Performance Shaping Factors as Operator Performance Measures for Validation and the Need for Robust Usability in Human Reliability Analysis

Torrey J. Mortenson and Ronald L. Boring
Idaho National Laboratory, Idaho Falls, Idaho, USA

Abstract: Performance shaping factors (PSFs) are those aspects of the task, environment, or individual that affect human performance. PSFs are frequently used in human reliability analysis (HRA) methods, both qualitatively to identify sources of operator errors and quantitatively to serve as multipliers on nominal human error probabilities. In human factors applications such as control room modernization, common operator performance measures include task completion time, workload, and situational awareness. The list of common measures omits many of the PSFs that are predicted to impact performance. In this paper, we review the applicability of PSFs as operator performance measures for use in validation studies, highlight some difficulties in the current treatment of PSFs, and discuss the need for robust usability exercises as a component of both early stage design and HRA. We examine the clarifying effect that usability testing can bring to design and HRA modeling. Additionally, as modern control rooms become increasingly modular, continual usability testing will be a critical component of the future of HRA and control room design, and a path toward usability can help aid in data collection for HRA and possibly open up new populations to study.

Keywords: Human Reliability Analysis, Performance Shaping Factors, Usability, Validation.

1. INTRODUCTION

1.1 Validation as a Measurement of Success Conditions

A key component of the use of human reliability analysis (HRA) is to inform regulators, operators, and risk analysts of vulnerabilities that are related to the human components of the dynamic systems of nuclear power plants (NPPs). Seldom is HRA considered alongside human factors to shape the design of these systems. To reconcile HRA with design, it helps to begin the consideration of these items in simplest terms; namely that the final goal of validation and verification is to define, measure, and quantify the success conditions related to the necessary operational actions. Once these success conditions can be laid out, the human factors engineer can maximize successes and minimize failures. The complex methods and processes behind both human factors engineering (HFE) and HRA further this simplified end result.

Validation of a system design or process is not the typical purview of HRA. Rather, HRA tends to focus on documenting human contributions to risk. For as-built (i.e., existing) systems, HRA establishes that a system or process falls within reasonable safety bounds for human activities. For new systems, HRA may be used to identify and mitigate such vulnerabilities. Despite the fact that HRA falls predominantly into the arena of risk assessment rather than design, HRA has tremendous potential to inform validation and design activities. As a system is designed, HRA may (and, we might argue, *should*) be used to improve the quality of the validation.

1.2 The Problem - Absence of Insight and Information with Regard to PSFs

We begin with performance shaping factors (PSFs). PSFs impact a component of a human-centered system in such a manner as to modify the potential for success or error for a given task. PSFs are a common component of the HRA process, and are a key effort in attempts to measure the risk of human failure in a manner which mimics that of a system's physical components. However, the use of PSFs in HRA models requires accepting a large range of variability in the predicted model of error due to the fuzzy nature of both our epistemic understanding and aleatory noise in the measurement of PSFs.

Framing PSFs within the context of impacting a given success condition can be problematic due to the difficulty of capturing the true impact a given PSF can have on human performance. It is not novel to calculate reaction times, and there are measures for fatigue and other physiological PSFs (and even some of the cognitive ones); however, there does not appear to be a clear progression, for instance, from a slowed reaction time to an error state, in general. Reaction times are generally not consistent among humans, or among various points of time for the same person. This creates a large amount of intra- and inter-individual uncertainty around a central estimate. Even if all crew members are well-rested and fit for duty to start work, for example, any variation could be compensated for at multiple points in the action chain. Additionally, when we are able to link the causal chain adequately to determine that a PSF was a cause or contributor, it is difficult to generalize that instance outside of the particular user in question due to the fact that we may not have a clear picture of the cognitive processes at play for the user, as well as the role the PSF played. Not only do we lack cognitive penetrability to the mental constructs represented by the PSFs, we also lack the amount of data required to extrapolate error rates to a larger population. The quantity of variation in potential PSFs appears to be rivaled only in the multitude of potential coping mechanisms or compensatory behaviors (i.e., recovery actions) that individuals are capable of manifesting. Therefore, the identification of a cause doesn't always have power of generalization external to the event, which can lead to a more challenging task for HRA analysts.

As an example, recent research on fatigued driving made a similar link in that the fatigue was measurable and manifested in a predictable fashion; however, the cognitive processes that created an error state are murky, and it is difficult to state that driver A, due to a fatigued state, performed a task in such a way that led to an error due to cognitive faults X and Y [1]. In validating errors, either in a prospective or retrospective investigatory fashion, we need to be cautious and accept the realities that while we can measure certain physiological characteristics and link them to other physiological actions or inactions, we do not always have the requisite insight to understand, and make conclusions, regarding the cognitive mechanics at play.

1.3 Refining PSFs and Their Use

While PSFs have flaws in terms of the generalizability of their predictive function, we do not need to undermine their value or to throw them out altogether. Rather, there is a need for adjustments in the use of PSFs, an industry-wide doubling down on data collection efforts, and a realignment of where a given PSF is of the most value. Currently, in using PSFs as a component of risk models in licensing a design as satisfactory for use, we may be misapplying the power of PSFs. There are certainly key uses for them at this stage, such as ensuring the design achieves accessibility for all users in relation to the physical characteristics of the work and user base. However, maximizing the use of PSFs necessitates employing them at the zero day of design. The focus of this paper is on the use of PSFs at this crucial early design phase. We believe that the PSFs serve not only to account for errors and determine appropriate safety margins but also to inform the design of systems to eliminate sources of errors and minimize any consequences of those errors.

Performing iterative usability tests consistently with systems design and other formative evaluation activities allows analysts to identify relevant PSFs to the particular work tasks and to identify mental models and cognitive load issues before an error state occurs. A thorough experimental strategy focused on usability can use quantitative measures to account for the quantifiable, and qualitative measures during post-test debriefings can inform the design of the systems in dramatic ways, as demonstrated in recent studies at Idaho National Laboratory [2]. These debrief sessions, performed writ large as a measure of control room modernization efforts, have the potential to eventually generate a set of principles derived from cognitive processes of operators and understanding of their mental models that could greatly improve workflows and overall odds of a success condition. So, PSFs are not without value, but there needs to be an attempt to realign their use in the process and to integrate the insights that can be gained from them into the overall systems design. An iterative process can help define the PSFs for which job tasks or users are more susceptible. These engagements can also allow the analysis or design teams to model various success conditions and failure states in order to mitigate future human error as greatly as possible in a no-risk environment.

1.4 Validation and Verification, Generally

Notions of validity and verification are not newcomers to skepticism in HRA and other components of human factors. As stated by Fuld, "technical meanings of verification and validation are often a point of debate" [3]. Murkiness aside, there does need to be a delineation of these terms in practice. Verification has been stated to be a process regarding the truthfulness of a thing or process [3]. While that is a potential definition, it does seem to lay a large burden on analysts claiming to verify something, especially in safety-critical domains such as nuclear. Ultimately, we are set to verify X against Y, where Y is the arbitrary component that possesses some quality or value of 'truth' that has been accepted. The difficulty with using this definition of verification with regard to PSFs and the broader HRA process is that we do not have firm resources to verify against. Therefore, we need to first accept that we are not currently in a position of having something objective to use across the industry for verification, and begin a process of discovering how we can identify or develop these items for future verifications.

Validation, for Fuld [3], has a more malleable definition that shies away from difficult to manage concepts like 'truth' by stating it is the process of confirming the strength of something. This is something we should be able to currently perform, albeit with some minor adjustments to existing procedures. Validity as a concept can draw both ancient and modern use cases from philosophical logic and its operations. In logic, validity is the measure of the form of an argument and is the primary guideline for a complete argument. Validity being the form, and soundness the content, there is some rough analog in the concepts of necessary and sufficient. Validity or a logical form for your argument is necessary. Without it, even the most compelling argument is illogical, and soundness is sufficient. Only both will end with a quality argument that can be considered defensible and free from fallacy. Similar to mathematics, an error in form can cause many problems in argumentation. While we cannot necessarily use argumentative validity in control room designs, safety analysis, or other aspects of HRA, we can use the notion of form and structure to validate processes in design and operations. While it would be too far to assume we can assign a form of design or operation across *all* of the nuclear industry, perhaps we can indeed arrive at an understanding of a process which should be used across the industry in terms of giving some foundational support to our activities in HRA.

The literature in HRA regarding PSFs and the component struggles with defining validation and verification is reminiscent of a prominent article on the so-called Chinese Room in the study of philosophy of mind criticizing the notions of syntactical processes of artificial intelligence and the notion of computation as consciousness [4]. In the article, Searle uses a thought experiment to illustrate that the syntactical nature of following an ensuing order of processes to achieve a result does not demonstrate understanding, and therefore the Turing test is flawed in determining whether something has a mind. The thought experiment follows as such: assume an individual is locked into a room with a manual that instructs manipulations of Chinese symbols. People outside of the room are placing questions into a slot written in Chinese. These people believe they are inserting questions into a computer, and the person inside does not have contact with the outside world and does not know that questions will be coming into the room in Chinese. However, he follows processes from the manual and is able to create a result in Chinese and returns the program out the other side of the room. He has performed the necessary processes of translation, but he obviously has not attained understanding. There are further implications related to consciousness and intentionality, but for our use it highlighted some key issues that may be present in our understanding of HRA and its performance. We see stimuli going into the 'Chinese room' of our user's cognition, and we see outputs in terms of actions. At any point the underlying phenomena behind PSFs can impact these stimuli or the resulting actions, or even the room itself, and we are trying to quantify and manage that without any testing or examination of the process. Validation and verification are even more troubling pursuits when considered in this way, because we cannot access the user's cognition and therefore cannot validate the computations. Without substantive usability testing, all we are able to do is to arrive at informed guesses and anecdotes about the impact these PSFs may or may not have on the system.

2. PSFS, PERFORMANCE MEASURES, AND USABILITY

2.1 Clarification and Specificity

Whether using direction from the Technique for Human Error Rate Prediction (THERP) [5] or the Standardized Plant Analysis Risk-Human (SPAR-H) [6] methods, PSFs have been enumerated in some detail, and it has also been acknowledged that quantifying the human error components of these NPP systems can be difficult. The PSFs of interest in HRA applications can be numerous, though both SPAR-H and THERP guidance make efforts to limit the number of PSFs. These streamlined lists of PSFs should not be viewed as exhaustive, however, and they should also serve as a guiding point as attempts are made to calculate error rates as best possible. Unfortunately, this guidance can be difficult to apply and may miss the critical components of PSFs in pursuance of a number. While PSFs have been somewhat isolated to HRA, the notion of external factors impacting human performance is a focus of human factors generally, and usability particularly. With the former focusing on users and the latter focused on design, both use PSFs of a different name, so to speak. HRA is uniquely positioned to help guide usability efforts as a component of design due to the longstanding formal treatment of PSFs in the field. This renewed focus can help both camps as usability could benefit from HRA methods and insights, and HRA can benefit from a sharper picture of user interactions with systems.

PSFs catalog phenomena, but they also serve another key point in design, which is to maximize the opportunity for success conditions and to minimize failure points. PSFs are not simply measurements (i.e., dependent variables) but rather also something that can be manipulated or optimized to improve performance (i.e., independent variables). While these attempts to use PSFs to calculate human error rates are not without value, there are some key actions that can help provide clarity and specificity to the PSFs that may be particularly impactful for tasks in control room operations. Especially as a component of control room modernization efforts, display and task evaluation for modern human-system interfaces have the potential to generate their own form of PSFs and can impact operators in unforeseen ways. Iterative usability testing, from project inception, covering task analysis, information architecture design, and other critical use-case actions are critical to identify not just PSFs that may be unique to a given system, but may also highlight more general PSFs that occur in these systems that an analyst may not have been able to access if usability was only performed as a measure of validation. Therefore, performing iterative, low-risk usability studies as a parallel to the system design can be incredibly informative to identify and quantify possible human error probabilities. Data from the usability studies will be able to inform the HRA and can give actual error data in a low-risk scenario as well as open up opportunities for the system designers to design out any particular error focused points or deploy other mitigation strategies to minimize human error risk.

As discussed by Boring [7], performing verification and validation (V&V) can sometimes get swept up by new and exciting advances in psychological science. We would additionally highlight that usability studies and similar task analyses are bread and butter engineering psychology. In addition to the tried and true nature of these activities, they operate from a place of *tabula rasa* - or a place of a blank state. Usability testing is performed without assumptions of user's actions, needs, or preconceived notions. One aspect of V&V and broader operational/safety tasks in high risk industries, like nuclear, is that it is possible to get lost in the echo chamber of the expert and highly specialized nature of these activities, and to some degree we can design the human components of our systems while using unproven assumptions. While it is true that operators are highly trained and educated, and that there are significant risks involved in the performance of their work, it is crucial to realize that human beings are the ultimate interchangeable part. As a component of a system they are extremely malleable and resilient to shifting situations. Performing usability studies from a point outside of the expert bubble can open up the design to many things that may not have been considered in previous designs and may not have come out if the study was restricted. Especially as modern control rooms and digital interfaces are used with increasing frequency, system designers will be faced with a task of designing new paradigms for the first time in decades. As a whole, the industry will require data and a new approach to these new challenges.

2.2 Do PSFs Serve Our Needs in V&V?

PSFs, within the context of V&V, are designed to inform the ultimate determination of the human system's efficacy in operating within the confines of the success and error criteria established for the system [8]. They serve as a component of engineering verification as discussed in [5], however, with a slight tweak on the discussion regarding the criteria necessary for verifications. Fuld states that verification criteria exist within fixed limits and that some improper criteria may be more discretionary or contextual. However, as explained in [3] the qualitative aspects of usability studies have substantive value for design and informing considerations regarding users. Additionally, without a prior component similar to iterative usability studies earlier in the process, the fixed limits within which verification can occur are, as Fuld stated, arbitrary. This arbitrary aspect of the verification process is a potential issue because it has no foundational link outside of the expert level review. Due to the malleability of the human components of these systems it can be extremely difficult to confidently identify true PSFs to verify against without testing and identifying those characteristics unique to either the operators or systems. Adequate design for control rooms are not necessarily underpinned by exclusive principles. Consider modern smartphones and apps and how many people use these devices and software solutions with large rates of success. Without usability studies to inform the verification criterion, the arbitrary limits established are exposed to error themselves.

2.3 Performance Measures and their Applicability

In an odd sort of meta-V&V performance measures serve to perform verification and validation of error causation and PSFs by giving a measurement system to quantify certain characteristics of operators, and by extension, the systems. For many performance measures there are documented cases speaking to their potency and the performance of the testing methods. This gives a level of credence to these measures that is greater than that of some of the PSFs identified prior. Guidance regarding some of these measures and their use can be found in [9]. As discussed in that document, there is significant value to utilizing the performance measures at various stages of the design, however critically there are risks of over-complicating the models, and thereby obfuscating the actual factors, and potential errors that the measures are intending to isolate. Achieving simplicity and adopting the blank slate mentality rather than seeking specific prescribed error states or characteristics should remain a key focus of the HRA process. While some performance measures are fairly robust with regard to their literary pedigree there are still questions that need work. Analysts and practitioners should be cognizant of the reality that while they are examining errors, they can also commit errors. To quote Sherlock Holmes, "It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts." Therefore, beginning from the first principles of these activities should be the primary focus.

The repeated mantra of this paper has been that there is a need for iteration of usability testing and analysis at earlier states of the design and risk assessment process. Once usability tests have identified some particular PSFs and areas for improvement, there should be some effort to identify the modifiers that impact an operator's performance as well. The guidance in SPAR-H [6] with regard to multiplier calculations can inform these, but they too require some validation, for lack of a better term. With some repeat testing we can begin to assess aspects of training, experience, and other heuristics that impact, positively or negatively, success states of NPP operations, and that information can be used to generate more focused trainings or inform particularly useful experiences for operators. It isn't clear that there is an understanding of the cognitive impacts of training, experience, education, or a manner of gauging these items for efficacy currently, and building quantitative models with real data would be immensely valuable to future design and may inform future design choices in ways that might be beyond our current understanding of PSFs or use of the human operator components in these systems.

Therefore, it is suggested here that in addition to the performance measures currently used by the HRA community, prospective HRA and design exercises should look to add a robust usability testing framework to these performance measures. A useful explanation of some of the mathematical power of usability characteristics can be found in [11]. As explained, through robust modeling techniques it is possible to identify optimal user counts for maximum detection of usability issues in interfaces. It also can give some guidance as to how to structure some of the quantitative aspects of usability testing that could be extremely valuable for these uses.

2.4 Digitalization of Future Control Room Design and HRA

As discussed in [7] there is a view of usability and the related design items as ending with final V&V. The historical view is that after final licensing the control room will exist in its form until a future modernization effort. In previous generations that perspective was understandable as switching out analog components was significantly more difficult due to the physical nature of the control room displays and instruments. However, as control rooms become more modern and digital, the act of adjusting displays, overlays, or other human-interface components becomes a simpler change in the digital backend. The cost is purely time and expertise of the developers behind the system. In considering that, HRA practitioners and other interested stakeholders can help make constant design and improvement that was previously cost prohibitive in many cases a component of regular assessments. The digital nature of these interfaces is extremely compatible with more modular control room units that can be continually focus tested and swapped out in the interest of increased efficiency, safety, or reliability. Likewise, the licensing considerations of these hypothetical modular, digital interfaces should also be examined to allow valuable iteration of these modules. In this digital age we can see apps focus test, prototype, and roll out to a user base in a manner of months or sometimes weeks. While we likely cannot have a timeline so truncated, we should work to streamline the process of usability operations as much as possible in order to make use of current and future operator populations to continue to use performance measures within usability studies to effectively deliver the control rooms of the future.

3. THE THREE-LEGGED STOOL OF HRA, HFE, AND USABILITY

3.1 Characteristics and Complements

The field of human factors engineering (HFE) has likely been considered the primary, or overarching set of theories and ideas within which HRA and usability make their home. However, we believe that there is value in separating them into three equal components that share concepts and serve to inform each other. HFE has largely been defined as the loose adaptation of engineering principles to the human components of systems. This can encompass everything from visual design, to safety, and ergonomics. At its core, it is the examination and, through testing, the clarification of the limits and efficiencies of the human part of a system. Therefore, usability to some degree is derived from task analysis and other HFE concepts, but has recently, with the dramatic increase in the role digital devices play in our lives, become a field in its own right.

Usability can be loosely defined as the study of improving user successes with some system, whether physical or digital, accomplished through iterative use testing with quantitative and qualitative measured variables. Another possible definition is simply that usability is a quality variable that identifies the ease of use of a given interface system [12]. The field has become more robust and advanced in recent years, and there have been leaps in understanding testing paradigms and user psychology. While usability originally used items from HFE like Gestalt psychology and other design principles derived from sensation and perception research, it has grown to encompass user emotional states, stressors, environmental conditions, and more. Through a robust understanding and implementation of usability studies a design team can isolate and manipulate interfaces and settings to sharpen the understanding of factors that may influence a user's performance in completing various tasks. We call them PSFs, but they are not an isolated focus of safety-critical industries.

HRA is obviously the third leg of our stool, and has been defined earlier in the paper. However, in drawing on HFE principles and data from usability, HRA models can achieve significant power in predicting error, increasing user success, and maximizing efficiency through the integrated concepts from probabilistic risk assessment. HRA can be seen to be the end result or final step in a 'cradle-to-grave' design journey that wouldn't be possible without all three of the legs, and rather is severely flawed without the other two.

3.2 Aspects of Usability and Their Additions to HRA

HRA can benefit in some notable ways from employing usability testing as a component of a system's design. HRA has been largely isolated to calculating probabilities of human error in a system, but can also potentially quantify aspects that would increase success conditions in the user's actions. While each operational team may have differing usability priorities due to the different nature of control rooms, technologies, and more, there are some key activities that can be put to work to design control rooms with maximum success potential.

Information architecture (IA) is a broad field examining the way information can be organized, accessed, and catalogued. Impacting fields ranging from modern app development to library operations, there are information architects everywhere. The final deliverable of an IA process is some form of graphical representation of a user's mental model. The mental model is the construction of how information should be shown or organized according to the user. It can be shaped from the user's particular cognitive understanding of information items to past experiences interacting with systems. These representations can illustrate the relative hierarchies and taxonomies that users expect from the system and can be incredibly informative to design and engineering teams to design a system that the users expect. A very simple way to avoid error states can be to design a system that operates the way the user expects it to, and is laid out in a similar fashion.

User experience/user interface design (UX/UI) has been more commonly employed in development of websites, applications, and consumer software, but can be extremely useful to HRA and other HFE sub-fields. Similar to IA in designing a system the user expects, you can work to design a system that the user *prefers* to use. While it isn't likely that the operator can easily use another control room for their work, it has been shown that user's enjoyment of interacting with a system can impact their likelihood of success of use. Therefore, employing user experience testing and employing sound interface design principles can help avoid issues down the road.

In exercising these tasks, not only will the design become better for use, but your process can also yield valuable data on what worked and what didn't. These data points are real, representative, and can give the strength to HRA that has been missing in previous processes. Currently, we can use SPAR-H or a similar methods' error multipliers, or through usability testing we are able to identify real multipliers in the data that can be relied on to build the eventual HRA model. Through these processes, a best practices or design guide may be possible to achieve for the industry, but as we stand now our treatment of PSFs and the ensuing multipliers for design remains underspecified.

Performing these exercises early in the design stages also can have immense value when your team is able to mitigate or eliminate future error or increase efficiency down the line. There are significant changes that may come out of usability testing and those adjustments get more expensive the longer the project design goes on. As Frank Lloyd Wright famously said, "You can use an eraser on the drafting table or a sledge hammer on the construction site."

3.3 Modularity as a Reality

As a further point on the reality of necessary continuous usability testing in modern control rooms, it is important to understand a critical difference in where control rooms were and where they are going to be in the future. As the embrace of digital interfaces increase among nuclear industries and automation similarly increases it is important to note that we are not designing modular systems, though we may think we are. Instead, we are designing systems with modular components. Digital components are, by the very reality of their concept, modular. What would have taken engineers, manufacturers, and raw materials to create in the 1960s can now be done with some lines of code, and tested in a machine learning predictive environment. Therefore, usability does not end after licensing and should be a constant focus of the modern nuclear utilities and agencies. We should embrace, with skepticism, new design ideas and have an infrastructure ready to support change in control rooms of the future. We can now perform, and should demand, updates to our interfaces more than once every forty to fifty years.

4. DISCUSSION AND CONCLUSION

4.1 Without Robust Usability Testing the HRA Model is Flawed

Critically, the HRA modeling processes laid out in THERP, SPAR-H, and others appear reliant on the PSFs as drivers of the human error probabilities. This exposes the model to risk due to the, as-stated, epistemic uncertainty and aleatory nature of the PSFs as we understand them. The interrelation and interaction between PSFs themselves and between PSFs and tasks are complex and extremely difficult to generalize. Therefore, in the absence of a process or system to check the alignment of the tasks and the designs against user's understanding and capabilities we may be missing opportunities for understanding the impacts that PSFs may have on error or success, what specifically modifies those states, and with what frequency. It would appear then that our current system of performing HRA as a component of probabilistic risk assessment is based on multipliers that are educated assumptions, but may be inaccurate when compared with data from usability studies.

4.2 Usability Brings Known and Unknown PSFs to the Surface as Independent Variables

A recent debate in HRA is whether we should treat PSFs as independent of dependent variables [8]. Are these concepts that should be purely measured or manipulated? At least initially, it is not clear that we have a firm enough understanding of the PSF's role in causation and the relative effects on operator action to treat them in any manner other than independent. Additionally, there is a debate over how many PSFs a model should measure or take into account [10]. Usability testing gives clarity to both concerns and allows analysts to begin to see relevant PSFs as they will operate as independent variables in the study, as well as within the context of plant operations. As stated elsewhere in this paper, multiple PSFs can be present at any time, and some PSFs may be unique to your particular testing or operating environment. A possible solution to the problem of identifying the priority PSFs for the relevant tasks and to access the otherwise unknown PSFs is to employ testing to bring these concerns to the surface, and usability testing provides an optimal method to gather these much-needed data points. Our knowledge without these exercises ends at being hypothetical or anecdotal. There is no better way to analyze what will impact users than having them actually use the product. Even more strongly it can be argued that there is no way to analyze actual user impacts without having them actually use the product.

4.3 Iteration in Usability Can Allow PSFs to Become Dependent Variables

Once we have analyzed the data and identified the key PSFs for a particular situation, then we can begin understanding their effects on the user, and the resulting multipliers for the risk model. We do this by continuing the iteration of the usability testing but treating the PSFs as dependent variables and manipulating their interactions with the user base. We can gain data on the effects of PSFs on error by increasing or decreasing the presence of the PSF and measuring the resulting error rates. This information could inform everything from training decisions to work hours and more. The critical component, however, is that we are then in a position to create error in a safe environment and ensure the resilience of the design and users, and we can work to then mitigate any sources of said error. In safety critical industries in which error can have catastrophic consequences, the opportunity to gain this knowledge is invaluable and it may very well be irresponsible to not take advantage of it.

4.4 Usability Testing Can Only Sharpen the HRA Model

One of the largest benefits to a focus on usability testing is the reality that it literally cannot harm the project and can only make the project better. Whether we are able to identify PSFs we weren't aware of or make the design more capable of success conditions, we are always going to be in a more advantageous place than we were before. Aside from the aforementioned clarity of our epistemic position, our HRA model becomes significantly more powerful and statistically significant when backed by these data observations.

4.5 Generalizability of PSFs as a Function of the Availability of Data

A key difficulty in safety-critical, high-expertise industries can be finding representative users to test the product or designs. This issue is not insignificant and is a serious hurdle to significant usability testing. However, through usability testing we can include non-operators in studies and then evaluate the comparisons of the group's error rates. We may be overly restricting our own data pools by focusing on what we deem as necessary expertise to perform control room operations. While it is clear that reactor operators are highly educated and skilled individuals, we may be able to isolate some tasks that could make use of a larger population of non-operators to test. Human beings are the ultimate interchangeable part and are extremely resilient to system variation, but there clearly are simultaneous notions about digital interfaces that are relevant to everyone. Facebook, Google, Apple, and others rely on the homogeneity of users to sell products to billions of people. We can take a lesson from these commercial giants by opening some aspects of our design process to broader audiences in order to gain insights and power of our HRA models and eventual designs.

4.6 Next Steps

The next steps in the process of HRA should be the adoption of usability ideals and principles across the organizations involved in control room modernization, process improvement, or current operation. The current model of performing HRA has some room for improvement and could be made to be significantly more accurate with the inclusion of these testing forms. We also need to accept that our current operator population is too small and spread out to be used in any significant fashion to enhance our current modernization efforts. We should begin to examine other populations for opportunities to gain insight into our users and how to best design future interfaces. It bears mentioning that this prospective focus of HRA is not a common perspective of how HRA can impact a system and its design; however, we feel that there are significant overlaps between HRA as performed and usability principles. In addition to the use cases and similarities, we believe that this refocusing may give new avenues to data collection for HRA, new populations to survey, and greater designs for the industry as a whole.

Disclaimer

The opinions expressed in this paper are entirely those of the author and do not represent official position. This work of authorship was prepared as an account of work sponsored by Idaho National Laboratory, an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights. Idaho National Laboratory is a multi-program laboratory operated by Battelle Energy Alliance LLC, for the United States Department of Energy under Contract DE-AC07-05ID14517. This research was funded through the Laboratory Directed Research and Development program at Idaho National Laboratory.

Acknowledgements

Torrey Mortenson would like to thank Dr. Ronald Boring and Dr. Thomas Ulrich for their advice, discussions, and editing help in the completion of this paper, as well as helping a philosopher take the dive into scientific study.

References

[1] Croft, R., Howard, M., Jackson, M., Kennedy, G., Owens, K. Cognitive components of simulated driving performance: Sleep loss effects and predictors. *Accident Analysis and Prevention, 50*, pp. 438-444. (2013).

[2] Ulrich, T., Qualitative or Quantitative Data for Nuclear Control Room Usability Studies? A Pragmatic Approach to Data Collection and Presentation. Human Factors and Ergonomics Society International Annual Meeting. (2018).

[3] Fuld, R., V&V: What's the Difference? *Ergonomics in Design*, pp. 28-33. (1997).

[4] Searle, J., Minds, Brains and Programs. *Behavioral and Brain Sciences, 3*, pp. 417–57. (1980).

[5] Guttmann, H., Swain, A. "Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Operations, NUREG/CR-1278," US Nuclear Regulatory Commission, 1983, Washington, D.C.

[6] Gertman, D., Blackman, H., Byers, J., Haney, L., Smith, C., & Marble, J. "The SPAR-H Method, NUREG/CR-6883," US Nuclear Regulatory Commission, 2005, Washington, D.C.

[7] Boring, R., Envy in V&V: An Opinion Piece on New Directions for Verification and Validation in Nuclear Power Plants. Human Factors and Ergonomics Society 59th Annual Meeting. (2015).

[8] Boring, R., Blackman, H., and Rasmussen, M. Preliminary considerations for human reliability analysis for verification and validation of new designs. PSAM Topical Conference on Human Reliability, Quantitative Human Factors, and Risk Management, Munich, Germany. (2017).

[9] Boring, R.L., Joe, J.C., Ulrich, T.A., & Lew, R.T. Operator Performance Metrics for Control Room Modernization: A Practical Guide for Early Design Evaluation. INL/EXT-14-31511. (2015).

[10] Boring, R., How Many Performance Shaping Factors are Necessary for Human Reliability Analysis? PSAM.

[11] Landauer, T., Nielsen, J., (1993) CHI '93 Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. pp. 206-213.

[12] Nielsen, J., (2012, Jan. 4). Usability - 101: Introduction to Usability. Retrieved from https://www.nngroup.com/articles/usability-101-introduction-to-usability/