkbelasting dat dat ook al wat meer belangrijker hierin wordt en hier staat 'potential human-system peformance issues and risks identified' ja daar wil je eigenlijk wel weer een simulatieomgevingen dan hebben waarin het dichter komt bij het echte systeem en waarin ook bijvoorbeeld bepaalde onderdelen van je systeem, en dan heb je het in dit geval bijvoorbeeld over het beveiligingssysteem wat eronder zit, of de functionaliteit, dat die wat beter gerepresenteerd zijn in die simulatie omgeving. Want je kan ook zeggen, kijk in 4 zouden het ook bij wijze van spreken nog bijna wat geavanceerdere mock-ups kunnen zijn. En ik denk ook dus dat daarmee dus de ja, de representativiteit waarin je, of eigenlijk heeft het ook wel te maken met die TRL niveaus, hoe volwassen is die TRL waarin je dat op dit moment beproeft. En daar zie je ook dat uhh het is nog echt in die ontwikkeling waardoor je uh ja dat ook weer moet matchen met je simulatieomgeving. Ja ik denk dat het een beetje lastig is hoe ik dit uitleg.

Interviewer: Nee ik vond het al heel knap dat je binnen 10 minuten eigenlijk al wel had gezien waar het systeem ongeveer zit want bij de vorige interviews duurde dat al 45 minuten. Het is alleen maar goed om zo hardop na te denken. En ik denk dat het ook best lastig is. Ik merk dus dat je eigenlijk al vrij snel een soort van feedback aan het zoeken bent, een soort van simulaties die je uitvoert, misschien nog niet heel volwassen simulaties maar wel al een soort van simulatie om te kijken van zijn m'n requirements op orde...

Interviewee 1: Ja.

Interviewer: Maar je zegt tegelijkertijd ik zit nog niet in HRL 5, het is niet dat we echt aan het valideren zijn of onze eerste concept uhh of de werkbelasting oké is etc.

Interviewee 1: Nee.

Interviewee 2: Dat staat er ook nog niet hè of wel? Ik probeer te mappen uhh die TRL 5, dan wordt het gevalideerd in een relevante omgeving. Ik weet niet precies wat dat is. Het is in ieder geval niet de operational environment, maar wel relevant.

Interviewer: In de originele HRL stond mission-relevent maar dat komt waarschijnlijk omdat het bedacht is door een werkgroep waarin mensen van kerncentrales en van defensie zitten. Dus ik denk dat dat beïnvloed is daardoor.

Interviewee 2: Maar het is nog wel gesimuleerd, dus nog niet echt... dus wel een test-omgeving of een uhh

Interviewer: Ja. Dus wat ik hierbij voor me zie, is operationeel is echt dat je het toepast in het echt en relevant, volgens mij is dat gewoon een gesimuleerde omgeving maar die dus wel specifiek is, waar bepaalde taken ofzo in kunnen worden uitgevoerd. Misschien is het ook wel een beetje een overbodig woord?

Interviewee 1: Nou ik kan me er wel wat bij voor stellen. Ik bedoel, wij denken vaak nu aan 3 type simulaties en ik denk dat er ook wel meer middelen zijn waar we aan moeten denken maar de laatste stap is eigenlijk een trainingsomgeving, een trainingssimulatie en daar zit je eigenlijk bijna met 1-op-1 systemen kan je zeggen, dat is vanaf 7 denk ik zo'n beetje. Daar kun je die voor toepassen. En als je dan terugkijkt, dan heb je voordat je zo'n systeem, ja eigenlijk je echte operationele systemen bouwt, wil je een eerste, wil je daar ook een soort van proof of concept van hebben en dat zit dan met name in 6 uhh omdat je uhh want dat is de fase die je moet doorlopen voordat je naar 7 gaat, voordat je het echt gaat ontwikkelen, in een operationele ontwikkeling uhh dus dan heb je dat nog eerst in 6 en dan heb je daarvoor dus nog een omgeving die je gebruikt om snel te prototypen en om requirements te formuleren die weer feedback geven op level 6. En dat is denk ik de omgeving waar we nu in zitten. Dat is een simulatieomgeving waar je gewoon snel in kunt wijzigen. Dus je krijgt er wel een wat beter gevoel of begrip bij als operator maar het zijn geen, als het ware, cardboard mock-ups meer.

Interviewer: Meer Wizard of Oz?

Interviewee 1: Ja eigenlijk ook niet. Het is in dit geval uhh ja het is een omgeving denk ik met name waarin je uhh wat wel lijkt op die uiteindelijke operationele omgeving maar verschilt in dat het qua representativiteit toch net een tikje abstracter is of minder vergelijkbaar.

Interviewer: Oké.

Interviewee 2: Maar zou je dan zeggen is dat een lab omgeving of een relevant environment?

Interviewee 1: Ja... ze schrijven het steeds een beetje anders op hè.

Interviewee 2: Ja...

Interviewee 1: Ja, daar kun je denk ik, ik denk dat 4 en 5 eigenlijk wel een beetje bij elkaar misschien wel horen.

Interviewer: Oké.

Interviewee 1: Omdat ik ook, wat ik al zei, ik denk dat 4 een heel lang traject is.

Interviewer: En komt dat dan omdat daar design concepts completed staat of wat maakt dan dat je het gevoel hebt dat je niet eerst 4 af kan sluiten en dan door naar 5?

Interviewee 1: Ik denk dat ik ook wel geleid wordt door wat er staat onder TRL. Omdat er staat 'validated in relevant environment' en bij de ander staat 'validated in lab'. Dus in principe zou ik ook zeggen, ik zou onder level 5 zelfs ook al, nog steeds zeggen dat je daar bezig bent met design concepts.

Interviewer: Uhu. Oké. En ook echt nog met de design concepts of met het design.

Interviewee 1: Met het design, ja. Maar dat is, ik zou zeggen het design en het design concept is... wat is het verschil ertussen?

Interviewer: Ik zou zeggen dat er misschien nog een onderscheid is omdat design concept, dat is iets wat misschien, dat is een idee, terwijl het design misschien al wat meer tastbaar is. Maar dat is mijn persoonlijk interpretatie, misschien is dat niet zo, misschien is een design concept hetzelfde als een design...

Interviewee 1: Nou ja, ik denk, ik interpreteer hem als nog een idee in principe. En ja ik interpreteer het als hetzelfde eigenlijk. Je hebt wel een design idee maar uiteindelijk blijft het een design idee totdat je het gaat implementeren of totdat je gaat kiezen dat je het gaat implementeren. En de stap om bijvoorbeeld te zeggen nou ik ga dit implementeren is wanneer je verschillende keuzes uitsluit uhh en ik denk dat je die keuzes pas uitsluit in uhh ja misschien zelfs level 6. Want je beproeft nog steeds wel in verschillende omgevingen, waar je er nog steeds achter kan komen van hé het moet nog aangepast worden, of het wordt op zijn minst nog wel getweaked als het goed is.

Interviewee 1: En dan gaat het over naar de implementatie, en dan heb je echt je design en dan ga je implementeren. Maar oké, dat is misschien ook wel een interpretatie of een smaakkwestie in hoe je het ziet.

Interviewer: Ja maar ik ben daar wel heel erg geïnteresseerd in, in de interpretatie en de smaakkwestie. Want wellicht dat dit nog iets te abstract is waardoor er te veel interpretaties kunnen ontstaan. Misschien is dat goed, maar misschien is dat ook verwarrend want het is uiteindelijk ook iets wat je wilt gebruikten om iets te communiceren, van 'oké dit is het doel' en 'dit systeem is in HRL 5' dat moet natuurlijk ook iets nuttigs zeggen. Dus als er te veel interpretaties mogelijk zijn, dan wordt het wellicht minder nuttig. Dus dat is zeker goede informatie.

Interviewee 1: Nou ja, is ook wel grappig als je kijkt naar, in principe, als je de kolommen met elkaar vergelijkt dan zou je ook aan de hand daarvan een bepaald level kunnen interpreteren en waar ik bijvoorbeeld nu uhh ik zag eigenlijk dat in level 6, dan staat er real-time simulations, dan heb je bijvoorbeeld ook de situational awareness vragenlijsten terwijl onderin zag ik bijvoorbeeld ook naast de TLX staan, op level 4, oké dat zijn eigenlijk de vragenlijsten als het ware uhh maar ik interpreteerde het in de zin van, omdat het naast de tlx is, dan denk ik aan, het is uhh sowieso maak ik eigenlijk niet echt een onderscheid tussen wanneer je op welk niveau welke technieken moet gebruiken dus ik denk dat je in principe ook SAGAT kan gebruiken op level 4 en 5. Ik denk dat de technieken uhh ja dat je die eigenlijk ook in verschillende fases wilt gebruiken, met name om... ik moet zeggen ik ken ze ook niet allemaal...

Interviewer: Ja het is ook luchtvaart-georiënteerd dus ik kan mij voorstellen dat jullie wellicht andere methodes gebruiken.

Interviewee 1: Nou in principe, nou bijvoorbeeld, ik weet niet wat QUIS is ofzo. Er zitten ook veiligheidsgerelateerde metingen in. Die gebruiken we niet direct. Bijvoorbeeld de HAZOP ofzo.

Interviewer: Worden die niet gebruikt in het algemeen of bedoel je als HF specialist gebruik je die niet, maar er is dan weer een veiligheidsspecialist die het wel gebruikt?

Interviewee 1: Ja precies. Dat laatste inderdaad.

Interviewer: Oké.

Interviewee 1: En als je zegt, los van je meetmethodes, die kun je vanaf 4 tot 6 wel een beetje gebruiken, want in principe kan je altijd zeggen, de meting die je dan verkrijgt zegt wat over je werkbelasting op een level 4 maar het geeft ook aan hoe dat uhh het is niet misschien echt een representatieve werkbelasting in de zin van stel dat je uiteindelijk dit echt implementeert en dat je in een echte operationele omgeving dus wat zou dan je werkbelasting zijn maar het geeft wel wat aan over uhh een inschatting die je hebt. En ik denk dat je, als je het hebt over level 4 tot 6, dan kun je altijd wel een inschatting krijgen van die werkbelasting of die Situational Awareness, alleen het zit op andere abstractieniveaus. Dus hoe meer je naar boven gaat, hoe accurater het gaat worden naar een vergelijkbare meting van werkbelasting of situational awareness in vergelijking tot in het echt.

Interviewer: Ja. Nou fijn dat je dat er in ieder geval uit haalt. Want dat is inderdaad een van de basics die er aan ten grondslag ligt. Fijn dat dat wel echt duidelijk naar voren komt. En dan nog even uhh want je had het al even over die tools... heeft zo een toolbox uhh zou je dat kunnen gebruiken? In jullie werk? Zo van oké we zitten nu ongeveer op prototype niveau dus dit zijn de methodes die we zouden kunnen gebruiken.

Interviewee 1: Uhh ja. Ik denk sowieso handig om dat overzicht te hebben van verschillende meetmethodes. Maar zoals ik net al zei, ik zou eigenlijk meer die tooling over de hele spectrum van 4 tot 6 toepassen.

Interviewer: Oké. Dus eigenlijk tussen validatie in het lab en een wat geavanceerdere situatie zit qua methodes niet per se een verschil?

Interviewee 1: Uhh nee.

Interviewer: Oké.

Interviewee 1: Want in principe wat we nu ook al hebben gedaan. We hebben toevallig vorige week al een onderzoek gedaan. We zijn vorige week bezig geweest op een verkeersleidingspost. Daar hebben we het eerste user interface getoetst met treinleiders. En daar hebben we uhh nou dat is een eerste keer in principe, dat het user interface concept is getoetst en dat hebben we ook gedaan nog met een eye-tracker. Dus ik heb eigenlijk een afstudeerder die ook op het vraagstuk zit van wat is uhh het meten van bijvoorbeeld situational awareness tijdens een uhh bij het maken van een nieuwe user interface. En ik kan ook zeggen dat we eigenlijk nog in verschillende fases dus die user interface zich nog moet ontwikkelen. Dus we hebben zelfs dus al in een eerste fase hebben we ook een eye-tracker toegepast om te kijken ja wat doet dat nou met situational awareness maar ook met usability, zo ver als mogelijk ook met werkbelasting. Dus dan kan je ook zeggen, dus wat je normaal denk ik hier ook ziet, als je real-time simulations hebt uhh ja dus als je kijkt naar onze casus, we hebben een eerste keer gemeten, ook met een real-time simulation maar uhh het is een simulatie omgeving die wat abstracter is in de zin van uhh ja de knoppen voelen nog wat anders aan, het systeem reageert wat anders maar je kunt wel als treinleider treinen uhh handelen zoals je wilt en je hebt een user interface wat dynamisch is. Het is dus wel een eerste keer dat we het hebben gedaan en we moeten daar nog feedback op krijgen dus gezien die simulatie zou ik zeggen volgens deze categorie zit je dan al in 6 en uhh

Interviewee 2: Waarom 6? Ik zou zeggen 5...

Interviewee 1: Uh ja nou omdat er real-time simulation staat.

Interviewee 2: Oja, ik kijk naar 'evaluatie van een prototype'

Interviewee 1: Ja.

Interviewer: Dus die matchen niet helemaal?

E & Interviewee 1: Nee.

Interviewee 2: Ik dacht, ik kijk wel meer naar het doel, naar de evaluatie van het prototype, en jij kijkt naar het middel 'real-time simulatie'.

Interviewee 1: Nee, waar ik naar kijk is eigenlijk dus wat we hebben gedaan en per kolom kijken in hoeverre het overeenkomt met wat we hebben gedaan. Dat probeer ik nu te doen eigenlijk.

Interviewer: Dus je zou dan zeggen het zit dan in 6 maar het voelt nog niet helemaal als een HRL 6?

Interviewee 1: Klopt. Ja. En ook met name omdat als we het in 6 gedaan zouden moeten hebben dan zou ik een andere, dan zie ik een andere simulatieomgeving voor mij, om dat in 6 te doen.

Interviewer: Oké. En waar zou je...

Interviewee 2: Als dat al gebeurt hè, dat weet je niet of je dat altijd doet. Volgens mij wordt onze discussie moeilijker door de verschillende simulatoren inderdaad.

Interviewee 1: Ja dat koppel ik dus aan die TRL niveaus dus de simulatieomgevingen.

Interviewee 2: Wat op zich uhh wat ik begrijp is dat je na HRL 6, dan heb je uhh dan weet je nou dit gaat werken op technisch gebied en op human gebied en dan ga je het voor het echt maken en dan kan je je prototype weggooien eventueel en dan ga je naar 7 toe. Dus uhh...

Interviewee 1: Ja.

Interviewer: Dus eigenlijk zou je zeggen dat simulatie al eerder moet voorkomen in de HRL? Maar dan op een hoger abstractie niveau?

Interviewee 2: Dat zou kunnen ja.

Interviewee 1: Nou ja, in principe zou het misschien ook, ik weet niet wat je precies begrijpt onder gaming maar in principe is uhh wij gebruiken binnen ProRail ook wel games in de zin van uhh eigenlijk wat we nu ook doen, we noemen een simulatie ook een game eigenlijk. En ja in principe met gaming heb je ook verschillende soorten games. Mijn proefschrift is daar over gegaan. Dus dat je ook analoge hebt maar ook digitale. Dus dan zat je uhh ja uhh ik denk dat we sowieso een beetje zitten van, wat je net ook al zei [Interviewee 2], volgens mij zitten we met name over het stukje onder 6, boven 6 vind ik hem best wel straight-forward in de zin van dat je dan systeem gaat bouwen en nouja dat je ook al alle meetmethodes gaat gebruiken die er onder staan. En ook de doelen die erin staan dus bijvoorbeeld dat je naar human system performance gaat kijken, die zijn allemaal duidelijk. Ik denk met name waar ik nog mee zit, is onder de 7.

Interviewee 2: Ja, hoe je dit doet en...

Interviewee 1: Ja.

Interviewee 2: Kan ook best dat het... ja dan misschien niet een beetje iteratief is inderdaad als je uiteindelijk maar je design fully matured hebt.

Interviewer: Ja het is een eindstatus dus het zegt inderdaad niet zoveel over wat er allemaal moet gebeuren om die eindstatus te bereiken, afgezien van de methodes die er aan de rechterkant staan.

Interviewee 2: Het is wel inderdaad vrij veel wat er staat in die kolommen hè. De vraag is of we dat altijd overal doen.

Interviewer: Ja dat is inderdaad ook wat ik tot nu toe al heb gezien, dat het wel een ideaal plaatje is wat waarschijnlijk over het algemeen, met tijdsdruk ook en commerciële druk niet wordt gevold. Want zou je bijvoorbeeld zeggen dat er bij jullie ook in al die ontwikkelniveaus naar training wordt gekeken, naar manpower & personnel, ...

Interviewee 1: Dat wou ik net aankaarten inderdaad. Dus nou... minder.

Interviewer: Welke categorie specifiek?

Interviewee 1: Training met name. Dus er is wel gewoon bewustzijn dat er moet worden gekeken naar training en dat daar op voorbereid moet worden maar voor de rest is dat eigenlijk niet echt uhh komt dat niet echt ter sprake tijdens de ontwikkeling. Het is wel een soort van ja dat moeten we doen en dat staat op de kaart dat dat ontwikkeld moet worden maar er wordt verder nog niks gedaan met training in die fases, onder 7.

Interviewee 2: Nee op het moment dat je echt zeker weet dat je het gaat doen, dan wordt er vaak gekeken hoe gaan we het implementeren hoe veel tijd is er nodig voor training, wie gaat het doen, enzovoorts.

Interviewer: En volstaat dat of zou het iets toevoegen om daar eerder mee te beginnen?

Interviewee 2: Nou vaak duurt het ontwikkelen van een systeem al vrij lang. Je moet eigenlijk uhh ja uhh weet ik eigenlijk niet. Ja het zijn best wel lange trajecten bij ons dus uhh ja...

Interviewee 1: Ja ik zou zeggen, ik zou hem persoonlijk niet gebruiken onder 7, ik denk dat het op zich, kijk de categorieën die er worden genoemd zijn natuurlijk wel goed als een leidraad om daar bewust van te zijn. Maar ik denk uhh wat je gaat krijgen dus training een beetje op het laatst, daardoor organisation and staffing denk ik ook, is ook wat minder belangrijk...

Interviewee 2: Ja behalve de staffing heeft ook te maken met hoeveel mensen heb je nodig hè.

Interviewee 1: Ja maar dat hangt er van af denk ik welke technologie je gebruikt. Bijvoorbeeld, als je het hebt over een ander project waar we mee bezig zijn, waar we kijken naar de rollen, ja daar is staffing wel belangrijk maar over ERTMS dan in principe...

Interviewee 2: Ja nee oké. Soms is het onderdeel van je ontwerp zeg maar, hoeveel mensen je nodig hebt, en wat voor soort mensen. Dan wordt ook die organisation & staffing en roles & responsibility weer erg belangrijk.

Interviewer: Oké. Dus het is project-afhankelijk?

Interviewee 1 & 2: Ja.

Interviewee 1: Ja en in principe, de vraag wordt er wel altijd gesteld rondom van wat voor competenties heb je dan nodig of wat voor impact heeft dat. Daar is het project dan zich wel bewust van maar er wordt nog niks gedaan in eerdere fases.

Interviewer: Oké. Training vult meer het gat achteraf?

Interviewee 1: Ja. Maar ook wat het doet met je staffing, en met de competenties die je hebt dat uhh want ook met staffing denk ik van ja moet je toch bijvoorbeeld wat andere uhh enerzijds is het trainen natuurlijk, tot mensen daar de bepaalde competenties krijgen maar staffing is dan bijvoorbeeld ook over uhh hier staat wel 'strategies' maar ik versta hier bijvoorbeeld ook onder dat nieuwe mensen die je aanneemt, dat je je staff eigenlijk als het ware inricht, dus dat je mensen op een andere manier aanneemt, op basis van andere criteria aanneemt.

Interviewer: En in het algemeen, want 'strategies' komt wel vaker voor in dit schema, wat verstaan jullie daaronder? Hoe interpreteren jullie dat?

Interviewee 2: Dus je bedoelt die 'effectiveness of strategies? Ja hier komt het ook terug...

Interviewer: Ja, strategies to address environmental constraints, strategies to accommodate manpower & personnel, strategies to... ja bij uhh

Interviewee 2: Ja meer hoe je dat gaat aanpakken. Het kan nog vrij abstract zijn denk ik.

Interviewer: Uhu. Ja?

Interviewee 2: Ja.

Interviewee 1: Ja.

Interviewer: Oké. En misschien kunnen we afronden met uhh ja zou je zoiets als de HRL toe kunnen passen? En op wat voor projecten zou je dat dan toe kunnen passen? Waar zou het nuttig voor kunnen zijn? En als jullie denken dat het nergens nuttig voor is, dan mogen jullie dat ook gerust zeggen want dat is ook nuttige feedback.

Interviewee 1: Nou ik denk sowieso dat het al heel erg goed is om hiervan bewust te zijn en je zou het, ik denk dat we het ook wel zouden willen toepassen uhh alleen ik denk wat we ook nog zouden doen, ja in principe op dit moment lopen er ook verschillende soorten processen als het ware, waar [Interviewee 2] ook in het begin over vertelde dat je binnen innovatie hebben we allemaal stappen hoe je innovatie doorloopt, in principe wordt er ook vaak gerefereerd naar zo'n V model vanuit systems engineering, nou en vanuit NASA is er ook weer een ander model ontwikkeld rondom conceptontwikkeling en eigenlijk is het een beetje de vraag van ja uhh ook bijvoorbeeld binnen ERTMS wordt er gebruik gemaakt van zo'n systems engineering V model en daar moeten we het ook op kunnen matchen. Dus je wilt ook eigenlijk dat dit model, en ik denk ook wel dat die op bepaalde, dat die ook wel goed te matchen zou kunnen zijn, maar die wil je dus ook kunnen matchen met andere modellen.

Interviewer: Ja precies.

Interviewee 1: En ik denk ook dat het heel waardevol kan zijn omdat je het eigenlijk in een zekere zin als een checklist kan zien.

Interviewee 2: Ja precies. Dat wilde ik ook zeggen. Ik zie het als een checklist, ook voor de project manager, van wat moet er nog allemaal gebeuren.

Interviewee 1: Ja.

Interviewer: Oké, mooi. Ja ik herken ook dat punt dat het duidelijk moet zijn hoe het linkt aan andere bestaande processen en modellen. Dat is denk ik een belangrijk punt. En hoe komt het dan dat er zoveel verschillende processen worden gebruikt en dat er niet 1 soort van ProRail proces is die al die ander uhh ja eigenlijk overbodig maakt? Heeft het misschien ook nut om allemaal verschillende processen en modellen te gebruiken of is het iets wat organisch is ontstaan?

Interviewee 2: Uhh het heeft niet altijd nut, dat is 1. Ik weet het niet, misschien heeft het soms wel nut, het is vaak organisch ontstaan en uhh en misschien dat het voor sommige dingen wel nut heeft om bepaalde modellen te gebruiken hoor, dat weet ik niet maar je kan je afvragen of iedereen daar over nadenkt in het begin, dat weet ik niet. Maar het is denk ik een feit dat dat wel gebeurt in grote organisaties. Denk je niet [Interviewee 1]?

Interviewee 1: Ja ik zit even na te denken. Sowieso zijn het complexe projecten. Je werkt ook vaak met verschillende leveranciers die brengen ook elke keer weer een eigen model mee. We zijn ook nu bijvoorbeeld met NLR bezig voor een project, nou die heeft eigenlijk ook weer een eigen model. Ik denk, en wat trouwens ook weer uit de luchtvaart komt, maar je ziet gewoon dat er toch best wel wat modellen, verschillende modellen bestaan. Je hebt verschillende leveranciers maar je hebt daarnaast ook bijvoorbeeld collega's die dan andere, verschillende ideeën met zich meebrengen. Ik denk, er is waarschijnlijk ook geen 1 ultieme model. Dat is ook wel het punt.

Interviewer: Ja dus ze zullen allemaal bepaalde zwakke en sterke punten hebben...

j: Ja.

Interviewer: Maar het is waarschijnlijk wel handig om een soort van communicatie tool te hebben om bijvoorbeeld met alle leveranciers te kunnen schakelen?

Interviewee 1: Ja.

Interviewee 2: Zou het nog leuk zijn, want je hebt nou zo een mooi uhh HRL overzicht om ook zo'n tooloverzicht nog even uhh te maken. Dat je een tool kan plaatsen want dan weet je ook meer waar het over gaat in zo'n uhh het is meer omdat ik niet alle tools inderdaad ken, of sterker nog heel veel tools niet ken, alleen als er gewoon Engelse woorden staan...

Interviewer: Ja het zijn heel veel afkortingen. Dus een tooloverzicht, niet alleen maar met afkortingen en namen maar met de beschrijving?

Interviewee 2: Ja ook al 1 regel hoor maar uhh 'om dit en dit in kaart te brengen' of 'om dat en dat te toetsen' maar misschien is het gesneden koek voor sommige mensen...

Interviewer: Oh nee, dat denk ik zeker niet. Ik denk ook uhh dit heeft altijd een handleiding nodig, zo'n tool als deze, een stukje met wat meer informatie. Dus ik denk dat het zeker een goede opmerking is. En is er dan verder nog iets, een classificatie, die zo'n toolbox nog handiger zou maken? Of alle tools op 1 hoop, eventueel per level of per 3 levels is voldoende?

Interviewee 1: Nou de tools die nu op uhh ik denk 1 tot 3 is inderdaad wel duidelijk, dat is dan beetje desktop analyse en gefocust op de CWA en bij 4 tot 6 zou ik eigenlijk die tools allemaal op 1 hoop gooien en daar een onderscheid maken van type tools, in de zin van zijn ze bijvoorbeeld psychofysiologisch, zijn het vragenlijsten...

Interviewee 2: Maar ook misschien, wat brengt het in kaart, of wat doet het. Want ik zie hier nog die' impacts on the human in the system', waar gaat het over?

Interviewee 1: Ja je zou hier zelfs nog elk bolletje kunnen toewijzen van hoe wordt dat gemeten?

Interviewer: Ja deze hoort eigenlijk bij die eerste kolom, human in the system. Omdat die human in the system categorie nogal veel omvat en ik vond dat zelf niet helemaal duidelijk wat dat dan allemaal kon zijn. En die wordweb geeft dan aan wat human in the system allemaal inhoudt, ook als een soort van geheugensteuntje.

Interviewee 2: Oké ja ik zie nu dat je dezelfde kleur hebt gebruikt.

Interviewer: Dus het zou ook interessant zijn om tools aan die verschillende HF impacts te koppelen?

Interviewee 2: Ja vind ik wel.

Interviewee 1: Ja. Ja want in principe noem je inderdaad wel de tools maar heel veel zijn toch wel bijvoorbeeld te linken aan 1 van deze punten.

Interviewer: Ja. Misschien een over-the-shoulder-observation dat kan van alles onderzoeken maar zo'n...

Interviewee 1: Nou misschien wel interessant, ik kan mij ook wel voorstellen dat je ook hier een onderscheid in kan maken, dat je zegt in level 1 tot 3 zit je heel erg op het stukje cognitieve control uhh dan stukje 4 tot 6 dan zit bijvoorbeeld zoiets als trust, acceptance nouja maar ook bijvoorbeeld heel veel van deze zitten erin maar ik denk specifiek voor 7 en hoger zit je echt op denk ik meer fatigue enzo, ik denk dat je fatigue minder goed kan onderzoeken in 4 tot 6, dat zou je veel meer bovenin willen onderzoeken.

Interviewer: Oké. Dus het is niet zo dat je ze allemaal onderzoekt op elk abstractieniveau?

Interviewee 1: Nee. Ik denk dat je wel veel onderzoekt in 4 tot 6 maar bij bijvoorbeeld 1 tot 3 zou ik met name gaan kijken naar uhh ja oké je zou gewoon desktop analyses kunnen doen, dat hebben we ook gedaan, wat het doet met de werkbelasting en situational awareness.

Interviewee 2: Ja maar dan meer kwalitatief sowieso.

Interviewee 1: Ja precies.

Interviewer: Dus 1 tot 3, waar zou de nadruk dan op liggen?

Interviewee 1: Uhh ik denk, ja tot nu toe, hebben we daar met name gekeken naar workload, situational awareness en SRKs. Dat zijn denk ik... en Error.

Interviewer: Oké.

Interviewee 1: Misschien dat je zelfs ook mental model erbij kan nemen... dat is ook nog wel vrij abstract, dat kan ook op dat niveau zitten. En ik denk als je dus uhh 4 tot 6 zou je eigenlijk al kunnen zeggen, zou je alles kunnen doen, behalve ik denk dus fatigue is dus minder een ding en vigilance ook. Ja en dan vanaf 7 en hoger, eigenlijk alles, want dan zit je weer eigenlijk dat je een systeem gaat implementeren en dat het je bijna 1-op-1 systeem wordt, of het wordt je echte systeem. En dan kan je die performance gaan onderzoeken, dus weer wat voor impact heeft dat nieuwe systeem op de mens.

Interviewer: Ja. Dat gaat waarschijnlijk door, ook als het op een gegeven moment geïmplementeerd is blijf je dat eigenlijk analyseren...

Interviewee 1: Ja. Dan kom je eigenlijk weer bij 0 uit als het ware.

Interviewer: Ja precies. En dan is het cirkeltje rond.

Interviewee 1: Ja.

Interviewer: Oké. Hebben jullie verder nog opmerkingen, brandende vragen of dingen die we nog niet hebben besproken?

Interviewee 1 & 2: Nee hoor.

Acknowledgments

ACROSS is a European 7th Framework European Community co-funded project.

ALICIA has received funding from the European Community's 7th Framework Programme under grant agreement no ACP8-GA-2009-233682.

NINA was created within the frame of the SESAR Programme co-financed by the European Union and EUROCONTROL under grand agreement.

SAFELAND has received funding from the SESAR Joint Undertaking under grant agreement No 890599 under European Union’s Horizon 2020 research and innovation programme.

STRESS has received funding from the SESAR Joint Undertaking under grant agreement No 699381 under European Union’s Horizon 2020 research and innovation programme.