SAND2020-5713C

UNDERSTANDING HUMAN READINESS LEVELS

George Salazar (NASA), Judi E. See (Sandia National Laboratories), Holly A. H. Handley (Old Dominion University), Richard Craft (Sandia National Laboratories)

The Human Readiness Levels (HRL) scale is a simple nine-level scale developed as an adjunct to complement and supplement the existing Technology Readiness Levels (TRL) scale widely used across government agencies and industry. A multi-agency working group consisting of nearly 30 members representing the broader human systems integration (HSI) community throughout the Department of Defense (DOD), Department of Energy (DOE), other federal agencies, industry, and academia was established in August 2019. The working group's charter was to mature the HRL scale and evaluate its utility, reliability, and validity for implementation in the systems acquisition lifecycle. Toward that end, the working group examined applicability of the HRL scale for a range of scenarios. This panel will discuss outcomes from the working group's activities regarding HRL scale structure and usage.

INTRODUCTION

The concept of Technology Readiness Levels (TRL) can be traced back as early as 1969 when NASA's Apollo program was at its peak. Readiness levels were formally defined and described in a paper published in 1989 (Sadin, Povinelli, & Rosen, 1989). Over time, the nine-level TRL scale that is widely used throughout government and industry emerged. The TRL scale indicates the maturity of a particular technology, ranging from Level 1 (basic principles observed and reported) to Level 9 (actual system proven through successful mission operations).

The TRL scale has many advantages. First, TRLs create a framework with a common terminology that facilitates the communication of technology status within and among diverse organizations (Mankins, 2009). Second, the definitions embedded in the TRL scale actively promote testing and verification to assess maturity. Third, TRLs can be used to gauge progress throughout development and plan anticipated levels of effort remaining to achieve a desired level of maturity. Fourth, TRLs can be used to manage program schedule and cost risks with a strategy designed to prevent integrating technologies into the system before they are technically mature. Ultimately, the TRL framework provides a common understanding of technology status that is used by management to make decisions regarding technology funding, development, and transition to major programs.

Human readiness levels. Despite the undeniable benefits of TRLs, they do have limitations (See, Craft, & Morris, 2019). One gap is the inability of the TRL scale to capture the human-related aspects of technology development and their critical role in the readiness of a technology for operational use (Endsley, 2014; Kosnik & Acosta, 2010; O'Neil, 2014; Phillips, 2010, Phillips, 2015). Accordingly, Human Readiness Levels (HRL) have been designed as an adjunct to the TRL scale in order to fully incorporate the human element of the system throughout the lifecycle. While the TRL scale provides assurance that the technological components of the system will function as intended, it does not address the interactions between the technologies and the humans in the system that are necessary for success. That is, a technology may be mature in a strictly technical sense; however, if it is not ready for people to use effectively, then its overall readiness for deployment could be much lower.

Purpose of the HRL scale. The goal of the HRL scale is to provide organization and program management with a simple number that indicates the state of integration within the system with respect to humans and technology (Table 1).

Table 1: Nine Levels of the HRL Scale

HRL	Description
1	Relevant human capabilities, limitations, and basic human performance issues and risks identified
2	Human-focused concept of operations defined and human performance design principles established
3	Analyses of human operational, environmental, functional, cognitive, and physical needs completed, based on proof of concept
4	Modeling, part-task testing, and trade studies of user interface design concepts completed
5	User evaluation of prototypes in mission-relevant simulations completed to inform design
6	Human-system interfaces fully matured as influenced by human performance analyses, metrics, prototyping, and high-fidelity simulations
7	Human-system interfaces fully tested and verified in operational environment with system hardware and software and representative users
8	Total human-system performance fully tested, validated, and approved in mission operations, using completed system hardware and software and representative users
9	System successfully used in operations across the operational envelope with systematic monitoring of human-system performance

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

The HRL scale is designed to support management decision making with respect to investments of time and resources, applications of human-focused technical standards, and testing considerations necessary to maximize readiness for human use. Potential benefits of the HRL scale include:

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

Structure of the HRL scale. The HRL scale has been structured to mirror approaches used in popular TRL calculator tools developed by the Air Force Research Laboratory and the Department of Energy (DOE). Namely, each TRL level is associated with a series of statements or questions that must be addressed in order to determine a technology's current maturity level. For example, TRL Level 1 in the DOE tool has nine questions; e.g., *Has a fundamental concept, innovation, or scientific principle/methodology or approach been developed?* The entire DOE TRL calculator tool has 125 questions spanning Level 1 through Level 9. The HRL scale was similarly constructed in accordance with the TRL approach. The entire HRL scale has 74 yes/no questions from Level 1 through Level 9. As an example, one of the questions in HRL Level 1 asks *Have potential key user characteristics been identified?* As in various TRL frameworks, the HRL scale also identifies the exit criteria and supporting evidence required to determine whether the questions at a given HRL level have been satisfactorily addressed in order to advance to the next level in the scale.

CONTRIBUTIONS

HRL Scale History and Utility

Judi E. See

History of the HRL concept. Figure 1 illustrates a timescale for the historical development of the HRL scale. The foundation for the HRL concept is the human views framework designed to supplement existing architectural frameworks commonly used to support defense acquisition efforts. A human view explicitly represents the human in the system and documents the unique implications humans bring to system design. At present, human views are not a required part of any accepted architectural framework. For that reason, other techniques to effectively incorporate the human dimension throughout design and development have been explored. Human readiness levels represent one such alternative to accomplish that objective.

The concept of an HRL scale was first proposed by Dr. Hector Acosta (2010) in a discussion panel at the Aerospace Medical Association annual meeting. As with the human views approach, a key driver was the need for early and frequent consideration of the human dimension throughout design and development. In the years since the 2010 meeting, considerable HRL research has occurred. The Naval Postgraduate School sponsored two master's theses focused on the HRL concept and human readiness assessment (O'Neil, 2014; Phillips, 2010). Phillips (2010) formally developed the first framework for a nine-level HRL scale and collected feedback from subject matter experts in HSI and defense acquisition identifying the value of the HRL concept to support HSI planning and program risk management. O'Neil's (2014) thesis developed the Comprehensive Human Integration Evaluation Framework (CHIEF) to assess HSI throughout design and development. CHIEF uses a five-level scale and system-specific criteria to assess the effect of HSI on total system performance.

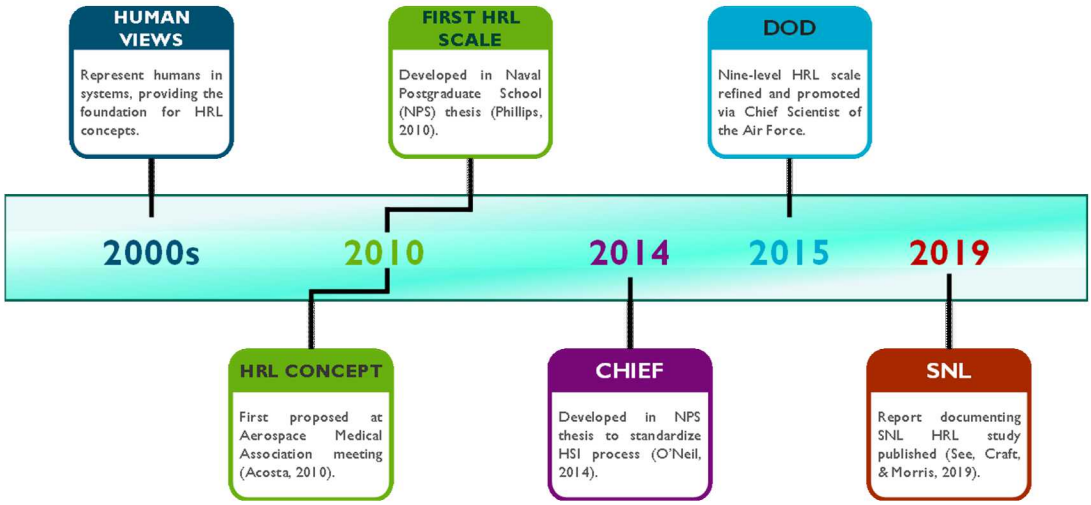

Figure 1. Timescale illustrating historical development of the HRL scale.

Additional progress was made to mature Phillips' (2010) original nine-level HRL scale during Dr. Mica Endsley's tenure (2013-2015) as Chief Scientist of the Air Force (Endsley, 2014). A Department of Defense (DOD) HSI working group was established in 2014 to refine the HRL scale with more detailed definitions and descriptions (Phillips, 2015). The working group included over 20 representatives from all four branches of the U.S. armed services, the U.S. Coast Guard, the Federal Aviation Administration, the National Aeronautics and Space Administration, and the Department of Homeland Security. The goal was to create a tool to convey HSI progress for program managers and identify the extent to which HSI requirements are incorporated (a.k.a. an HSI progress meter). Sandia National Laboratories leveraged the HSI working group's HRL descriptions to begin investigating HRLs in 2015, with the specific goal of identifying whether the concept has utility for its nuclear weapons mission work (See, Craft, & Morris, 2019).

From 2019 through 2020, Sandia National Laboratories, Old Dominion University, and the Naval Postgraduate School co-chaired another working group to complete the scale and test its utility. The working group consists of nearly 30 members from the broader HSI community in the DOD, DOE, other federal agencies, industry, and academia. This working group continued maturing the HRL scale levels and descriptions developed via previous efforts. The working group also supported evaluations of the utility, reliability, and validity of the HRL scale.

HRL scale utility assessments. In 2015, Sandia National Laboratories began investigating the utility of the HRL scale for its nuclear weapons mission work (See, Craft, & Morris, 2019). The scope of the study included an initial baseline assessment to understand how the laboratories conduct product development and the extent to which the human component of the system is incorporated. In addition, the study also gauged staff views of various options to support human readiness assessments, one of which was the HRL scale. Study results indicated that product design and development approaches at Sandia National Laboratories tend to neglect the human component of the system, with little to no consultation with human systems professionals for input. Feedback on the HRL scale was positive, though staff were concerned that it should be incorporated into existing frameworks to the extent possible so as not to burden already overloaded systems engineering processes.

The working group continued assessing utility in 2019 and 2020 by exploring whether the HRL scale is useful to detect HSI issues in a broad range of historical scenarios. Working group members selected scenarios with which they had first-hand knowledge of the evolution and outcome of a prominent HSI issue for a given technology. For each scenario, the HRL scale was applied to identify the first question in the scale that would have successfully detected the HSI issue. This analysis provided an indicator of the lifecycle stage at which HSI issues could be detected via application of the HRL scale. Other questions in the scale that would have also detected the issue were highlighted as well to provide an indicator of the number of opportunities to detect the issue. In this way, the working group was able to evaluate HRL scale sensitivity, diagnosticity, and completeness.

Inter-Rater Reliability and Usability

Holly A. H. Handley

The HRL working group conducted a small usability evaluation and an assessment of inter-rater reliability of the HRL scale. These evaluations were performed as part of the verification process for the HRL scale to ensure that it is usable for practitioners and can be applied to provide consistent results.

Usability. During a series of virtual workshops, approximately 15 working group participants were guided through an historical scenario that illustrated a particular HSI issue and shown how the HRL scale could be applied to identify the issue. Participants were shown how to navigate the series of yes/no questions associated with each level of the HRL scale in order to identify the first question that would have detected the HSI issue. In this case, the subject matter expert's analysis indicated the issue would have been identified via one of the questions in HRL Level 1. The working group then discussed the process and explored other questions in the scale that might also have detected the issue.

Afterwards, the working group used the HRL scale to address a second historical scenario involving ergonomics issues in a Coast Guard craft. For each issue, participants were asked to identify the first question in the HRL scale that would have detected the issue. Participants initially worked through the case independently and then collaborated in small groups to reach consensus. At the conclusion of this exercise, working group participants independently completed the System Usability Scale (SUS), a widely used survey instrument consisting of 10 questions regarding system usability (Brooke, 1986). Each question in the SUS is rated on a five-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The SUS questions were tailored for the specific purposes of the HRL scale usability assessment (Table 2).

Table 2: SUS Questions for HRL Scale

	System Usability Scale Question
1	I think that I would like to use the HRL scale frequently.
2	I found the HRL scale unnecessarily complex.
3	I thought the HRL scale was easy to use.
4	I think that I would need the support of a technical person to be able to use the HRL scale.
5	I found that the various functions in the HRL scale were well integrated.
6	I thought there was too much inconsistency in the HRL scale.
7	I would imagine that most people would learn to use the HRL scale very quickly.
8	I found the HRL scale cumbersome to use.
9	I felt very confident using the HRL scale.
10	I needed to learn a lot of things before I could get going with the HRL scale.

Inter-rater reliability. In addition to being usable, the HRL scale is intended to generate consistent results across diverse practitioners; i.e., multiple users evaluating the same issue should reach the same conclusion. Accordingly, the working group participated in a small test of inter-rater reliability, using a different historical scenario. Each working group member independently reviewed the scenario and applied the HRL scale in order to identify the first question that would have detected the HSI issue involved in the scenario. Results were then compared to the "ground truth" selections made by the subject matter expert for this scenario.

The fundamental research question for inter-rater reliability of the HRL scale is the accuracy with which each participant identified the first relevant HRL question relative to the "correct" level determined by the subject matter expert. For this analysis, the inter-rater agreement index r_{wg} was calculated to represent the extent of agreement between the expert and the working group participants (Demaree, James, & Wolf, 1993). This index ranges from 0 to 1, with a value of 0 signifying no agreement and a value of 1 indicating perfect agreement among the raters.

Olsen and Shorrock (2010) showed that inter-rater reliability can be a concern when users are unfamiliar with the tool or domain. In the current analysis, all working group members who submitted ratings were familiar with the HSI domains, but they had various levels of familiarity with the HRL scale. Hallgren (2012) confirmed that the more training inter-raters receive, the better their reliability, thereby producing more robust conclusions. Therefore, a goal of the HRL scale design is to ensure the scale is as self-explanatory as possible to support consistency of outcomes across users.

Validating the HRL Scale
Richard Craft

The working group is engaged in two major activities to validate the HRL scale. First, the working group has mapped the HRL scale to various existing HSI standards:

- NUREG-0711 *Human Factors Engineering Program Review Model*
- United Kingdom's Joint Service Publication (JSP) 912 *Human Factors Integration for Defence Systems*
- SAE6906 *Standard Practice for Human Systems Integration*
- ISO 9241 220 *Ergonomics of Human-System Interaction – Part 220: Processes for Enabling, Executing and Assessing Human-Centred Design within Organizations*

These mappings have demonstrated a sound technical basis validating HRL scale contents. For example, JSP 912 indicates that a human factors lead must be identified. Similarly, the HRL scale indicates that human systems professionals with requisite expertise must be engaged and funded to support the lifecycle of the effort. JSP 912 specifies that user roles in the lifecycle must be identified, and the HRL scale specifies that basic task descriptions for user roles must be developed. Finally, JSP 912 requires assessment of risks, which may occur through analysis of legacy and comparable systems. Likewise, the HRL scale indicates that potential key human systems issues throughout the lifecycle must be identified, and human performance on legacy or comparable systems must be analyzed to understand key human-technology interactions.

Second, the working group is currently evaluating multiple HSI tools to further explore consistencies, gaps, and linkages (Table 3). This analysis provides an additional validation of the HRL tool by demonstrating that it includes the same types of elements found in comparable HSI tools. In addition, it highlights the similarities and differences among the tools to support evaluation of their unique contributions to the HSI assessment process. The working group is currently evaluating the possibility of a comprehensive HSI framework that incorporates the unique elements from each tool in order to arrive at the top-level HRL rating to support organization and program management decision making.

Table 3: Working Group Evaluation of HSI Tools

HSI Tool	Key Source
HSI Framework	Lacson, Risser, Gwynne, & Kosnik (2017)
HSI Risk Management Tool	Kosnik, O'Neill, and Zimmerlin (2018)
HSI Progress & Risk Specification Tool	Phillips, Stohr, Pharmer, and Seely (2015)
CHIEF	O'Neil (2014)
Multivariate Human Readiness Assessment	Garcia, Ganey, and Wilbert (2017)
Human Factors Readiness Levels (HFRL)	Del Giudice, Hale, and Johnston, (2015)
Human Capability Levels	Miller, Thomas, and Rusnock (2016)

ACKNOWLEDGMENTS

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

REFERENCES

Acosta, H. (2010, May). *Human readiness levels: Implementing HSI – Connecting some dots.* Panel discussion presented at the 81st Annual Scientific Meeting of the Aerospace Medical Association, Phoenix, Arizona.

Brooke, J. (1986). *System usability scale (SUS): A quick-and-dirty method of system evaluation user information.* Reading, UK: Digital Equipment Co Ltd.

Del Giudice, K., Hale, K., & Johnston, M. (2015). Development of a scale for quantifying a system's human factors readiness. *Procedia Manufacturing, 3,* 5192-5198.

Demaree, R. G., James, L. R., & Wolf, G. (1993). R_{wg}: An assessment of within-group interrater agreement. *Journal of Applied Psychology, 78,* 306-309.

Endsley, M. (2014). *Human system integration: Challenges and opportunities.* PowerPoint slides retrieved from http://www.acq.osd.mil

Garcia, A., Ganey, N., & Wilbert, J. (2017). Human readiness assessment: A multivariate approach. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61,* 106-109.

Hallgren, K. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. Tutorials in *Quantitative Methods for Psychology, 8,* 23-34.

Kosnik, W., & Acosta, H. (2010). HSI as a method for embedding human-centered design in Air Force rapid acquisition programs. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 54,* 2165-2168.

Kosnik, W., O'Neill, P., & Zimmerlin, Z. (2018). Human systems integration risk management tool. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 62,* 325-329.

Lacson, F. C., Risser, M. R., Gwynne, J. W., & Kosnik, W. D. (2017). The human systems integration framework. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 61,* 1720-1724.

Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica, 65,* 1216-1223.

Miller, M., Thomas, S., & Rusnock, C. (2016). Extending system readiness levels to assess and communicate human readiness. *Systems Engineering, 19,* 146-157.

Olsen, N., & Shorrock, S. (2010). Evaluation off the HFACS-ADF safety classification system: Inter-coder consensus and intra-coder consistency. *Accident Analysis and Prevention, 42,* 437-444.

O'Neil, M. P. (2014). *Development of a human systems integration framework for Coast Guard acquisition* (Unpublished master's thesis). Naval Postgraduate School: Monterey, CA.

Phillips, E. L. (2010). *The development and initial evaluation of the human readiness level framework* (Unpublished master's thesis). Naval Postgraduate School: Monterey, CA.

Phillips, H. (2015). *DOD HFE TAG: Human readiness level (HRL) working group.* PowerPoint slides from william.kosnik.1@us.af.mil

Phillips, H., Stohr, E., Pharmer, J., & Seely, O. (2015). HSI progress & risk specification tool (HPRST). Paper presented at DOD HFE TAG 69, Orlando, Florida.

Sadin, S. R., Povinelli, F. P., & Rosen, R. (1989). The NASA technology push towards future space mission systems. *Acta Astronautica, 20,* 73-77.

See, J. E., Craft, R., & Morris, J. D. (2019). *Human readiness levels in the systems engineering process at Sandia National Laboratories* (Report SAND2019-312). Albuquerque, NM: Sandia National Laboratories.