Advances in Intelligent Systems and Computing 956

Ronald L. Boring *Editor*

Advances in Human Error, Reliability, Resilience, and Performance

Proceedings of the AHFE 2019 International Conference on Human Error, Reliability, Resilience, and Performance, July 24–28, 2019, Washington D.C., USA

Springer

Advances in Intelligent Systems and Computing

Volume 956

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within "Advances in Intelligent Systems and Computing" are primarily proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink ****

More information about this series at http://www.springer.com/series/11156

Ronald L. Boring
Editor

Advances in Human Error, Reliability, Resilience, and Performance

Proceedings of the AHFE 2019 International Conference on Human Error, Reliability, Resilience, and Performance, July 24–28, 2019, Washington D.C., USA

Springer

Editor
Ronald L. Boring
Idaho National Laboratory
Idaho Falls, ID, USA

ISSN 2194-5357 ISSN 2194-5365 (electronic)
Advances in Intelligent Systems and Computing
ISBN 978-3-030-20036-7 ISBN 978-3-030-20037-4 (eBook)
https://doi.org/10.1007/978-3-030-20037-4

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Advances in Human Factors and Ergonomics 2019

AHFE 2019 Series Editors

Tareq Ahram, Florida, USA
Waldemar Karwowski, Florida, USA

AHFE International Applied Human Factors and Ergonomics Conference 2019 www.ahfe.org

10th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences

Proceedings of the AHFE 2019 International Conference on Human Error, Reliability, Resilience, and Performance, held on July 24–28, 2019, in Washington D.C., USA

Advances in Affective and Pleasurable Design	Shuichi Fukuda
Advances in Neuroergonomics and Cognitive Engineering	Hasan Ayaz
Advances in Design for Inclusion	Giuseppe Di Bucchianico
Advances in Ergonomics in Design	Francisco Rebelo and Marcelo M. Soares
Advances in Human Error, Reliability, Resilience, and Performance	Ronald L. Boring
Advances in Human Factors and Ergonomics in Healthcare and Medical Devices	Nancy J. Lightner and Jay Kalra
Advances in Human Factors and Simulation	Daniel N. Cassenti
Advances in Human Factors and Systems Interaction	Isabel L. Nunes
Advances in Human Factors in Cybersecurity	Tareq Ahram and Waldemar Karwowski
Advances in Human Factors, Business Management and Leadership	Jussi Ilari Kantola and Salman Nazir
Advances in Human Factors in Robots and Unmanned Systems	Jessie Chen
Advances in Human Factors in Training, Education, and Learning Sciences	Waldemar Karwowski, Tareq Ahram and Salman Nazir
Advances in Human Factors of Transportation	Neville Stanton

(continued)

(continued)

Advances in Artificial Intelligence, Software and Systems Engineering	Tareq Ahram
Advances in Human Factors in Architecture, Sustainable Urban Planning and Infrastructure	Jerzy Charytonowicz and Christianne Falcão
Advances in Physical Ergonomics and Human Factors	Ravindra S. Goonetilleke and Waldemar Karwowski
Advances in Interdisciplinary Practice in Industrial Design	Cliff Sungsoo Shin
Advances in Safety Management and Human Factors	Pedro M. Arezes
Advances in Social and Occupational Ergonomics	Richard H. M. Goossens and Atsuo Murata
Advances in Manufacturing, Production Management and Process Control	Waldemar Karwowski, Stefan Trzcielinski and Beata Mrugalska
Advances in Usability and User Experience	Tareq Ahram and Christianne Falcão
Advances in Human Factors in Wearable Technologies and Game Design	Tareq Ahram
Advances in Human Factors in Communication of Design	Amic G. Ho
Advances in Additive Manufacturing, Modeling Systems and 3D Prototyping	Massimo Di Nicolantonio, Emilio Rossi and Thomas Alexander

Preface

Welcome to the third installment of the International Conference on Human Error, Reliability, Resilience, and Performance (HERRP)! The conference was organized within the framework of the International Conference on Applied Human Factors and Ergonomics (AHFE) as an affiliated conference.

The 1st HERRP International Conference took place at the Westin Bonaventure Hotel, Los Angeles, California, USA, from July 17 to 21, 2017. This gathering featured a largely curated list of significant researchers in the field, with a particular emphasis on human reliability analysis. Unlike other risk conferences, which tended to be centered largely on probabilistic risk of hardware systems, HERRP had a decidedly human factors angle. The research presented explored human error from a human factors perspective, not solely a risk modeling perspective.

When you organize a new conference, you invite as many of your research colleagues as possible to participate. I am pleased that many of these individuals have chosen to continue to write papers for and attend HERRP. Additionally, I am delighted that the scope of the conference has expanded considerably. For example, in this installment of HERRP, we see an emerging significant focus on resilience. The interplay of resilience and reliability—indeed, the connections between human error, reliability, resilience, and performance research are coming together. As these different angles on a common topic enter into dialogue, it is my hope that these fields will converge and begin to answer research questions together.

The purpose of the HERRP conference is to bring together researchers and practitioners in different fields who broadly share the study of human error. The HERRP conference is intended to serve as an umbrella for human error topics by providing an annual forum for otherwise disjoint research efforts. As such, the conference is intended to complement but not replace existing specialized forums on particular facets of human error. The HERRP conference is distinctly interdisciplinary, encouraging the submission of papers in focused technical domains that would benefit from interaction with a wide human factors audience. Additionally, the HERRP conference aims to provide a yearly, high-quality, archival collection of papers that may be readily accessed by the current and future research and practitioner community.

The HERRP scientific advisory board invited papers related to a broad range of topics on human error, including but not limited to:

Human performance
Human variability
Human reliability analysis
Human performance shaping factors
Root cause analysis
Accident investigation
Human resilience and resilience engineering
High-reliability organizations
Safety management
Medical error
Driver error
Pilot error
Automation error
Defense in depth
Errors of commission and omission
Human error taxonomies and databases
Human performance improvement and training
Cognitive modeling of human error
Qualitative and quantitative risk assessment

Many of these topics and others are reflected in these proceedings. Contributions encompassed empirical research studies, original reviews, practical case studies, meta-analyses, technical guidelines, best practices, or methods. Papers encompassed traditional topics of human error such as found in the safety-critical industries like energy, manufacturing, and medicine. We also encouraged innovative explorations of human error such as security, defense, new human–technology interactions, and beneficial uses of human error.

The sections of these proceedings are grouped as follows:

Section 1	Theoretical Issues in High-Reliability Organizations
Section 2	Theoretical Advances in Human Error and Performance
Section 3	Human Error Considerations in Design
Section 4	Personal Resilience
Section 5	Human Reliability Analysis for Safety-Critical Industries
Section 6	Tasks and Errors
Section 7	Human Resilience in Aviation Systems

It has in my view been a very successful third installment of the HERRP conference, and I look forward to watching the evolution of this conference. I am grateful to the organizing committee of the 10th International Conference on Applied Human Factors and Ergonomics and the affiliated conferences for making this embedded conference possible. I also wish to thank the authors for their exceptional contributions to the conference, and to scientific advisory board for encouraging strong submissions:

Harold Blackman, USA
David Gertman, USA
Katrina Groth, USA
Xuhong He, Sweden
Stacey Hendrickson, USA
Yochan Kim, Korea
Barry Kirwan, France
Karin Laumann, Norway
Zhizhong Li, China
Peng Liu, China
Ron McLeod, UK
Myriam Merad, France
Naj Meshkati, USA
Aino Obenius-Mowitz, Sweden
Jinkyun Park, Korea
Manikam Pillay, Australia
Alice Salway, Canada
Carol Smidts, USA
Oliver Straeter, Germany
Claire Taylor, Norway
Patricia Trbovich, Canada
Matt Weinger, USA
Thomas Ulrich, USA
April Whaley, USA
Jing Xing, USA
David Yacht, USA

To err is human, and human error is consistently implicated as a significant factor in safety incidents and accidents. Yet, as pervasive and important as human error is, the study of human error has been fragmented into many different fields. In fact, in many of these fields, the term "human error" is considered negative, and terms such as human variability or human failure are preferred. Across differences in terminology and approach, the common link remains an interest in how, why, and when humans make incorrect decisions or commit incorrect actions. Human error often has significant consequences, and a variety of approaches have emerged to identify, prevent, or mitigate it. These different approaches find a unified home in this conference.

July 2019 Ronald L. Boring

Contents

Theorizing and Theoretical Issues in High Reliability Organizations

Investigating Collective Mindfulness in Mining: A Prospective Study in High-Reliability Organizations

Manikam Pillay[1](✉), Michael Tuck[2], and Karen Klockner[3]

[1] School of Health Sciences, Centre for Resources Health and Safety, The University of Newcastle, Callaghan, Australia
manikam.pillay@newcastle.edu.au
[2] School of Science, Engineering and Information Technology, Federation University, Ballarat, Australia
[3] School of Human, Health and Social Sciences, CQUniversity, Brisbane, Australia

Abstract. Mining is an important contributor to the social and economic fabric of our society. However, it continues to be regarded as one of the most dangerous industries. Compared to manufacturing, mining is more complex, which can pose additional challenges for mining and safety managers in terms of achieving sustainable safety outcomes. More advanced approaches are required. This paper first discusses the state of mining safety in Australia, followed by an examination of some of the complexities that characterizes the industry. It then introduces High-reliability organizations and Collective mindfulness as an advanced organizational safety management strategy that can be used to achieve sustainable safety improvement. A pragmatist research framework and two organizational theories follow this, which can be used to inform further research in these fields. The paper concludes with a research proposition which can be used to empirically investigate these concepts in mining organizations.

Keywords: High-reliability organizations · Collective mindfulness · Socio-technical system · Social construction of safety · Organizational safety management · Mining safety

1 Mine Safety Performance in Australia

Mining is a major contributor to national income, investments, exports and government revenues in Australia and continues to be recognized as a key driver of higher living standards in Australia. It currently contributes to 18 percent of the nominal Gross Domestic Product [1] and provides employment to over 187 000 workers [2]. However, it continues to be regarded as a poor performer in terms of its safety performance [3, 4] with 121 fatalities in the industry between 2003 and 2015, an average of 9 deaths each year. Figure 1 illustrates trends in fatalities in the Australian mining industry from 2003 to 2015. Some improvements were observed until 2010, but these have worsened or plateaued off, a trend that has been previously identified [3, 5].

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 3–12, 2020.
https://doi.org/10.1007/978-3-030-20037-4_1

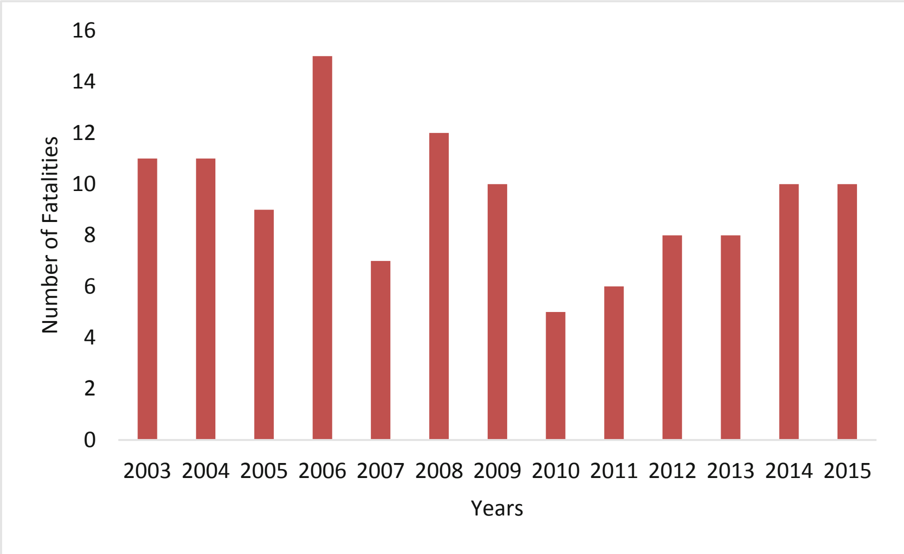


Fig. 1. Fatalities experienced in the Australian mining industry 2003–2015

These raise serious questions about the industry's capacity to maintain or improve safety performance [6]. They also suggest that many of the existing, contemporary approaches for mining safety have failed to drive sustainable improvements in the industry because they are outdated and have not kept pace with emergent organizational theory and practice. More innovative strategies are required [5, 7].

1.1 Mining: A Complex Industry

Unlike manufacturing, mining is seen as a more complex industry [3, 5, 8] due a range of environmental, organizational and social factors. Some of these include:

- uncertain rock strata and structure,
- optimization approaches for open-pit and underground mining,
- a high-percentage of contracting and sub-contracting worker arrangements
- mix of highly skilled and semi-skilled operators operating both old and newer mining equipment, some of which can be sophisticated and beyond the cognitive reach of operators, and
- increase in stakeholders, new technologies, interactions between the different entities involved and nature of events.

What this means is that any innovations in safety management in mining need to account for such complexities.

2 High-Reliability Organizations

High reliability organizations (HRO) represent a very selective group of systems that are able to sustain relatively high levels of safety performance while operating in a high risk, volatile and uncertain work environments all the time [9]. They were first conceptualised at the University of California, Berkeley, to explain how some high-risk organizations were able to maintain a near-error free operation despite operating with very complex technologies. They chose three industries to illustrate their case; a nuclear power generation stations, air traffic control and aircraft carriers [10–15]. These early studies suggested that systems could achieve HRO status by paying particular attention to organizational design and functions that allowed it to self-regulated and adapt to changing demands [10, 12, 13], continued training, re-training and learning throughout the organization [10–14], critical decision-making at all levels [10–13], maintaining reserve (slack) resources [10–12, 14], maintaining multiple sources and flow of information [11], and a culture focussed on reliability [11–13, 15]. These organizations also:

- maintained high levels of technical competence throughout the organization,
- constantly searched for improvement across a number of different dimensions of reliability,
- invested in staff who analyzed and maintained detailed records and analysis of any events irrespective of size,
- elaborated on and developed structured procedures and work practices which were closely linked to ongoing analysis and directed towards avoiding precursor conditions, and
- maintained formal structure of roles, responsibilities and reporting relationships that were strongly hierarchical; but flexible enough to be transformed and re-organized during emergencies to ensure those with most knowledge of the events at hand were involved in managing them [9, 16, 17].

2.1 Collective Mindfulness, Safety Culture and Safety Climate

The term collective mindfulness (CM) captures those attributes that contributes to a system becoming a HRO. CM represents a cognitive mindset that exists at all levels of a HRO and which assists it to effectively manage and control organizational risks. These attributes or ‘principles’ include:

a. pre-occupation with failure,
b. reluctance to simplify,
c. sensitivity to operations,
d. commitment to resilience, and
e. deference to expertise [5, 18].

These principles have been elaborated on elsewhere (see, for example [9, 17, 18]), so a brief explanation is provided here. Pre-occupation with failure is the active consideration and ongoing wariness of the possibility of failure that treats any failure or near-miss as indicators of potentially larger problems [12, 17]. Reluctance to simplify involves actively questioning received wisdom and operating assumptions to uncover blind spots [17]. Sensitivity to operations is about creating and maintaining a current, integrated understanding of operations in the moment, while commitment to resilience involves growing and maintaining organizational capabilities to adapt, improvise, and learn in order to recover from unexpected events [17, 18]. And deference to expertise occurs when decisions migrate to those with the greatest expertise with the problem at hand irrespective of formal rank [11, 14, 17]. In more recent years it has been posited that these CM represent behavioral capabilities of 'mindful organizing' (MO) which enable HROs to detect and correct errors before they become a major problem, bounce back from threats and adversities, and adapt to unexpected events [19, 20].

From the perspective of evolving safety management, CM represents an advanced strategy for managing organizational safety [5, 7], so is similar to safety culture in some respects [16, 21, 22]. Early studies [11–13, 15] have alluded to this, while recent authors e.g. [19] have investigated resilient safety culture through CM. Such studies generally measure safety climate, or the shared perceptions regarding safety policies, practices, procedures that an organization expects, rewards and supports with regard to safety [23]. A number of indicative factors depending on the industrial context are used for this; in mining these include things such as management values, safety communication, safety training and safety systems [24]; or management compliance, commitment and caring; supervisor compliance, commitment and caring, workgroup climate, personal involvement in health and safety, job risks, control and pressure, and attitudes towards safety [25]. Safety performance is generally measured through workers' compliance and participation in safety, and the initiating and enabling role of managers in promoting safety and compliance with procedures. However, most of these studies failed to take into account the interactions between the different factors, participants or informants; or the links between those interactions and the achievement of safety. Social interactions are important in negotiating and achieving safety outcomes in HROs [13], and cultivating group level interactions is important in driving mindful practices [26, 27]. In addition, safety climate has been suggested to be an antecedent for CM and can be used to measure mindful organizing as a predictor of safety outcomes in the workplace [20].

3 Research Gaps and Research Questions

While safety culture and safety climate have been embraced by many mine safety managers, professionals and policy advisors; the same cannot be said about CM. In this regard it remains underexplored as an organizational safety strategy in mining. This a significant gap which the authors believe can hinder progress in achieving sustained safety improvements in the industry. A central research question that can be asked is how can CM and mindful organizing be used to influence safety culture in the Australian mining industry? This can be broken up into a series of sub-questions such as:

a. How have CM and MO been conceptualized in the published literature?
b. Which factors and indicators of CM and MO can be used to influence group and organizational-level safety climate?

Answering some of these questions will be useful in developing an understanding of where a team, department, or organization is in terms of CM [5] and where organizational safety improvements can be best targeted.

Answering the above research question(s) empirically requires one to use an appropriate conceptual and theoretical framework to collect, analyze and present the data. The next section presents a pragmatic framework that can be used for this.

4 A Pragmatist Framework for Investigating CM and MO

Pragmatism is one theoretical framework (apart from positivism and interpretivism) which has been used selectively to inform some fields of research and practice, but which is still yet to be embraced in organizational safety management [28]. Some steps are being taken to address this anomaly [29, 30]. In contrast to the positivism and interpretivism, pragmatism holds the view that the research question that needs to be answered is more important than either the philosophical stance or the methods that support that stance. For pragmatists, knowledge claims arise out of actions, situations and consequences [31, 32], with the value of any knowledge generated depending more on the methods by which it is obtained, instead of one's view of whether truth and reality are discovered or constructed [29, 30]. Early proponents of this tradition of research assumed that the paradigm attributes are logically independent so provided a great opportunity to mix and match data collection approaches to achieve the most effective research. Current proponents e.g. Korte and Mercurio [33] have suggested it is useful in bridging the theory-practice research in fields such as human resource development. From a research and practice perspective, pragmatism is more flexible than positivisms and constructionism because it provides a greater freedom of choice of methods to shed light on the research problem at hand [30, 32]. For this reason it allows for the use of a range of methods to answer practical research questions. Getting an understanding factors which can influence CM and MO in mining organizations from a safety culture perspective will be useful in answering practical research questions around the utility of these as an organizational safety in mining. The authors strongly believe pragmatism is a useful framework which can be used to advance this understanding through empirical research.

5 Organizational Theories

Conducting an empirical research also requires the research to be embedded in an appropriate theory. In the authors views CM, MO safety culture and safety climate involves some level of inquiry into organizational behaviour, more specifically those aspects of organizational behaviour which are associated with the achievement of organizational safety outcomes. Two useful theories based on systems and social construction of safety have been previously suggested for investigating advanced approaches to safety [3, 28]. Both of these provide a good platform for investigating CM, MO and high-reliability concepts in mining.

5.1 Systems and Socio-Technical Systems Theory

Systems theory was first floated by researchers in organizational behaviour and management science, with early researchers classifying organizations into closed and open types. Closed organisations are mechanistic and characterised by high specialisation, rigid departmentalisation, narrow spans of control, high formalisation, limited information network with limited participation in decision-making by employees. Open systems, on the other hand, are not only complex but also more organic, highly adaptive and flexible. A number of theorists have suggested most organizations are complex adaptive system [34, 35]. Earlier on it was argued that mining is complex, so systems theory can be applied for research in this industry.

Extending on the systems thesis, authors such as Rasmussen and Svedung [36] also posited that organizations also operated in a dynamic environment, and risk management in such organizations required a consideration of the socio-technical arrangements amidst which those risks needed to be managed. According to this socio-technical theory (STS) the broader socio-technical system that organizations operated in comprised of several levels; including government, regulators and associations, company, management, staff and work [36]. Each of these played different roles in the system, and either influenced or were influenced by, other levels of the system in question.

5.1.1 Socio-Technical System of Mining in Australia

Extending on the theories above, mining organizations can be suggested to be part of a of broader STS which comprises of at least seven levels, illustrated in Fig. 2. The first three are part of the external environment and context, while the next three are internal to the organisations. The first level includes the government which sets the broad national safety policy based on the political aspirations of the party elected. The next level include state and territory the regulators who translate the government's aspirations into safety law and enforce this in industry. The third level includes Associations of employers and unions, and at is at this level that legal requirements are translated into advisory documents and made available to members. The fourth level includes the mining organization who oversee exploration, mining and production; set broad policies and frameworks for works, operations, and safety; and where senior managers translate the advisory documents into organizational policies, procedures and rules. The fifth level is represented by managers who work with supervisors are generally responsible for establishing and meeting targets for production and safety. They work hand in hand with safety personnel such as managers, coordinators and advisors in implementing broad-level organizational controls. The sixth level includes supervisors; at this level they could play two distinct roles; as a manager for either one specific contract or a number of mining projects, and it is here that they implement policies, procedures and controls stipulated by the organization. The other is as a worker at the seventh level where they themselves are expected to follow policies, procedures.

Each level is subjected to pressures and stressors from both above and below, so outcomes such as safety, reliability and mindfulness will be the either arise out of, or be influenced by, the interactions between the different levels.

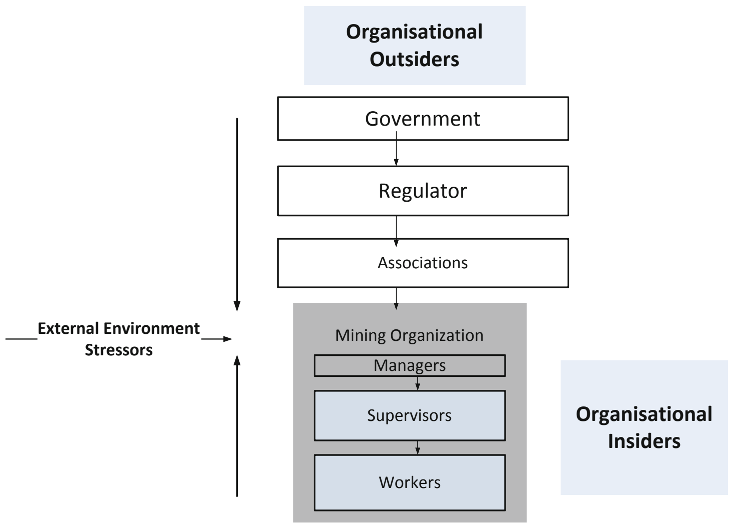


Fig. 2. The socio-technical system of mining in Australia, adapted from Pillay and Borys [37].

5.2 Social Construction of Safety

According to the new thinking, safety is a dynamic property that emerges out of the interactions different elements and subunits of a socio-technical system [38]. This suggests safety is a social construct, a process of discovery [39] and is consistent with HRO theory proposed by Rochlin [13]. It is also in tandem with organisational learning theorists such as Gherardi and Nicolini, who posited that safety was an organisational competence arising from a constellation of interconnected practices, was socially constructed, innovated and transmitted to new members [40]. Cook, O'Connor, Render and Woods [41] extended this argument by suggesting that safety was an emergent (instead of a fixed) state of a system as people continuously created and re-created safety by managing risks as they arose in the day-to-day conduct of work.

6 Concluding Remarks

While mining continues to be a major contributor to the Australian economy, it continues to be regarded as a poor performer in as far as safety is concerned. As we have argued here; mining is complex, existing strategies for managing safety in the industry have failed to drive sustainable improvements, and more innovations are needed. We than introduced HRO and CM as an innovation and opportunity the Australian mining community needs to embrace going forward. We also proposed a series of research questions as a starting point for the journey towards HRO and CM, a pragmatist framework and two theories upon which this research can be developed further. The authors are currently utilising the STS framework illustrated in Fig. 2 to conduct pilot and full-scale studies to benchmark HRO, CM and MO in the Australian mining sector.

Future papers will include a review of the literature on HRO and CM, development and testing of instruments for investigating these in mining, and findings from the pilot and full scale studies.

References

1. Minerals Council of Australia. http://www.minerals.org.au/corporate/about_the_minerals_industry
2. Australian Bureau of Statistics (ed.): Mining Operations, Australia 2013–2014. Australian Bureau of Statistics, Canberra (2015)
3. Pillay, M., Tuck, M.: Permit-to-work systems as a health and safety risk control strategy in mining: a prospective study in resilience engineering. In: AHFE 2017 International Conference on Human Error, Reliability, Resilience, and Performance, 17–21 July, pp. 145–154. Springer, Heidelberg (2017)
4. Nyoni, W., Pillay, M., Jefferies, M.J., Rubin, M.: Organizational factors, residual risk management and accident causation in the mining industry: a systematic literature review. In: Arezes, P. (ed.) Advances in Safety Management and Human Factors, vol. 791, pp. 14–23. Springer, Cham (2018)
5. Pillay, M.: Disasters, accidents and safety: a review of theory and applications for advancing mine safety management. In: Foster, P. (ed.) 35th International Conference of Safety in Mines Research Institutes (ICSMRI 2013), pp. 367–376. IOM Communications Ltd., London (2013)
6. Cliff, D.: Ensuring health and safety through the entire mining cycle. AUSIMM Bull. **3**, 68–71 (2016)
7. Pillay, M.: Accident causation, prevention and safety management: a review of the state-of-the-art. Procedia Manuf. **3**, 1838–1845 (2015)
8. Komljenovic, D., George, L., Mustafa, K.: Organization: a new focus on mine safety improvement in a complex operational environment. Int. J. Min. Sci. Technol. **27**, 617–625 (2017)
9. Hopkins, A.: Learning From High Reliability Organisations. CCH Australia Limited, Sydney (2009)
10. Rochlin, G.I., La Porte, T.R.: The self-designing high-reliability organization: aircraft carrier flight operations at sea. Naval War Coll. Rev. **40**, 76–90 (1987)
11. Roberts, K.H.: Some characteristics of one type of high reliability organization. Organ. Sci. **1**, 160–176 (1990)
12. La Porte, T.R.: A Strawman speaks up: comments on the *limits of safet*y. J. Contingencies Crisis Manag. **2**, 207–212 (1994)
13. Rochlin, G.I.: Safe operation as a social construct. Ergonomics **42**, 1549–1560 (1999)
14. Roberts, K.H.: Managing high reliability organizations. Calif. Manag. Rev. **32**, 101–113 (1990)
15. Weick, K.E.: Organizational culture as a source of high reliability. Calif. Manag. Rev. **29**, 112–127 (1987)
16. Pillay, M., Tuck, M.: Safety culture and high reliability organizations: convergent approaches for smarter mine safety management. In: SME 2019 Annual Conference and Convention. Society for Mining Engineering (2019)

17. Enya, A., Dempsey, S., Pillay, M.: High reliability organisation principles of collective mindfulness: an opportunity to improve construction safety management. In: Azeres, P. (ed.) Advances in Safety Management and Human Factors, vol. 791, pp. 3–13. Springer, Cham (2018)
18. Weick, K.E., Sutcliffe, K.H., Obstfeld, D.: Organizing for high reliability: process of collective mindfulness. In: Boin, A. (ed.) Crisis Management, vol. III, pp. 31–66. Sage, Los Angeles (1999)
19. Klockner, K.: Strategically developing a resilient safety culture: organizational mindfulness and mindful organizing. In: AHFE 2017 International Conference on Safety Management and Human Factors, The Westin Bonaventure Hotel, Los Angeles, California, USA, 17–21 July 2017, pp. 111–121. Springer, Heidelberg (2017)
20. Vogus, T.J.: Mindful organizing: establishing and extending the foundations of highly reliable performance. In: Cameron, K., Spreitzer, G.M. (eds.) The Oxford Handbook of Positive Organizational Scholarship, pp. 664–676. Oxford University Press, Oxford (2011)
21. Reason, J.: Safety paradoxes and safety culture. Int. J. Inj. Control Saf. Promot. **7**, 3–14 (2000)
22. Boin, A., Schulman, P.: Assessing NASA's safety culture: the limits and possibilities of high-reliability theory. Publ. Adm. Rev. **68**, 1050–1061 (2008)
23. Zohar, D.: Safety climate in industrial organizations: theoretical and applied implications. J. Appl. Psychol. **65**, 96–102 (1980)
24. Froko, I.U.-F., Maxwell, A., Kingsley, N.: The impact of safety climate on safety performance in a gold mining company in Ghana. Int. J. Manag. Excellence **5**, 556–566 (2015)
25. Parker, A.W., Tones, M.J., Ritchie, G.E.: Development of a multilevel health and safety climate survey tool within a mining setting. J. Saf. Res. **62**, 173–180 (2017)
26. Aase, K., Skjerve, A.B., Rosness, R.: Why good luck has a reason: mindful practices in offshore oil and gas drilling. In: 6th International Conference on Organizational Learning and Knowledge, pp. 193–208. University of Trento (2005)
27. Dahl, Ø., Kongsvik, T.: Safety climate and mindful safety practices in the oil and gas industry. J. Saf. Res. **64**, 29–36 (2018)
28. Pillay, M.: Exploring resilience engineering through the prescription and practice of safe work method statements in the Victorian Construction industry, pp. 339. School of Health Sciences, Doctor of Philosophy, University of Ballarat, Mount Helen, Australia (2013)
29. Klockner, K., Pillay, M.: Theorizing and theory building in the safety sciences: a reflective inquiry. Saf. Sci. **117**, 250–256 (accepted). https://doi.org/10.1016/j.ssci.2019.04.023
30. Pillay, M.: A pragmatist framework for revising higher education curriculum in occupational health and safety. In: Chova, L.G., Martínez, A.L., Torres, I.C. (eds.) 11th International Conference on Education, Research And Innovation (ICERI), pp. 6763–6770. International Academy of Technology, Education and Development (IATED) Academy, Seville, Spain (2018)
31. Shields, P.M.: Pragmatism as a philosophy of science: a tool for public administration. Res. Publ. Adm. **4**, 195–225 (1998)
32. Creswell, J.W.: Research Design: Qualitative, Quantitative and Mixed Methods Approaches. Sage Publications Inc., Thousand Oaks (2009)
33. Korte, R., Mercurio, Z.A.: Pragmatism and human resource development. Hum. Resour. Dev. Rev. **16**, 60–84 (2017)
34. Schneider, M., Somers, M.: Organizations as complex adaptive systems: implications of complexity theory for leadership research. Leadesrh. Q. **17**, 351–365 (2006)
35. Cowley, S., Borys, D.: Stretching but not too far: understanding adaptive behaviour using a model of oganisational elasticity. J. Health Saf. Res. Pract. **6**, 18–22 (2014)

36. Rasmussen, J., Svedung, I.: Proactive risk Management in a Dynamic Society. Swedish Rescue Services Agency, Karlstand (2000)
37. Pillay, M., Borys, D.: Towards an exploration of safe work method statements as a safety management strategy in construction: background and theory. In: Seventh International Conference on Construction in the 21st Century (CITC-VII). (2013)
38. Cook, R., Rasmussen, J.: "Going solid": a model of system dynamics and consequences for patient safety. Qual. Saf. Health Care **14**, 130–134 (2005)
39. Wildavsky, A.: Searching for Safety. Transaction Books, New Brunswick (1988)
40. Gherardi, S., Nicolini, D.: Learning in a constellation of interconnected practices. J. Manag. Stud. **39**, 419–436 (2002)
41. Cook, R.I., O'Connor, M., Render, M.L., Woods, D.D.: Operating at the sharp end: the Human factors of complex technical work and its implication for patient safety. In: Manuel, B.M., Nora, P.F. (eds.) Surgical Patient Safety: Essential Information for Surgeons in Today's Environment, pp. 19–30. American College of Surgeons, Chicago (2004)

Management Tool for Reliability Analysis in Socio-Technical Systems - A Case Study

Salvador Ávila Filho[1(✉)], Cassio Brunoro Ahumada[2], Érica Cayres[1], Adriano Puglia Lima[1], Cesar Malpica[3], Camille Peres[2], and Edmara Drigo[1]

[1] Polytechnic Institute, Federal University of Bahia, Rua Professor Aristides Novis, 2, Salvador 40210630, Brazil
avilasal@ufba.br, emcayre@gmail.com, edmarasd@gmail.com, adripugli@hotmail.com

[2] Mary Kay O'Connor Process Safety Center, Texas A&M University, Bizzell St. 400, College Station, TX 77843, USA
cassioahumada@gmail.com, peres@tamu.edu

[3] PDVSA Petroleo SA, Av. Professor Libertador, 169, Caracas 1010-A, Venezuela
cme2710@gmail.com

Abstract. The objective of this work is the development of a management model for system reliability analysis influenced by technical factors and human factors (Socio Technical Reliability Analyzes - STRA). Such analysis is made from the understanding of technology, operational and organizational contexts and the determination of human reliability, equipment reliability, operational reliability and process known as integrated reliability or sociotechnical reliability. With this management tool, equipment and process failures as well as human failures will be reduced, and the result will be increased reliability and operational availability.

Keywords: Sociotechnical reliability · Human reliability · Reliability mapping · Decision tools

1 Introduction

The accelerated process of modernization of industry over the past 20 years has increased the complexity of production systems. The industrial automation provided by this type of modernization has brought several advantages among them to reduction of the human intervention [1] guaranteeing a faster and optimized response of the system.

Such complex systems in case of consecutive faults in the barriers or in case of unusual events can cause accidents of high impact for the Society. Regardless of technological advances, failures always occur, even if the best operating and maintenance practices are adopted, why? Due to Systemic failures, that are associated with human factors, organizational factors, routine practices, and process technology constraints need to be investigated.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 13–25, 2020.
https://doi.org/10.1007/978-3-030-20037-4_2

Due to the inevitable nature of the failure, it is necessary to develop a management tool capable of helping managers make decisions that reduce the occurrence of failures, increasing the reliability and availability of the system. The objective of this work is to develop and apply the necessary procedures to compare reliability standards making possible the control of the systems that will result in the increase of sociotechnical reliability. The Socio Technical Reliability Analyzes (STRA) is carried out from the operational context, human characteristics and technological restrictions. The aim is to increase the competitiveness of the production system by reducing failure and increasing reliability. The STRA tool is being developed considering the technology, operational and organizational context, levels of operational complexity and reliability in industrial systems.

2 STRA Tool

The systemic failure is a complex subject and need new concepts to be discussed and validated. Avila [2–4] tried to include this issue in Conferences of human reliability, Courses and Seminars to Brazilian Industries. In the same time, the research group GROD in was concentrated in real tests done by researches, master study and published in papers with the application of SPARH [5] and STRA. To understand this topic is important become clear some concerns: the importance of adapt culture to apply a project that include reliability criteria; the discussion about systemic failure; and particularities from human reliability tests and social-technical.

2.1 Culture of Reliability and Project

The research on the systemic fault in Reliability Culture intends to: elaborate a scenario that demands solutions; investigate the influence of Blame Culture with underreporting on failure and how to migrate from Blame Culture to Just Culture through adjustments in task and behavior; apply dynamic risk management system with practical tools and build projection of future events; apply tools that make possible to measure the quality of task at the workplace; analyze the fault and map the reliability to direct resources to control critical regions of sociotechnical system.

Current scenario analysis demands a high level of reliability due to the serious consequences of climatic phenomena, social and geopolitical. Emerging climate and social changes are causing events that were considered rare to begin to appear. In order to understand and program actions to avoid these events it is necessary to break the paradigms and open the “pandora” box as the analysis of human factors and culture implications. The need is great for quick actions due to the greater possibility of occurrence of top events, such as accidents and even disasters with many fatalities.

Some concepts are important to include reliability criteria in venture projects. The analysis of components configuration in the equipment or production system will facilitate achieving the goal of reliability in the new processes, critical equipment and products. The Models for Reliability are built from past plant histories and equipment failure database, that are updated with current fault results. These models serve as a guide to initiate failure studies with techniques such as FMEA, FTA [6, 7] or root cause study for systemic failure, that includes human and organizational factors (FMEAH, FTA with Human factors) [6, 7].

Increased product technological complexity and external competition require engineering concepts to investigate the failure and set the level of acceptable reliability for products and systems. The expected performance of the system depends on the specification established in the design and performed during the operation of the equipment. If there is a difference between the project and operation, it cannot guarantee performance in terms of failure rate. Thus, the probability of a product or system performing its function, as designed, and for a certain period, in a given environment depends on the maintenance of the design specifications during the operation.

The Project Criteria for reliability are: (1) design simplification of components (human or technical); (2) redundancy in safety barriers; (3) over-specification to keep the reliability target; (4) safety barrier to prevent failure where the design is prepared to avoid failure and process losses; (5) maintainability where equipment's should be designed with parts that can be conveniently replaced, effectively maintenance; (6) and good project concepts (tolerance, operating conditions, size of systems, & spare parts).

The Fault can be related to human behavior in a social-organizational environment, so we need tools that try to avoid the recurrences: standard task analysis [8] and failure diagram assessment (logic diagram, connectivity, chronology and materialization) [4]; risk analysis including human factors [8, 9]; risk analysis of stress influencing on operational routine decisions [7]; risk analysis of progressive stress on health [11]; analysis of cognitive quality and bad habits using the tools executive function [12] and archetypes [13].

2.2 Systemic Failure

The emergence and accelerated growth of new technologies has increased the level of complexity in production systems. According to Perrow [14], complexity is associated with productive systems that have multiple interconnections.

Despite the reduction of human labor in industries, automation and complexity present in the most modern production mechanisms do not have 100% reliability. The occurrence of failures in complex systems represents financial losses and, in some cases, when there are catastrophes, human losses with bad reputation, leading to the collapse of the company. It is necessary to investigate the reasons for a failing system and, through the identification of root cause, to propose a model to reduce these faults [15].

Failure analysis is multidisciplinary [16] and is widely used to investigate accidents, since failure indicates the fragilities of the system. More specifically, the application of this analysis has considerable relevance in the areas of operation and maintenance, especially in high risk systems such as petrochemical [15].

There are a number of causes that lead to system failure: poor system operation, inadequate maintenance, design failure, assembly failure, manufacturing failure, and human failures. Studies show that 60% to 90% of complex system failures are attributed to human actions in the work context [15], including errors in the human-machine interface, omission, incorrect decisions, violation and bad habits in routine.

Failure analysis should not only generate corrective actions and measures. In addition to training and review of protocols, the analyst should investigate the reason, the root cause of failure [15] that is directly and indirectly related, and carry out preventive actions. In order to do that, it is necessary to consider failure from a technical-organizational point of view and from a sociocultural and human point of view [16].

2.3 SPARH, Human and Organizational Factors

The discussion on some Human and Organizational Factors (FTO) must be carried out throughout this text in the presentation of the characteristics about the work environment, the organizational culture and reliability culture, the job position and the risk of accidents-crises resulting from complex decisions. FTO is discussed to develop an algorithm for calculating human and sociotechnical reliability and then to present new concepts and methods in the analysis of characteristics common to human and organizational factors [18]. The FTOs are discussed in documents that analyze human error to enable the calculation of their reliability. The variability of FTOs will be discussed from aspects of organizational change [19], worker life cycle and materials life cycle, type of technology, type of operation, linguistic aspects and organization credibility.

Meanwhile, the API770 [20] is a document linked to the American oil industry which aims to discuss human errors to guide preventive and corrective management actions. The main objective of this document is to present concepts about human errors, stress factors and human performance factors, also exemplifies cases of the chemical industry that cause human error and respective losses. The discussion presents common situations that induce error, also called human factors.

The Nuclear Industry has a high risk of occurrence of impacting events in society despite the redundancy of critical systems and despite the low frequency. In fact, after the Fukoshima accident [21], it is confirmed that the barriers built for rare events are not enough to contain the hazard, that is, scientists predict the approach of the black-swan [22]. Blackman et al. [5] worked to quantify human reliability in nu-clear industry tasks by creating SPARH. Although it is quite useful there must be confirmations for the calibration of the human factor each class weights.

The authors of the papers that discuss human factors [5, 20, 23] consider that managerial and organizational aspects are included in the discussion about human factors. The discussion on Bayesian Network to calculate the probability of human error [24] works by differentiating the human factors, the managerial factors and organizational factors.

According to Blackman et al. [5], SPARH is a standardized method for risk analysis and human reliability where an initial probability of human error is corrected with the following characteristics: distinction between task of action and diagnostic; time as an influencing factor redefined for low power and plant shutdown events; refined dependence; uncertainty calculation methods and others.

2.4 Sociotechnical Reliability (STRA) and Reliability Mapping (FPSO)

One concern that seeks the processes and products improvement in direction of organizational excellence has been the reliability of systems that is related with: system or component fault; work in a specific period of time; operating conditions without failure occurrence [25]. Reliability tool assess the failure modes of an equipment or system to predict and reduce failure rate [16].

The equipment reliability should not be analyzed alone but together with other sectors of a productive system for greater efficiency and better results. Then, reliability of each sector or function is interconnected and, depends on the level of automation, type of process, management style and organizational environment [3].

After understanding the relationships between the functions and sectors of these systems, it is possible to study integrated reliability and map reliability throughout the system. This knowledge allows the formatting of factors that increase team and equipment performance leading to greater organizational efficiency. The success of integrating reliability is guaranteed and achieved through the adoption of a new organizational standards and new management styles [3].

The objective of integrated reliability is to facilitate the development of this integrated tool to analyze failures, accidents, tasks and hazards, including aspects of operation, process, maintenance and human factors. Integrated reliability depends on the level of complexity, high, average and low of production systems [4].

To calculate socio-technical reliability it is necessary to calculate human reliability first. The analysis of human reliability is the study of human errors, applied in critical security systems [26]. Human reliability is defined as the probability of a worker failing to fulfill a task in a given period of time, in appropriate environmental conditions and with resources available for execution [27]. Just as equipment and systems are flawed, so is man. However, these failures are measurable, through predictive methods that include a variety of tools, as SPARH [5].

Some applications of human reliability and social technical are presented in Table 1 [3, 28–34].

Table 1. Applications and calculations of HR, ER and CSST in industry

	Activity	Cmpx	Ref.	Feature	EqR%	HR%	CSST %
1	HR in the Chicken Industry	Medium	[28]	Family influencing work	Giro Freez MTBF 251	98,9	NA
2	EqR and HR in Nuclear Industry	High	[29]	High risk of health and leakage-discredit	Crusher 4: 53,4	2ª est.: 85,6	NA
3	Pumping propene refining	High	[3, 30]	Lack of commitment of the organization	Pump 92	84	86
4	HR Chemical industry HCl	High	[31]	Loading HCl and environment	Pump 72	71	81
5	HR, CSST platform, pre-salt, oil	Medium	[32]	Social Relations and Rituals	Operation NA	NA	NA
6	HR Release order product Manuf.	Low	[33]	Service being analyzed, releases order	Operation 99	8	NA
7	HR Coconut Water Industry	Low	[34]	Failed turn in panel	NA	98,5	NA

HR - Human Reliability; ER – Equipment Reliability; CSST - Socio-Technical Reliability System; NA - Not applicable or not calculated; Cmpx - complexity; Ref – Reference.

An exercise demonstrating the use of the STRA tool was carried out in a set containing twelve ships type FPSO that produce oil. Through meetings between specialists in the field of oil exploration, platform manager and researchers in the area. An important data related to the operational context was obtained where shutdown occurrence of the FPSO vessel has a cyclic period of 1 year. With this and other data from the meetings, it was possible to carry out the mapping of the reliability of the system. The reliability mapping of the FPSO vessel system will help identify the root cause, impact and complexity of failures, and quantify reliability parameters by optimizing resources to the maintenance [9]. It was detected that the system reliability is 17% and the greatest occurrence region of failure was the control panel with ST Reliability of 52%.

2.5 Organizational Culture, Operational and Social Attractiveness

Some situations and factors are discussed in the literature due to the possibility of causing human error and accident [5, 20, 23]. We can cite the relationship between the following situations involving culture (organizational, security, regional and global) with human errors: conflicting priorities between security and production; inadequate communication or feedback; conflict between policy and practice; violation of population stereotype; standardization of deviation; cultural flexibility; style in team coordination; organizational change and external environmental factors resulting from social phenomena. These human-organizational factors define the level of social attractiveness that must be estimated in this tool to correct the calculated the Sociotechnical reliability.

3 Methodology

The methodology of the study is divided into seven sequential steps. (1) Step one corresponds to the identification of the operational context from expert survey; (2) with the results collected via questionnaire, a block diagram is constructed by mapping all functions, processes, and activities that have significant influence on the overall reliability; (3) after the blocks are defined, we perform a system complexity analysis; (4) subsequently, reliability calculations are conducted for each individual block including human, equipment, operational and process reliability; (5) results from previous steps are then combined to estimate an integrated system reliability; (6) the next step of the methodology would be the validation based on the economy and real field data; (7) the final stage is to draw a visual reliability block diagram with suggestion and possible intervention plans (Fig. 1).

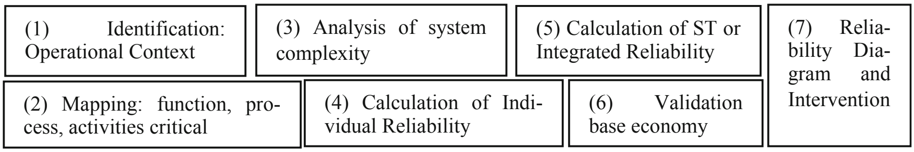


Fig. 1. Methodology.

4 Case Study and Discussions

The case study shows the application of the methodology for a refining unit with hydroprocessing and hydrocracking technology. Refineries are continually being challenged for cleaner technologies that are highly efficient and cheaper [35]. Changes in crude oil market prices have forced refineries to reevaluate their process options by targeting investments in hydrocracking with catalytic systems. These low-cost changes are related to operational flexibility with increased return on investment.

When it comes to reliability analysis of refineries, very often the technical aspects are predominant in this type of studies [36]. However, although it is of extreme importance, companies usually fail in integrating social, human factors, and organizational aspects in conjunction with equipment reliability studies and task analysis. The authors acknowledge the complexity of such task and propose a method to include social and technical elements into reliability block diagrams. The purpose of this case study is to explain how this methodology can be applied. At the current stage, due to time constraint, real plant data were collected only for the operational context identification phase (stage 1). The data utilized for the subsequent steps were estimated based on expert knowledge and internal communication with refinery specialists.

4.1 Identification of the Operational Context

At the initial stage, a survey was applied to the technical manager of an oil refinery referring to the operational context, technology, and human issues (see Table 2).

Table 2. Illustrative questions of the expert survey

What kind of production, capacity, and automation?	What's the product and raw material?
Main characteristics of culture and management?	What's technology?
How many steps does hydrocracking start?	How do you rate communication on shift?
In your opinion in this plan are variables and actions visible or hidden?	

From the survey, we classified the facility under study as a medium-sized oil refinery located in Latin America. The facility has a total production capacity of 232000 bpd, and its operation started in 2004. The production process is continuous with 8-hours shifts and 10 days of scheduled shutdown per year.

The refining production process focuses mostly on obtaining light materials, without sulfur content. It has hydrocracking technology, highly automated and operates with closed critical grids. Regarding management and culture aspects, the plant experiences underreporting and centralization of information in a light degree. Variables and actions in this type of industry are difficult to visualize and analyze.

The collected information served as a basis for the context formulation, additional data used in this case study was obtained from expert knowledge and internal communication with collaborators.

4.2 Mapping of Principal Functions, Activities, and Processes

The second stage of the research is the mapping of the main functions/activities/processes that affect production. Therefore, they cause a plant shutdown or decrease productivity to a point approaching a shutdown. To obtain this information, the following questioning was sent to the specialist, which led to the completion of the table.

According to the characteristic of the treatment of deviations and failures in the routine, please define the 10 main functions/processes/activities that affect the sociotechnical reliability in its installation for the construction of the diagram in Table 3.

Table 3. Structure of the system reliability diagram including human factor (only >10%)

Subject	%	Functions/processes/activities	% of influence
Maintenance	30	Compressors, pumps & valves, pressure vessels & pipe flange	35% valves & pumps; 35% compressors; 8% pressure vessel; 7% pipe flange
Process	30	Coker, hydrocracking, atmospheric distillation	20% Distillation; 15% Coker; 40% hydrocracking
Utilities & Effluents	20	Water quality, water-oil separation, steam generation, solids, emissions	20% Water quality; 18% Steam generation; Water-oil separation 15%; 9% Solids; Emissions 7%
Managerial & Culture	15	Decision, underreporting	20% Decision; 9% Underreporting
Logistics	5	Transportation of materials and spare parts	30% Transportation of materials and spare parts

Pareto was used as criterion for the cut line, subject x function, with 10 being the maximum of possible blocks, those with influence greater than 10% were chosen. In this way, it was possible to map the main functions/activities/processes that impact the industry in question. The study led to the diagram resulting from this article where the equipment-process-functions chosen for the analysis are indicated in diagram (◊).

The analysis of impact of the failure on the system will define the type of reliability to be studied in each system that makes up the block diagram. So, if the main problem (over 60% of impacts) is the decision of control panel operators, it is a question of human reliability. If the main problem relates to critical equipment and material failure, the main problem lies in the reliability of equipment and structures. When the main problem lies in the planning or execution of the task, referring to poorly drawn procedure, it is operational reliability. And finally, when there is no main problem, and process, equipment, man, and operation influence the reliability of the system, it is considered that the result is reliability of a sociotechnical system. It is possible to identify in the reliability diagram, choice of the type to be analyzed (CSST, HR, EqR, OpR, PrR).

4.3 Complexity of the System

In order to proceed with the study, it is necessary to calculate the complexity of the system. According to [2, 14, 37] complexity may be associated with aspects that cause noise in the mental map, impairing decision making in the task, it may also be related to the number of recycle lines [14] or communication difficulty in social relations [17].

According to the exploratory study, the complexity was calculated. Considering (1) level of automation with closed loop. How many (2) major equipment relative to the auxiliaries, which increase the complexity of the task. Besides this, the (3) level of attention required in the task that depends on the selected workers. The (4) recycle lines of the process that reduce the prediction of the quality of the chains, (5) the accomplishment of the task, which does not occur exactly like the procedure. After this analysis, it was verified that the level of complexity is high.

4.4 Calculation of Individual Reliabilities

To know the reliability of each block, a specific questionnaire was used that takes into account human factors, equipment, operation and process. Enabling integrated reliability calculation. For the construction of the questionnaire SPAR-H, API770 was used, among other texts and methods as basis. The questionnaire for the calculation of individual reliabilities is described in Table 4.

Table 4. Survey for calculating individual reliabilities (some questions)

Kind	Description
Complexity	What is the level of automation? How much equipment influences complexity? What is the level of attention in the tasks? What is the influence of the recycle lines? What is the level of noncompliance with the procedure?
Process R	What are the 3 variables that cause the loss of product quality, performance reduction, partial stops for safety, quality or environmental impact?
Human R	Is the time available for the task adequate? What level is stress? What is the level of complexity? What is the level of experience and training of employees? Are the procedures available and consistent with the task? Is ergonomics suitable for the job? Are the employees fit for the job? Is the level of work process adequate?
Equipment R	Relate the equipment that causes stoppage or loss of severe performance in industrial plants and which statistical model is applied? To estimate reliability through the exponential function [24]
Op R	Use complexity definition assumptions for the operational reliability calculation

Based on the context description and the activities classification, critical functions/processes, we analyze which reliability dimension causes more than 60% of equivalent failures. According to the equations referenced in Table 5, the individual reliability was calculated, allowing the calculation of the reliability of the refining unit. Then obtaining the numbers to compose the reliability diagram as described in Table 6.

Table 5. Equations for the calculation of reliability (*individual, integrated*)

LC = Cmpx = (1) * (2) * (3) * (4) * (5) * 100000	CSST = $(PrR)^2$ * EqR * (HR * Op R) ^1/2	HR = NHEP * PSFc/NHEP * (PSFc −1) + 1
OpR = (100-(LOG(LC)/4) *100))	PrR = (R/ΔL)	

Table 6. Results of individual reliabilities

Kind	Block	%
Human Reliab.	Decision;	99
Operational Rel.	Coker; Logistics	98; 99,5
Equipment Rel	Compressor of H2	98
Sociotechnical System Reliab.	Water quality; Steam generation; Water-oil separation; Fractional distillation; Hydrocracking; Pumps -valves;	99; 98; 99; 99; 99; 98

4.5 Refinery Reliability Calculation

It is understood as integrated reliability, the analysis of reliability of the whole system [29]. The reliability of the system is calculated taking into account factors of complexity and social attractiveness as described in Table 7.

Table 7. Refinery reliability calculation

Refinery SST Reliability = ,99 *,99 *,98 *,99 *,99 *,98 *,98 *,98 *,99 *,995 = 0,873	
Social Attractiveness = 1,0 *,873 = ,873	λ = 8 shutdown unscheduled for year
2,5 day for shutdown, 20 days of shutdown every 355 days, Availability = 97%	

4.6 Validation Base Economy

In the case of the industry in question, the article presents an exploratory research, and it is not possible to validate it at the present time. At this moment, the study and compilation of the tool was done, bringing a study of the application of the tool with the experience of the specialists.

4.7 Result and Suggestions

As a result of the study we obtained the block diagram with the indication of the reliability map of the system, which allows the manager to make more assertive decisions in the improvement of his system and decrease the variability of the process.

The blocks chosen to make up the diagram are those that imply a loss of production similar to a unit shutdown or real shutdown. The second part is given by calculating the reliability of each block. Finally, the analysis of the reliability of the whole system, taking us to the image below.

◊ Indicates the mapping of functions; * Indicates the type of reliability of each block

○ Indicates the individual reliability calculation (Fig. 2).

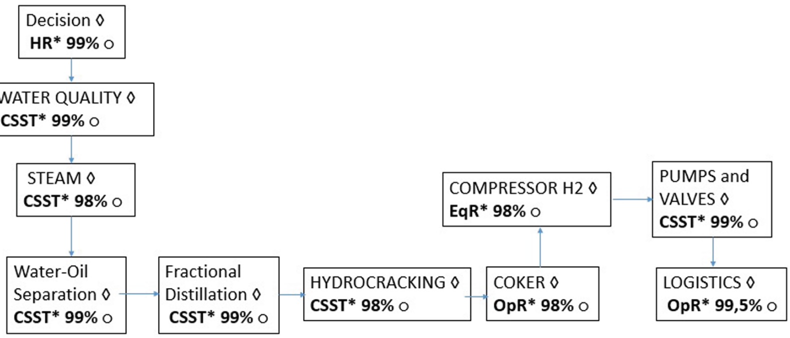


Fig. 2. Reliability block diagram of refinery

This method tries to approximate the priorities to carry out the interventions, but has limitations on the subjective issues related to the culture installed in the routine of the operation. The latent movements that can cause failure or accident deserve specific analysis including the field of forces of culture [38] and the analysis of the operator's speech [39].

5 Conclusion

The STRA tool was developed considering human, operational, process and equipment reliabilities. This implies in a more complex tool, that take into account technical factors and also observes human factors, culture and organization. It is necessary continue the validation the tool in real case to suggest interventions and to study the influence of these in improving production.

This tool indicates the direction of probable cause to establish barriers, but does not indicate the root cause of the problems. This means that the actions carried out need to be adjusted through other techniques that take into account the operational culture and its biases, being necessary to apply the analysis of the operator's discourse to increase reliability in the sociotechnical system.

References

1. Ogle, R., Morrison, D.T., Carpenter, A.: The relationship between automation complexity and operator error. J. Hazard. Mater. **159**(1), 135–141 (2008)
2. Ávila, S., Dias, C.: Reliability research to design barriers of sociotechnical failure. In: Annual European Safety and Reliability Conference (ESREL), Portorož (2017)
3. Ávila, S., Santos, D.: Integrated reliability: failure assessment to process, operation, equipment and human integration (POEHint), a case of propene pump operation. In: Sixth International Conference on Engineering Failure Analysis, Lisbon (2014)
4. Ávila, S.F.: Complex Process Failure Analysis. In: Brazilian Congress of Chemical Engineering. Rio de Janeiro (2012)
5. Boring, R., Blackman, H.: The Origins of SPAR-H Method's Performance Shaping Factors Multipliers. Idaho – EUA. Agosto (2007)
6. Ávila, S., Ferreira, J., Sousa, C., Kalid, R.: Dynamics operational risk management in organizational design, the challenge for sustainability. In: 12th Global Congress on Process Safety, Houston (2016)
7. Ávila, F.S.: Dependent layer of operation decision analyzes (LODA) to calculate human factor, a simulated case with PLG event. In: Proceedings of 7th Global Congress on Process Safety – GCPS, Chicago (2011)
8. Ávila, S., Sousa, C.R., Carvalho, A.C.: Assessment of complexity in the task to define safeguards against dynamic risks. In: 6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, Las Vegas (2015)
9. Ávila, S.F., Bittencourt, E.: Operation Risk Assessment in the Oil & Gas industry, Dynamic Tools to Treat Human and Organizational Factors, Chapter 1. Nova Science Publishers. New York (2017)
10. Ávila, S.F., Pessoa, F.L.P., Andrade, J.C.S.: Social HAZOP at an oil refinery. Process Saf. Prog. **32**, 17–21 (2013)
11. Ávila, S.F.: LESHA – multi-layer progressive stress & impact assessment on health & behavior. In: Proceedings of 9th Global Congress on Process Safety, San Antonio (2013)
12. Ávila, S., Costa C.: Analysis of cognitive deficit in routine task, as a strategy to reduce accidents and increase industrial production. In: European Safety and Reliability Conference (2016)
13. Ávila, S.F., Menezes, M.L.A.: Influence of local archetypes on the operability & usability of instruments in control rooms. Industrial production. In: Proceedings of European Safety and Reliability Conference – ESREL. Zurich (2015)
14. Perrow, C.: Normal Accidents: Living with High-Risk Technologies. Basic Books, New York (1984)
15. Ferreira, M.G.: Analysis of the reliability of the CO2 compression system. In: A Petrochemical Plant Considering The Technical-Operational And Human Factors. Masters dissertation. Salvador (2018)
16. Ávila, S.F., Dias, C.: Reliability research to design barriers of sociotechnical failure. In: Annual European Safety and Reliability Conference (2017)
17. Ávila, S.F.: Failure analysis in complex processes. In: Proceedings of 19th Brazilian Chemical Engineering Congress, Búzios (2012)
18. Cerqueira, I., Drigo, E., Ávila, S., Gagliano, M.: C4t safe behavior performance tool. In: International Conference on Applied Human Factors and Ergonomics. Springer, Cham (2018)

19. CCPS Guidelines for Managing Process Safety Risks during Organizational Change. Wiley, New Jersey (2013)
20. Lorenzo, D.K: API 770 – A manager's guide to reducing human errors: improving human performance in the process industries. API Publishing Services, Washington (2001)
21. Aoki, M., Rothwell, G.: Coordination under uncertain conditions: an analysis of the Fukushima catastrophe. ADBI Working Paper Series. No. 316. Tokyo, October 2011
22. Taleb, N.N.: The Black Swan: The Impact of the Highly Improbable, 2nd edn. Penguin, London (2010)
23. Crowl, D.A.: Human Factors Methods for Improving Performance in the Process Industries. Wiley, Hoboken (2007)
24. Ashrafi, M., Anzabi Zadeh, S.: Lifecycle risk assessment of a technological system using dynamic Bayesian networks. Qual. Reliab. Eng. Int. **33**(8), 2497–2520 (2017)
25. Lafraia., J.R.: Manual of reliability, maintainability: and availability (2001)
26. Boring, R.L.: Fifty years of THERP and human reliability analysis. In: Probabilistic Safety Assessment and Management, Helsinki [S.l.]: IAPSAM & amp; ESRA, vol. 11 (2012)
27. Pallerosi, C.A.: Human reliability: new qualitative and quantitative analysis methodology. In: International Reliability Symposium, Florianópolis. Electronic Annals…: ARS, Florianopolis, vol. 42, p. 31 (2008)
28. Dórea, S.S., Ávila, S.F.: Failure investigation in operations of chicken industry with based human and technical factors. In: Proceedings of Brazilian Association of Risk – ABRISCO/PSAM, Rio de Janeiro (2015)
29. Avila, S., Alcoforado, A.C.: Application of SPAR-H in the ore crushing operation. Brazilian Association for Risk Analysis, Process Safety and Reliability (ABRISCO), PSAM Topical Meeting on Safety and Reliability of Oil and Gas Exploration and Production Activities (2015)
30. Ávila, S.F.: Reliability analysis for socio-technical system, case propene pumping & quot. J. Eng. Fail. Anal. **56**, 177–184 (2015)
31. Ávila, S., Souza, M.L., Ramalho, G., Passos, I.C., Oliveira, A.R.: Analysis of the reliability model for failure mode interdependent, a discussion of human error. In: Organizational and Environmental Factors National Congress of Mechanical Engineering, Salvador (2018)
32. Ávila, S.F.: Risks in the platform: social and human guidelines avoid accidents and alter standards in oil production, Revista Proteção (2013)
33. Fonseca, M., Carvalho, P., Ávila, S., Kalid, R.: Integrated reliability analysis in task of emission production orders in ERP system: a case study on a company's sports segment. In: European Safety and Reliability Conference, Glasgow (2016)
34. Sá, G.: Analysis of human reliability in sociotechnical systems: a case of agroindustry. Completion of Course Work, Salvador (2019)
35. Brossard, D., Koldachenko, N., Maesen, T., Torchia, D., Yoon, H.A.: Hydroprocessing to maximize refinery profitability. In: Proceedings of American Fuel & Petrochemical Manufactors, Annual Meetings (2012)
36. Nakamanuruck, I., Rungreunganun, V., Talabgaew, S.: Reliability analysis for refinery plants. KMUTNB Int. J. Appl. Sci. Technol. **10**(1), 61–70 (2017)
37. Ávila, S.F.: Review of risk analysis and accident on the routine operations in the oil industry. In: 5o LCPS – 5TH Latin America conference on process safety, Cartagena (2013)
38. Ávila, S.F., Sousa, C.R.O., Silva, C.C.: Worker behavior changing by dynamic culture. In: 9o Global Congress in Process Safety GCPS, San Antonio (2013)
39. Drigo, E.: Analysis of the operator's speech and its instrument of communication between shifts as a tool for the management system. Dissertation (2016)

Formerly Unrevealed Public Records Should Change the Account of What Occurred on June 30, 2013

Fred J. Schoeffler[1](✉), Lance Honda[2], and Joy A. Collura[3]

[1] Sheff LLC, Pine, USA
dougfir777@yahoo.com
[2] Prineville, USA
[3] Yarnell Hill Fire Revelations, Congress, USA

Abstract. On 30 June 2013, nineteen men, the Prescott Fire Department (PFD), Granite Mountain Hot Shot Crew (GMHS), quasi-military Wildland Fire Crew, died on the Yarnell Hill Fire in Arizona. Little has been done to dispel the September 2013 Serious Accident Investigation Team's (SAIT) Serious Accident Investigation Report (SAIR) no fault conclusions. This infamous event defies reason. Newly revealed Public Records and Freedom of Information Act (FOIA) Requests evidence indicate the YH Fire "story" is inaccurate. These newly discovered records suggest (1) was there a rogue firing operation; (2) disparate treatment regarding Public Records; and (3) Agency employee "guidance" and direction to not discuss the YH Fire contradict professed "lessons learned" objectives. Attempts by so-called "leaders" to deter WFs from discussing and truth-seeking of this tragedy abound. The "wildfire fatalities are unavoidable" argument is contentious. The authors maintain that wildfire fatalities can be reduced notwithstanding that many naysayers continue to fervently disagree.

Keywords: Yarnell Hill Fire · Wildfire · Hot Shot · Human Factors · AFUE

1 Introduction

As noted in previous papers, the continuing need for research on the Yarnell Hill (YH) Fire fatalities as well as this paper's focus on newly revealed public records evidence, combined with causal Human Factors and Human Errors, requires that some sections of the current paper may reiterate previous work germane to the topics [1, 2]. Additional information and analyses will be included as follows: (1) was there a rogue firing (burning out) operation in the Sesame Street and Shrine Fuel-Fire Break Corridor, referred to in the SAIT-SAIR as the "the two-track road (an old fuel break) between Sesame Street and Shrine Road … preparing for burnout along the dozer line" [3]; (2) disparate treatment regarding Aerial Firefighting Use and Effectiveness (AFUE) Public Records Requests, and (3) USDA Office of General Counsel (OGC) and the USDA USFS employee "direction/guidance" *limitation* from discussing the YH Fire is the opposite of stated internal/external alleged "lessons learned" as evidenced in several email threads.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 26–39, 2020.
https://doi.org/10.1007/978-3-030-20037-4_3

WFs will continue to struggle over making sense of the June 30, 2013, YH Fire; the "how and why" of the 19 Granite Mountain Hot Shot (GMHS) fatalities. The newly revealed Public Records evidences strongly suggest some plausible "how" and "why" prongs of this enigma. Our goals are to analyze *why* so many men died needlessly and *why* key evidences are concealed that are germane to the ongoing issue of professed "lessons learned" yet their USFS employees are restricted in talking about the tragedy.

1.1 Background

Wildland firefighting is recognized legally and logically as "inherently dangerous" [4]. Countless individual, organizational, cultural, and systems goals, are known to gradually lead the unwary drifting into failure. Wildland fire fatalities are inevitabilities based on the interactions of complex environmental, social, cultural, and individual behaviors. The DIVS A and GMHS decisions were flawed because of these complexities and their own limitations [2]. Was there a rogue firing operation below them in the Sesame Street and Shrine Fuel-Fire Break Corridor? Were they (un)aware of this because they had no lookout or because those doing the firing never informed anyone? Or both? Plus, DIVS A made a major tactical move with his only resources, the GMHS, however, he never informed his supervisor as required. Moreover, the GMHS choice to vacate their Safety Zone (SZ) at the worst possible time without posting a lookout, nor advising Air Attack of their intentions; a clear sign of the normalization of deviance, drifting into failure; elements in a long chain of bad decisions with a fatal entrapment outcome [1, 2, 5, 6].

Confusion why the GMHS were individually and then collectively fixated with abandoning their adequate SZ to (un)willingly reengage and commit themselves at the worst possible time persists [2]. Were they aware of a possible firing operation below them? These factors may have impeded their decision-making: (1) cultural influences of municipal fire departments with collateral wildland fire responsibilities to "save structures" at all costs (herd tunnel vision); (2) known, obvious unsafe GMHS leadership concerns; (3) weak Crew cohesion that relied on consensus-seeking ("absence of leadership"); and (4) marginal to dubious situational awareness of impending danger [1, 2].

1.2 Motivation and Goals

The authors allege that the YH Fire SAIR was deceptively labeled as "Factual" by the Serious Accident Investigation Team (SAIT-SAIR) [1–3]. A notable Human Factors expert, Dr. Ted Putnam wrote: "Generally[,] the goal of accident reports is to convey as much of the truth of an event that is discoverable. … Sometimes investigators deliberately distort or do not report all the causal elements. Such biases lead firefighters to distrust the resulting reports, which can hamper our efforts to stay safe" [7]. The authors will persistently seek the truth in the matter despite SAIT-SAIR "factual conclusions."

"Wildland firefighting is a high-risk occupation, evidenced each year by deaths or injuries in the line of duty … [we need to] identify factors responsible for past fatalities, [to] mitigate those factors in future fire seasons" [8]. Wildfire fatalities are regrettably unavoidable. The authors have vowed to exert due diligence to reduce wildfire deaths. Mistakes often occur through the faulty actions of many and it is mostly the accrual of many smaller errors that finally result in the bigger error [5, 9]. So it was for the GMHS.

Emphases are placed on the following key areas: (1) was there a firing operation; (2) Wildland Firefighting Rules and Human Failures; (3) concealing and failing to release <u>all</u> of the YH Fire AFUE Study and other records; (4) USDA Office of General Counsel (OGC) and USFS "guidance" and "direction" limitations to their employees regarding the YH Fire; and (5) Conclusions. The authors allege USDA dishonesty re: the AFUE records, and deceitful employee YH Fire "direction/guidance," resulting in what Vaughan refers to as "incomplete lessons learned" [5]. Forced USDA silence upon their employees barred the SAIT and ADOSH from exploring the vital causal human factors.

2 The Sesame Street and Shrine Fuel/Fire Break Corridor Firing Operation

At least twenty (20) people, including experienced WFs and FFs and the two YH Fire hikers, Tex Gilligan and Joy A. Collura watched a video in July 2013 at the Yarnell, AZ Library of the Sesame Street and Shrine Corridor firing operation. It involved "two men wearing Nomex and using drip torches … above The Shrine … along a visible rock wall." This was also viewed by one FF on YouTube before it was removed. The video abruptly vanished without a trace, like other YH Fire evidence. There were also burnt "fusees" (firing devices akin to large road flares) found on a 2014 site visit and a GMHS family member found "accelerants" with specialized dogs another time [10].

Consider now an enhanced Google Earth image of the Sesame Street/Shrine Fuel/Fire Break Corridor looking northwest, at parallel chutes aligned upslope on June 30, 2013. Did the firing operation-fire behavior funnel into the GMHS deployment zone?

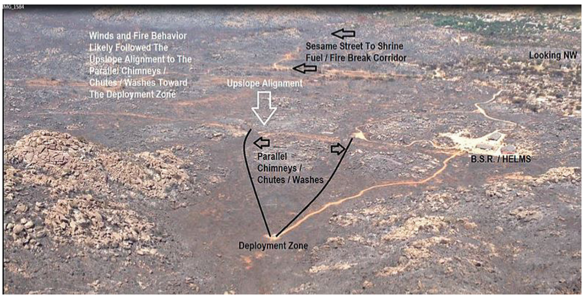


Fig. 1. Enhanced snippet image using Paint, aligned northwest, denoting the Sesame St. and Shrine Fuel/Fire Break Corridor with probable firing operation and upslope alignment with parallel chutes into the GMHS deployment zone. Source: Google Earth and Joy A. Collura.

Past *YFD Fire Chief Peter Andersen* (deceased) noted: “we built an emergency escape route for Yarnell *in case there was a burnout like this* … in that area below The Shrine, west of The Shrine, they had dozers back there widening that so that it would create a fire break, … too little too late” [11]. (https://youtu.be/UFObh-fNOl8) Then consider the ABC15 Crossfade of ABC15 Helicopter Video clip no. 18 of the YH Fire (WTKTT). Published 3/24/15 (https://youtu.be/NSYpnPMfPmc), June 30, 2013, and parallel ‘Google Earth’ imagery. ‘Orange firelines’ imagery represents fire visible in the video footage and not the totality of the ABC News 15 fire video [12]. Opening and viewing all videos herein as (“https”) hyperlinks will certainly elucidate.

Fig. 2. Jerry Thompson photo IMG_1898 on June 30, 2013 at 1624 h. looking NNW; two separate/distinct smoke columns (plumes) suggests a firing operation. The Sesame/Shrine Corridor (left); Youth Camp (middle); Harper Canyon (right) Source: InvestigativeMEDIA

Fig. 3. a & b (left) June 30, 2013, 1629 h. Both photos w/Google Earth overlay with GMHS action points. Source: B. Lauber & WantsToKnowTheTruth. (right) June 30, 2013, 1631 h. This is the Sesame St. and Shrine Corridor area. Source: ABC News15 & WTKTT

Fig. 4. a & b (left) June 30, 2013, 1348 h., Yavapai County Morin dozer building fireline in The Shrine area. Source: Anonymous By Request FF. (right) June 30, 2013, 1437 h. Peeples Valley FD Water Tender and unknown Municipal FD Type 6 Engine in The Shrine area. The BRHS are to the left in the trees out of view. Source: Anonymous By Request FF.

Fig. 5. a & b (left) Enhanced view of clearly visible, very active fire behavior in smoke column (plume) on June 30, 2013, 1544 h. beyond Lakewood Dr. noted on the sign in Glen Ilah. Source: Yavapai County Records request 7/19/14 A30-20306 T. Carrigan photos (IMG_4254); (right) Photo of two GMHS Crew Carriers in clearing in lower left on June 30, 2013, 1544 h. Note the SAIT photo caption along the bottom, regarding specific details on times, posted as a SAIR exhibit but was never used or mentioned in the SAIR. Source: SAIT & Joy A. Collura

3 Wildland Firefighting Rules and Human Failures

All WFs are trained in specific rules, crucial to follow to ensure good direction, leadership, safety, and vigilance. The Standard Firefighting Orders, organized purposely and sequentially, are to be carried out sensibly on all wildfires [1–3]. The 18 Watch Out Situations (i.e. guidelines), are faced on all fires, more to warn of impending dangers. The authors and experienced WFs contend that knowing and abiding by the wildland firefighting rules works. They encourage sound leadership and safe decisions [13, 14]. There are no documented wildfire fatalities when the Standard Firefighting Orders are followed and the cautionary 18 Watch Out Situations ("10 & 18") are mitigated [15].

The most critical of the established Wildland Fighting Rules are listed in the (NWCG) Incident Response Pocket Guide (IRPG) [16]: (1) *Ten Standard Firefighting*

Orders; (2) *Eighteen Watch Out Situations*; (3) *Lookouts - Communications - Escape Routes – Safety Zones (LCES)*; *Downhill Checklist*; (5) *Common Denominators of Fire Behavior on Tragedy/Near-Miss Fires*; and (6) *Wildland Urban Interface (WUI) Watch Outs*. "If firefighters follow the Standard Firefighting Orders and are alerted to the 18 Watch Out Situations, much of the risk of firefighting can be reduced" [17]. Jointly they save tens of thousands of wildland firefighter lives each-and-every fire season [1–3]. It is well-known and accepted in the WF community that these Wildland Firefighting (WF) Rules work when applied consistently and with sound, decisive judgement [1–3, 6, 13, 14].

"The National Interagency Fire Center (NIFC) urges investigators to *withhold some findings from the public,* and to *avoid analyzing whether crews violated fundamental fire-line rules*," reports The AZ Republic [18]. (emphasis added) Why would NIFC suggest that? It gets worse. There are even high level criticisms by some alleged "leaders" through the WFLLC that the WF Rules "cannot work … not going to keep us safe" and we need to rely on "luck decision conversations" [15, 19]. The authors and most WFs and FFs agree on the validity of the WF Rules and discount the latter stance regarding "luck decisions." According to these WF human factors researchers, "Downhill fireline construction is hazardous in steep terrain, fast-burning fuels, or rapidly changing weather" [20]. The present authors state that these listed major *environmental* causal factors required precautionary decisions by the GMHS on June 30, 2013, as well as decision-makers on most fatality wildfires that failed to include them in their intentions. Abiding by and heeding these defensive WF Rules was missing in each of them [15].

3.1 Individual Blame and Organizational Fault Logics

The authors will resume following Catino's Individual Blame Logic (IBL) and Organizational Fault Logic (OFL) to explain the probable, yet puzzling, causal human failures that resulted on tragic wildland fires [1, 2, 21]. The IBL suits public reaction to identify accident cause(s) and transgressor(s) while OFL is an organizational and functional tactic, aimed at finding the factors within the system behind the occurrences. In the OFL method, expectations are that similar events will not recur or will rarely occur once these influences are removed [21]. The authors allege both IBL and OFL logics were present regarding the GMHS from 2009 to 2013 [1, 2, 6]. These logics apply to *all* WF fatality fires with like human errors stressing that *all WF supervisors are responsible for the safety and welfare of those they supervise - regardless of the situation*!

3.2 Willingness to Properly Refuse Risk and Turn Down Protocol

In the quasi-military wildland fire realm, one is to obey orders unless unsafe, illegal, unethical, or immoral [1, 2, 6]. Paraphrasing the related IRPG section [16]: a WF is obligated to identify alternative(s) to unsafe assignment(s) and refuse them if needed and must be told by a supervisor that others have refused it. This protocol is vital to effective risk management. Most times, resources are relegated to "Division Siberia," somewhere already put out, and finally returned to their home unit with a poor rating. As expected, many Resources fear using this crucial protocol and would likely engage contrary to safer judgement, notably *those that want to be everyone's friend* instead of being "The Boss" fulfilling their supervisory duties. An 'old school' USFS Ranger gave sound advice when he often quipped that "Boss spelled backwards is double S-O-B."

3.3 Plan Continuation and Confirmation Biases

Plan continuation bias, a very powerful, unconscious cognitive preference, impeded the GMHS' ability to recognize they needed to change their course of action. It was '*52 minutes from the blowup to the burnover,*' (link) [22] yet they failed to notice obvious cues indicating that conditions were *exponentially* (link) changing. Avalanche fatality scholars note people generally have a strong bias for sticking with what they have now and let their minds default to what is given or what has already been decided. They rely on stay-the-course impulses most times, often with deceptively satisfactory results [23].

3.4 Steady Drift into Failure via Bad Decisions with Prior Good Outcomes

Several instances of GMHS hazardous attitudes and actions support their drift into failure addressed in 2016 [6, 9]. It spanned from their first official Hot Shot status season in 2009 with a repeated attitude of having to "prove themselves" or "one-up" other HS Crews on fires, due to their Municipal FD status, up until the YH Fire in 2013 [6].

On the 2012 Holloway Fire in NV, their Crew Carriers, visibly parked in the unburned minus drivers, were saved by an OR Contract Engine Crew in this VIMEO video, long-removed from the website (http://vimeo.com/48411010) [24]. From 2:41 to 3:05, the nozzleman pans left, so notice a handline on the hillside. Freeze-frame the video and you will notice a GMHS black helmeted WF (lookout?) running downhill. This was the second of three times another Crew had to "save" their Crew carriers [6].

At the 2016 AZ Wildfire Academy, a former GMHS (2011–2013) recounted to the primary author that they had a similar Holloway Fire near miss. Initially line spiked and told by the IMT to be "fire-ready" even when sleeping. The Acting Supt. *posted no lookout* and woke up to a distant glow, ignored it and went back to sleep. They awoke hours later with the fire upon them. They quickly fired out around themselves [6, 25].

Author Kyle Dickman reported that one of the reasons GMHS Brandon Bunch applied for a transfer to another Crew prior to the 2013 season is because he was sick and tired of the GMHS Supt. always acting like he had 'something to prove.' "The more seasons Bunch worked for the [GMHS], the more he felt that under Marsh's command, the Hotshots were always having to prove themselves" [26] (p. 54).

It appears that the GMHS *normalized* deviance from the start of being a Hot Shot Crew as they steadily drifted into failure right up to the end on June 30, 2013 [1, 2, 5].

This was posted on InvestigativeMEDIA by "WFF wife" on 8-11-16: "As the wife of a WFF, there were so many times of obviously disregarding safety protocol. Even McDonough mentioned it, and there are documented cases of this being their nirmal. [*sic*] Yeah, they probably got away with it a lot of times, but that suffering will catch up to you, the safety precautions, fire orders and watch out situations are there for a reason" [27]. This has strong validity, especially if "WFF wife" was a GMHS "wife."

3.5 Consequences of Inattention as Causal Factor

Credible research on (in)attention indicates that when someone is otherwise engaged, at times they fail to "see" otherwise noticeable, fully visible, yet unexpected objects or events, (i.e. 'inattentional blindness') (IB). IB also leads one to miss items that one *needed* to experience [28, 29]. If an event meets their expectation(s), then they may be more likely to exhibit IB for a sudden, possible critical visual event [22, 28, 29]. IB is very likely what occurred as they, *minus a lookout*, hiked downhill into the uphill fire behavior from the probable firing operation (Figs. 1, 2, 3, 4 and 5). The YH Fire - 2013 - SAIR Fig. 19 (WTKTT) video clarifies fire activity (https://youtu.be/Jl1l8EyDric) [10].

4 GMHS and USFS Aerial Firefighting Use and Effectiveness (AFUE) and Other Germane Yarnel Hill Fire Records

On or about July 1, 2013, AFUE Team Leader Panebaker placed YH Fire AFUE records on a 'hard drive' and gave it to a 'SAIT team member.' In July 2013, ABC NEWS Investigative Producer James Meek learned of an 'Aerial Firefighting Study Group' in Yarnell on June 30, 2013, 'recording data' and filed numerous FOIA Requests [30] In 2015, the primary author first became aware that the USFS utilized an AFUE Study on the YH Fire on June 30, 2013, and from a colleague, that he had a 3-ring binder of Air To Ground (A/G) radio transmission transcripts. The primary author's AFUE FOIA Requests and eventual lawsuit were based on that insight [31]. Unexpectedly, the colleague turned Quisling, altered his declaration after Agency coercion, and was swiftly 'rewarded' with a promotion and a transfer - a common USFS tradition. Co-author Collura filed a 2015 FOIA Request for AFUE records and got AFUE email records in 2016.

The YH Fire was basically an AZ State Forestry wildfire, however, the USFS had the USFS Blue Ridge Hot Shot (BRHS) and other personnel assigned to the fire plus the USFS AFUE Study, during which extensive audio and video recordings were made of aircraft operations. Many Air-to-Ground (A/G) transmissions with WFs on the fire-lines, were audible in the background in varying degrees of conception, contrary to the SAIT-SAIR conclusion of a "gap of over 30 min" [3]. With few definitive, in-depth information sources on the obscure AFUE, the Wildland Fire Safety Training Annual Refresher (WFTAR) website admits the AFUE is sponsored by the USFS [32].

4.1 USDA Office of General Counsel and USDA Forest Service Email Threads Regarding the Existence of USFS YH Fire June 30, 2013, AFUE Records

In an August 20, 2013, email essential new information was revealed from the USFS Deputy Fire Director to the USFS Fire Director, George Vargas of the Office of

Regulatory and Mgmt. Services (ORMS), and other OGC employees that the AFUE audio and video records *do in fact exist*. "Benny, *George Vargas has custody of the disc with the video/audio files for the WO. He is cced.*" (emphasis added) OGC Counsel "Benny" Young wrote to USFS Fire Director: "… I understand from our [Albq.] office that *a AFRUE (?) [sic] was flying at the time of the incident ... and capture[d] audio and video of a portion of the tragedy. ... now in the hands of our office. ... please make sure nothing happens to those tapes. Also, please have some copies carefully made for preservation purposes*" (emphasis added) [33]. Copies?

Astounding! They divulged to Co-author Collura these AFUE records existed and yet denied they existed to the primary author's FOIA Request. A presumption of "good faith" legal precedent given to Federal Agencies means that they <u>should</u> readily admit to retaining and releasing records, however, that was not the case, suggesting deception.

4.2 Unaccounted for YH Fire and GMHS Evidence

An AZ Dept. of Public Safety Officer and a Yavapai County Sheriff's Deputy took late day June 30, 2013, 'aerial photos' of the fatality site with two 'cellphones' for still shots and videos. SAIT Human Factors Investigator Brad Mayhew and other SAIT members were given a photo disk from YCSO. These photos/videos have always been known to be in the possession of the SAIT. *Unethically*, none of them were granted in response to records' requests for this evidence, nor did they supply them to ADOSH [30, 33].

At least four of the GMHS GPS units they *always* carried, plus the one Caldwell wore on his pack strap, seen in numerous photos/videos, have been unaccounted for, foiling any chance of GPS track evidence [30]. Nor has the "checking on your comfort level" *Mackenzie video - <u>altered</u> from 29.76 s to two-nine-second videoclips*; less than 30-s from the end of the first video and the start of the second one; *exactly 16:01:40.24 to 16:02:10.00* [30]. Nobody takes nine-second video clips on a fire.

GMHS Christopher Mackenzie's still-working Canon digital camera vanished from the Maricopa County Medical Examiner's office (along with other things) and never officially entered into the YCSO evidence chain. It was discovered by a family member when delivered to them by PFD. *Crucial photos, videos, and audio* ended up on a CD that his father handed back to PFD Wildland BC Darrell Willis at Mackenzie's funeral service [30]. Who finally received that critical auditory evidence and where is it now?

Former Peeples Valley FD FF (now Chief) Brandon wrote of his missing June 30, 2013, records: "… on [July 2, 2013] they [SAIT] … debriefed us. This was two days after the incident with the 19. They went through and looked at my pictures and took a flash card of them. And I don't know what they did with them, … because my pictures had timelines on them, so they could see what happened at what time. Then they took information off our cellphones … And those had timelines on them, too" [33].

It appears the unaccounted for evidentiary records are *calculated* versus *incompetence.*

5 USDA USFS and OGC Employee "Direction" and "Guidance"

5.1 The Touhy Regulations and the Federal Housekeeping Statute

Consider briefly the USFS/OGC emails on their USFS employee YH Fire direction (Collura FOIA 2016-FS-R3-04243-F). USDA OGC Attorney Hattenbach wrote to several Southwest USFS Fire and OGC employees ("FS Employee Interviews in Re: [YH] Fire, ECM#7805646"). "*… [ADOSH]'s requests for interviews should be treated as Touhy requests,* handled in accordance [to] 7 C.F.R. 1.214." (emphasis added) [33]

Now consider the "*Touhy Regulations*." Prior to obtaining testimony from Government employees, one must first overcome the regulations, a daunting barrier to obtaining depositions and pretrial discovery. It derives from the U.S. Supreme Court's decision in U.S. ex rel. Touhy v. Regan, 340 U.S. 462 (1951) (citations omitted) [34].

In Touhy, the Supreme Court reversed a contempt order against an FBI agent who had defied a deposition subpoena against a Department of Justice regulation issued under the Federal Housekeeping Statute, 5 U.S.C. § 301 [33], which allows agencies to adopt rules about *"the conduct of [their] employees ... and the custody, use, and preservation of [agency] records, papers, and property"* [35]. (emphasis added)

The Jackson-Rosenfield LLP law firm concluded: 'Within the last few decades, *federal courts have reined in this expansive reading of Touhy* and clarified that, while Touhy regulations may empower a Federal Agency head to decide whether the agency will comply or resist a subpoena, [the Agency head may oppose the subpoena] with a *governmental privilege or the rules of evidence or procedure. Federal appellate courts are split ... over [subpoena] noncompliance decisions* …; some ruling the actions should be reviewed per the [FRCP], and others that it should be judged under the [APA]'s "arbitrary and capricious" standard, … [33, 36]. (emphasis added)

Consider now the "*Federal Housekeeping Statute*" 5 U.S. Code § 301 - Departmental regulations: "The head of an Executive department or military department *may prescribe regulations for the government of his department, the conduct of its employees, the distribution and performance of its business, and the custody, use, and preservation of its records, papers, and property.* This section does not authorize withholding information from the public or limiting the availability of records to the public." (emphasis added) This 61-word law means that the Federal Agencies have *plenary authority to do whatever they wish with their employees, their business, and everything to do with records and property.* Thus, Touhy and the Housekeeping Statute, ensured that the <u>*USFS employees would never be interviewed*</u>.

5.2 USDA USFS Blue Ridge Hot Shot (BRHS) Crew Emails

Consider an April 2016 email (a bit redacted) from BRHS Supt. Brian Frisby to USFS National Human Dimensions Specialist Joseph Harris regarding the YH Fire Staff Ride. BRHS Supt. Frisby noted that "*the [YH Fire Staff Ride] picture being painted is <u>very different than what we remember</u>" and that "there was so much that went on that day that [was] <u>swept under the rug</u>"* and "*the <u>human factors</u> that day were <u>off the chart</u>.*" (emphasis added) [33]. <u>*No*</u> BRHS has ever been interviewed or deposed, even by

ADOSH, ostensibly to "*protect them.*" BRHS Supt. Frisby and the 2013 BRHS want to and need to share what they experienced. Their recollections of June 30, 2013, would ensure true versus the SAIT's "incomplete" lessons learned. The BRHS "recollections and discussion" accounts are a must read - filled with many revealing details on tactics, strategy, and human factors [33]. *See the 10-15-18 post, Fig. 10 email image.*

5.3 2013–2014 Agency Direction/Guidance to Not Talk About the YH Fire and 2014 Weather Channel Special Report and YH Fire Contrary Lessons Learned

The USFS and OGC email threads and discussions about how they treat their USDA USFS employees when it comes to prohibiting them from interviews about the many "lessons learned" of an epic wildland fire tragedy are disturbing and hypocritical. Initially, they approve of their employees sharing YH Fire tragedy "lessons learned," yet later emails specify restrictions to constrict them from debating or talking about it [33].

And then there is the March 2014 Weather Channel special report on the YH Fire featuring Britt Rosso, the Wildland Fire Lessons Learned Center (WLFLLC) Center Manager. "… I am here today to talk to you about the Yarnell Fire … we are all struggling with how to process what happened on June 30th, 2013. *Know we're all struggling out in the fire community about where the lessons, what are the take-home messages, what can we learn from this incident. ... [it is] important ... to talk about it; ... it's okay to talk about it and it's important that you do talk about it. Share what you've learned by reading the reports, by watching the videos and have an open, honest, respectful dialogue. Be willing to listen to other's opinions and have that respectful dialogue with your fellow firefighters ... about Yarnell. This is where the learning's going to happen, is with you and your brothers and sisters out there in the field. ... We'll be learning about the Yarnell Incident for years to come. ... just take the time and be patient and work through this together*" [37]. Well intentioned, but oblivious to "the memo" to NOT discuss the YH Fire, because he is encouraging us - pleading with us - to talk about it for the benefit of "lessons learned" Maybe it's because they are Park Service and not USFS - different Agencies.

As you read the many USDA OGC and USFS "Direction" and "Guidance" email threads, and listen to this WFSTAR Weather Channel video and Rosso's message, you will understand the nuanced hypocrisy. The email contents are totally contrary to what he discussed regarding "sharing" and "lessons learned." Other emails were equivocal [33]. The share-and-talk-about-it-but-don't-share-and-talk-about-it irony is more than mere haziness or paradox. This is better known academically as the accomplished practice of Orwellian Doublespeak and Doublethink [2, 38] as was the NIFC direction [18].

6 Conclusion and Recommendations

The authors made a good faith attempt to share new information regarding some of the events that occurred on June 30, 2013, Yarnell Hill Fire and the Granite Mountain Hot Shot Crew. New evidence is revealed almost daily. There are an untold number of WFs, FFs, citizens; and family, friends, and loved ones of these same individuals that possess a lot of valuable information about the YH Fire and the GMHS in the form of photos, videos, and narratives that need to be brought forward and shared. From these valuable insights, lacking in the SAIT-SAIR, will guide true lessons learned toward reducing wildfire fatalities when WFs/FFs know the real human factors hidden within this epic tragedy. Surely, one of them will be to routinely know, understand, and abide by the WF Rules versus the contrary USFS's "incomplete" alleged "lessons learned" ostensibly responsible for a culture and system fraught with arrogance since Mann Gulch [7].

The authors revealed numerous photographs of separate and distinct smoke columns (plumes) that - from numerous veteran WFs' perspectives - clearly suggest a firing (burn-out) operation occurred in the Sesame Street and The Shrine Fuel/Fire Break Corridor as the former Yarnell Fire Chief stated the area was intended for. The authors feel confident that this firing operation occurred. Integrity and dedicated resolve to assess and present the numerous YH Fire records will finally reveal its veiled truths.

Paradoxically, Dr. Putnam's paper on wildfire fatality investigations and reviews can be found few places on the Internet, (i.e. coauthor Collura's site and Academia.edu) [7]. The WFLLC refuses to publish it despite more than a dozen attempts by Dr. Putnam and the primary author to convince them contrary to their official mission statement: "*Our mission is to promote learning in the wildland fire service by providing useful and relevant products and services that help to reveal the complexity and risk in the wildland fire environment*" [15]. (emphasis added) One would think that a "Lessons Learned Center" would leap at a chance to publish a research paper on a topic as valuable as "Accidents, accident guides, stories and truth" [7]. Likewise, they fail to *publicly* admit that the primary author's sixty-plus donated deployment and fatality investigations initiated the Incident Reviews section in 2002.

Wildfire fatalities *continue* to occur from the <u>same</u> causal factors. Staff Rides are a valuable asset in the "lessons learned" tool box to reduce them, however, when based on deceptive "investigations," how valuable are those "lessons learned?" An overlooked statement: "[they] should avoid being a recital of a single investigation report. Such reports *rarely address the human factors that affect individual decision-making*. … providing participants *with a variety of information sources* is important" [38]. (emphasis added) The YH Fire *requires* different "information sources" to be factual.

Acknowledgements. The authors wish to express their gratitude to: Dr. Ted Putnam for his Human Factors insight; WantsToKnowTheTruth; the two Yarnell Hill Fire Eyewitness Hikers; sources mandated silence due to 'direction' and 'guidance;' all those troubled by the event; and WFs, FFs, and others that encourage us. To those who lost family, friends, and loved ones, the authors and most WFs and FFs think about those young men every day. They inspire us to research and write toward reducing fatalities.

References

1. Schoeffler, F.J.: Epic human failure on June 30, 2013. In: Boring, R.L. (ed.) Advances in Human Error, Reliability, Resilience, and Performance, vol. 589 (2017)
2. Schoeffler, F.J.: It could not be seen because it could not be believed on June 30, 2013. In: Boring, R.L. (ed.) AHFE 2018, AISC, vol. 778, pp. 231–243 (2018)
3. Yarnell Hill Fire Serious Accident Investigation Team (SAIT): Yarnell Hill Fire Serious Accident Investigation Report (SAIR), 122 p. (2013)
4. U.S. Forest Service: Fire Suppression: Foundational Doctrine, 21 p. (2005)
5. Vaughan, D.: System Effects: On Slippery Slopes, Repeating Negative Patterns, and Learning From Mistake?. Organ. at the Limit. Starbuck & Farjoun; Blackwell, New York (1999)
6. Schoeffler, F.J.: Human factors influenced the 30 June 2013 YH fire fatalities. CAWRT, RT-130 Refresher. PDF, 48 p. (2016). Academia.edu
7. Putnam, T.: Accidents, accident guides, stories, and the truth. In: Proceedings of 11th International Wildland Fire Safety Summit, 4–8 April 2001, International Association of Wildland Fire, MT, USA (2011). https://www.yarnellhillfirerevelations.com/blog. Accessed 15 Oct 2018
8. Mangan, R.: Wildland firefighter fatalities in the United States. 1990–2006. NWCG, PMS 841. Boise, ID. NWCG, Safety and Health Working Team, NIFC, 28 p. (2007)
9. Reason, J.T.: Managing the Risks of Organizational Accidents, vol. 252. Ashgate, Farnham (1997)
10. InvestigativeMEDIA. (https://www.investigativemedia.com/yarnell-hill-fire-2/)
11. Dougherty, J.: Former Yarnell Hill Fire Chief Peter Andersen Interview 8 October 2013 (2014). YouTube: https://youtu.be/UFObh-fNOl8
12. WantsToKnowTheTruth (WTKTT): ABC15-clip-18-crossfade (2015). https://youtu.be/NSYpnPMfPmc
13. Coolidge, H.: Rules and Risk in Wildland Firefighting. Hypocrite Reader, vol. 52 (2015)
14. Holmstrom, M.: Common denominators on tragedy fires - updated for a new (human) fire environment. Wildfire Mag. Int. Assoc. Wildland Fire (2016). (https://www.iawfonline.org/article/common-denominators-tragedy-fires-updated/)
15. Wildland Fire Lessons Learned Center (LLC): - Incident Reviews, Online (2018)
16. NWCG, Incident Response Pocket Guide (IRPG). PMS 461, NFES 001077 (2018)
17. U.S.D.A. Forest Service, Fire and Aviation Management: Risk Management (2017)
18. Chappel, B.: Report On Arizona Hotshots' Deaths Finds a Communications Gap; NPR (28 September 2013; 4:10 PM ET) (NPR article link)
19. Collura, J.A.: Yarnell Hill Fire Revelations.com website (2018–2019)
20. Useem, M., Cook, J., Sutton, L.: Developing leaders for decision making under stress. Acad. Manag. Learn. Educ. **4**, 461–485 (2005)
21. Catino, M.: A review of literature: individual blame vs. organizational function logics in accident analysis. J. Contingencies Crisis Manag. **16**, 53–62 (2008)
22. Wildland Fire Safety Training Annual Refresher (WFSTAR): Blow to Burnover (2018)
23. McCammon, I.: Heuristic traps in recreational avalanche accidents: evidence and implications. Avalanche News **68**, 10 (2004)
24. Drake, C.: We Saved GMHS Buggies (2013). VIMEO: http://vimeo.com/48411010
25. PFD: Personnel Records (2015). Archived: https://web.archive.org/web/20151031112325/. https://www.prescottenews.com/pdfs/GMHS1.pdf
26. Dickman, K.: On the Burning Edge: A Fateful Fire and the Men Who Fought It, 277 p. Ballantine Books. New York (2015)

27. WFF Wife (2016). https://www.investigativemedia.com/families-of-the-fallen-granite-mountain-hotshots-are-not-getting-the-answers-they-need/#comment-344419
28. Krietz, C., Furley, P., Memmert, D., Simmons, D.J.: Inattentional Blindness and Individual Differences in Cognitive Abilities. Pub Med (2015)
29. Snook, S.A.: Friendly Fire: The Accidental Shootdown of U.S. Black Hawks Over Northern Iraq, 257 p. Princeton University Press, Princeton (2002)
30. Dougherty, J.: IM YH Fire I-XXVI; i.e. Meek, ABC, AFUE, FOIA, photo, GPS unit
31. Ross, J.: Firefighter Demands Info on [YH] Fire Disaster Courthouse News (2017)
32. NWCG: RT-130 (2019). https://www.fs.fed.us/managing-land/fire/aviation/afue
33. Yarnell Hill Fire Revelations. https://www.yarnellhillfirerevelations.com/blog
34. US ex rel. Touhy v. Ragen et al.: Cornell Law School LII. Accessed 2019
35. Cornell Law School LII: 5 U.S. Code § 301 - Departmental regulations. Pub. L. 89–554, Sept. 6, 1966, 80 Stat. 379). Accessed 2019
36. Jackson Rosenfield LLP: Dealing with Touhy: Compelling Non-party Deposition Testimony from a Government Agent (2014–2017)
37. Lutz, W.: Beyond Nineteen Eighty-Four: Doublespeak in a Post Orwellian Age. National Council of Teachers of English, Urbana, Illinois (1989)
38. NWCG: Wildland Fire leadership Development Program. The Staff Ride (2019). https://www.fireleadership.gov/toolbox/staffride/index.html

Winning Hearts and Minds in High-Reliability Industries: The Role of Human Performance in Persistent Front-Line Behavior Change

Damien Armenis(✉)

Cortexia, Brisbane, Australia
Damien@Cortexia.com.au

Abstract. Within high-reliability industries, traditional approaches to safety behaviour change using conventional pedagogical methods of teaching often fail at achieving persistent and lasting results on the front-lines. A drastic step change is needed to address outdated practices for a group of industries beset with 'safety fatigue'. By leveraging evidence-based countermeasures and tools developed in the military, sports and cognitive psychology, and other high-performance fields, safety becomes embedded within a self-improvement approach rather than an edictal and prescriptive introduction of procedures. The results of this approach from a number of high-reliability and high-risk companies demonstrate the effectiveness over traditional more prescriptive approaches to front-line behaviour change. This paper clearly demonstrates that by making safety a by-product of high-performance, presenting countermeasures as individual tools within an individual's arsenal, and resonating with the target audience; persistent behaviour change in high-reliability industries can be achieved.

Keywords: Human factors · Human performance · Safety behavior · Safety training · High-reliability · Behavior change · Safety culture · High-performance · Safety performance · Operational performance

1 Introduction

Within high-reliability industries, teams interact with technology and operate in complex environments. Risk varies from low to high with threats coming from a variety of sources in the environment [1]. Safety is paramount in these high-risk industries as incidents can and often do involve loss of life. Organizations that succeed in avoiding disasters in an environment where accidents can be expected due to the high level of risk and complexity are labeled as High Reliability Organizations (HROs). Weick and Sutcliffe [2] describe that HROs share five characteristics in assuring high-performance, which includes a: (1) preoccupation with failure; (2) reluctance to simplify interpretations; (3) sensitivity to operations; (4) commitment to resilience and (5) deference to expertise. Traditional methods of safety have focused on compliance towards company policies and the punishment of unsafe behavior with the organization taking action when accidents or injury occur [3]. In this paper, we argue that traditional approaches to safety behavior change using conventional pedagogical

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 40–49, 2020.
https://doi.org/10.1007/978-3-030-20037-4_4

methods of teaching often fail at achieving persistent and lasting behavior change and instead propose the utilization of Threat and Error Management (TEM) in safety training. In the next section, we explain why traditional safety approaches are less effective and then describe how TEM can be applied in high-reliability industries for the production of persistent behavior change and the high transfer of training for frontline operators.

TEM is widely used in aviation to improve safety and is being adapted for use in medical settings [4, 5]. TEM describes adverse events in terms of risks present in an operational environment (i.e. threats) and the actions of operators that potentiate or exacerbate those threats (i.e. errors) [6]. Hudson, Verschuur, Parker, Lawton, and van der Graaf [7] proffer that there are four major forms of human error: slips (actions not carried out as intended, such as pressing the wrong button), lapses (missed actions and omissions, such as forgetting to lower the undercarriage on landing), mistakes (an error brought about by a faulty plan with the individual believing that they are doing the right thing) and violations (unintentional or deliberate deviations from rules, procedures, instructions and regulations with the breaches in these rules). The University of Texas human factors team [1] classified threats as latent (e.g. national culture, organizational culture, vague policies) or immediate (e.g. environmental factors, individual factors, team factors). According to these researchers, external threats (e.g. adverse weather and aircraft malfunctions) increase the likelihood that operators will make an error [8]. Moreover, common causes of errors include fatigue, workload, poor interpersonal communications and flawed decision making [9]. Given the ubiquity and inevitability of threat and error, the key to safety is the effective management of threats and errors rather than its prevention [10]. This involves training personnel to understand the nature and extent of error, change the conditions that induce error and, determine behaviors that prevent or mitigate error; ‘countermeasures’.

Traditional approaches to improving safety presume that things go wrong because of identifiable failures or malfunctions of specific components: technology, procedures, operators and the organization [3]. While adverse events may be treated by a traditional approach, there are a growing number of cases where this approach will not work and leaves one unaware of how everyday actions achieve safety. Importantly, new methods focus on the encouragement of safe behavior and how performance goes right in spite of uncertainties and ambiguities that pervade complex work situations. That is, the focus has moved from ensuring that as few things as possible go wrong (Safety I approach) to ensuring that as many things as possible go right (Safety II approach) [11]. A Safety II approach incorporates a proactive stance by continuously trying to anticipate events while a Safety I approach is reactive by responding when something happens (e.g. accident) or when an unacceptable risk exists. A Safety II approach attempts to understand the conditions where performance becomes difficult to monitor and control whereas a Safety I approach carries out investigations to identify causes and contributory factors of accidents. TEM is a method that aligns with a Safety II approach as it is a proactive approach that has a focus on self-improvement whereby operators manage operational threats errors with the use of effective countermeasures.

Traditional approaches often do not work in creating lasting safety behavior change [12]. In-line with Vroom’s expectancy-theory of motivation [13], operators weigh up the advantages and disadvantages of various types of work behavior and then choose the behavior with the best outcome. Moreover, the individual knows that by placing more effort towards this behavior will increase the chances of attaining this valued outcome. As an example, an operator may be aware of being fatigued or having an unrealistic workload planned for the day, but chooses to push through due to weighing up the advantages of the behaviour. The operator may desire the esteem of colleagues (e.g. to be seen as a team player), or the esteem of the supervisor (e.g. to not hold up the work) and thus, persevere without relevant countermeasures to deal with the threat at hand. Thus, we argue that traditional approaches to safety training need to be enhanced with a high-performance approach which provides tools for the operator regardless of the choices made.

2 A High-Performance Focus Over Compliance

The interpretation of safety has changed since the 1970s, which brought about an interest for "just cultures" [14]. A just culture aims to balance the need to learn from failure through the reporting of errors, adverse events and accidents with the need to take disciplinary action [15]. Research shows that an act is very rarely carried out with the intention to inflict damage and if an operator perceives that their reports of accidents are treated unfairly, the willingness to report incidents declines [16]. A just culture addresses this paradox of accountability and learning by asserting that operators should not be punished for actions, omissions or decisions taken by them that are in-line with their experience and training but for gross negligence, willful violations and destructive acts [17, 18]. Operators can differentiate between legitimate and illegitimate behavior because each operator is aware of their intentions behind actions and consequences (e.g. did not intend to inflict harm) and the reasons behind any violations of safe operational procedures (e.g. imperfect knowledge, time constraints) [19]. Hence, a strong argument can be made that not all violations of safety procedures should be criminalized [14].

No procedure is perfect, and evidence has shown several problems with safety procedures that result in violations [7]. Often case, procedure problems arise because there is not enough time to perform all the tasks required, checks are seen as unnecessary or operators feel that there is a better method for carrying out a task. Introducing more procedures will not necessarily prevent accidents, nor do exhortations to follow procedures more carefully increase compliance or enhance safety [20]. Dekker [20] highlights the need to redefine safety – it is not the result of rote rule following but the result of individuals' insight into situations that demand certain actions and individuals being skillful at using a variety of resources to accomplish their work goals [21, 22]. Hence, safety is a by-product of high-performing individuals and teams rather than the result of absolute compliance with rules and regulations.

Operators need to be informed of the opportunity for rules to be adapted in certain situations [20]. It is the role of organizations to support operators in becoming skillful at judging when and how to adapt procedures [20]. Hudson and Verschuur [23] conducted a study that revealed that there are two types of operators in each organization in terms of how likely they are to adapt or violate procedures. Operators can be classified as sheep or wolves: sheep do not like violating and feel dissatisfied with their behavior even if they feel compelled by circumstances to bend or break the rules. In contrast, wolves are naturally opportunistic and have no problem with bending or breaking rules if it benefits them, such as getting the job done in record time. However, it is the wolves that often excel in problem solving within unique situations. Organizations are said to require sheep and wolves to ensure that the need for actual violation does not arise and to create conditions in which initiative is productive [23].

3 A Novel Approach to Safety Training

Employers are legally responsible for educating employees on workplace safety standards and the hazards that they may face while on the job through the provision of effective safety training. Trainers need to deliver training in an educational and engaging manner to enhance the retention of information and application of learning to workplaces [24]. Research shows that training is more effective when trainers use language, situations and examples operators will relate to and understand. Training becomes more meaningful and enjoyable to operators when trainers relate the training content to operators' day-to-day work experiences [25]. As a result, operators are more likely to pay attention throughout training and apply what they have learned. Moreover, training activities should be chosen that allow operators to relate their skills and knowledge to work and health safety issues relevant for the industry that they work in. For instance, trainers can incorporate scenarios or real-life situations that link to the concept being learnt whilst being appropriate for the target audience.

There are considerable benefits for trainers to connect and build rapport with operators [26, 27]. One meta-analysis revealed that training utilizing conversation or dialogue with trainees was highly engaging and was approximately three times more effective than the least engaging methods in promoting knowledge and skill acquisition [28]. Researchers assert that safety trainers can connect with trainees by going on ride-alongs before delivering training, as a way of not only observing and understanding the role of trainees, but also become accepted as an outsider who understands. For instance, Herbet [29] conducted an ethnographic field study with police officers to gather their opinions on police brutality. He observed that the police officers were initially reluctant to allow him to go on the ride alongs but then developed rapport and became accepted by the police officers as an outsider who understood their perspective. Herbet [29] described this as a transformation from "spy" to "okay guy" (p. 304).

Another effective training delivery tool is the use of informal storytelling [30, 31]. In an ethnographic field study, incident-reporting schemes were not integrated in railway technicians' practices and did not seem to serve their interests [32]. Hence, the number of reported occupational health and safety incidents was very low, which impeded the usefulness of such schemes. Informal storytelling, however, was found to be the preferred mode for technicians to address risks, with the circulated stories emphasizing attention, vigilance and carefulness. Telling stories of accidents and incidents is advantageous, as it allows for knowledge about recent events to be shared and what one might appropriately learn from them [22, 33, 34]. Storytelling can also extend to moral and emotional dimensions of unfortunate events, which recuperates persons, relationships and communities [35]. Moreover, the use of self-disclosure develops trust and, allows leaders to connect with and influence front-line employees on a deeper level. The most powerful learning rarely come from facts or figures as individuals learn experientially, through oneself and the vivid example of others [36]. Studies show that storytelling is a much more effective way to drive change in attitudes and behavior than increased rules and bureaucracy. Storytelling, however, should not substitute incident-reporting systems and instead, should be used in conjunction with incident-reporting to ensure feedback on root causes.

Operators require undergoing training to develop behaviors that act as countermeasures to common preconditions or conditions leading to human error (decision making and review of plans for example) [1]. The International Civil Aviation Organisation [37] has identified the most common human factors that contribute to human error. The top twelve factors were: (1) poor communication, (2) distraction, (3) lack of resources, (4) stress, (5) complacency, (6) lack of teamwork, (7) time pressures, (8) lack of awareness, (9) lack of knowledge, (10) fatigue, (11) lack of assertiveness, and (12) norms to deviate from safety procedures. We argue that effective countermeasures to these factors for high-reliability and safety-critical industries can be drawn from the tools and techniques that are implemented in other high-performance fields, particularly from sports psychology and the Special Forces. In this section, we describe a number of techniques athletes and military personnel utilize to maintain high-performance. There are a number of mental skills that successful athletes utilize for the long-term development of high-performance, immediate preparation for performance and during actual performance behavior [38], [39]. These mental skills include, the ability to: (1) maintain a positive attitude, (2) maintain a high-level of self-motivation, (3) set high and realistic goals, (4) deal effectively with people, (5) use positive self-talk, (6) use positive mental imagery [40], (7) manage anxiety effectively, (8) manage one's emotions effectively and (9) maintain concentration. These nine mental skills associated with athletic success are the same mental skills associated with performance in a wide variety of non-sport, performance situations [38]. Thus, relevant activities and information can be incorporated into the safety training of operators in high-reliability and safety-critical industries to develop these vital mental skills.

The National Research Council [41] declares that soldiers are required to counteract a range of stressors (e.g. pressure, ambiguity, weather conditions) in their operational environment. Their physical performance is maintained during prolonged periods of physiological and mental stress through eliciting certain behavioral and cognitive skills (maintaining alertness, clarity of thought, and decision-making ability for example) [42, 43]. Soldiers undergo stress inoculation training (SIT) to develop these skills through making information available and pre-exposing them to stressors to reduce the novelty of stressful tasks. This "increases the likelihood of a greater sense of predictability and control, and a consequent reduction in both physiological and emotional reactivity" (p. 258) [44]. In SIT, soldiers are trained to become aware of the stress environment so that when a specific stressor occurs (e.g. exposure to excessive heat) it prompts the individual to prepare (e.g. drink water and dress appropriately). This training method sustains soldier performance before, during and after battle, by combatting stressors that can lead to mental and physiological fatigue [45]. SIT has been adapted for training in organizational contexts (i.e. Stress Exposure Training, SET) and thus, can be used within high-reliability and safety-critical industries [46]. SET involves training operators about the types of stressors that are likely to occur in the operational environment and develop the cognitive and behavioral skills to counteract these threats before a hazardous situation can strike. It is also utilized to increase operators' abilities to maintain high levels of performance under a variety of stressful conditions.

It is incomplete to only train operators on the types of countermeasures from other high-performance fields to effectively manage threats found in the operational environment. It is vital to train operators to use proven countermeasures that are also easy to remember and apply in the field. Cognitive psychology principles can be applied in training sessions to assist operators to encode and recall safety procedures, such as using mnemonics [47]. To illustrate, the SBAR (Situation, Background, Assessment, Recommendation) tool is an easy-to-remember mnemonic that provides a powerful framework for communication between health care team members about a patient's condition [48]. Also, it is crucial to make operators highly motivated to learn and maintain desired safety behaviors. As aforementioned, the attractiveness of behaviors and choice of action, depends on the perceived outcome of each behavior (desirable or undesirable) [13]. Hence, we argue that safety training needs to promote countermeasures that are conveyed in a way that resonates well with trainees and thus, produces persistent front-line behavior change. Trainers can incorporate storytelling to deliver TEM in an impactful manner (we often employ special forces personnel or athletes to discuss moments where they had to manage a number of threats with effective countermeasures). It is also essential for trainers to utilize other effective pedagogical methods when carrying out training. Integrating adult learning principles in safety training programs improves the safety knowledge retention and safety performance of participants [49, 50]. For instance, adult learners need to know how learning will be beneficial prior to undertaking training and are self-directed in that they are capable of making decisions on their own and dislike being directed or imposed by others in the training process [51, 52].

4 Behavior Change that Persists

So far, we have proposed an approach for conducting safety training for high-reliability industries by utilizing TEM in a way that resonates and connects with trainees whilst advocating for a high-performance focus. However, training research shows that the transfer of training is typically low; simply delivering training does not ensure its automatic transfer to the workplace [53]. Also, researchers have found that learning comes from a combination of sources - through formal training programs, on the job learning from colleagues and peers, trial and error in one's work and through the coaching and feedback from a manager. Lombardo and Eichinger [54] propose a '70:20:10' model in that individuals learn mostly on-the-job (70%), but also from others (20%) and through formal learning programs (10%). Thus, we propose the use of in-situ coaching in high-reliability industries as a strategy to optimize the transfer of safety training. Safety coaches, who are skilled at making observations, know what to look for as trainees develop their skills and, use dialogue and questioning skills to provide direction, help trainees to develop their skills out in the field [55]. Coaches are also needed to regularly enforce and support newly learned behaviors whilst encouraging high levels of motivation in trainees for learning to transfer into the workplace. We also propose that TEM training be permanently integrated with the existing training curriculum for training to improve performance over the long-term and to become fully ingrained in the work culture [56]. Otherwise, in the absence of recurrent training and reinforcement of desirable safety behaviors, attitudes and practices decay [57].

We also suggest, and have begun implementing, continued observation and monitoring of threats, errors and effective use of countermeasures into the daily behavior of the workforce. In aviation, observations of crew are regularly taken to determine the cause of an error, the response to the error, who detected the error and, the ultimate outcome. This observational methodology, the Line Operations Safety Audit (LOSA), acts as an effective safety analysis tool for the proactive management of threats and errors [58, 59]. Over 20,000 domestic and international airline flights have used LOSA with the Federal Aviation Administration and the International Civil Aviation Organisation supporting the use of this methodology in ensuring airline safety.

We have adapted LOSA for use in non-aviation high-reliability industries to ensure the solidification of behaviors acquired after completion of our targeted workshops (designed and delivered based on the principles outlined in this paper). From mining to power distribution, our approach has resonated with over 1,500 safety-critical operators, resulting in significant reductions in near misses, incidents, fines and time off tools; and increases in reporting, and task efficiency.

References

1. Helmreich, R.L.: On error management: lessons from aviation. BMJ **320**, 781–785 (2000)
2. Weick, K.E., Sutcliffe, K.M.: Managing the Unexpected – Assuring High Performance in an Age of Complexity. Jossey-Bass, San Francisco (2001)
3. Hollnagel, E., Wears, R.L., Braithwaite, J.: From safety-I to safety-II: a white paper. The Resilient Health Care Net, The University of Southern Denmark, University of Florida, USA, and Macquarie University, Australia (2015)
4. Hickey, E.J., Nosikova, Y., Pham-Hung, E., Gritti, M., Schwartz, S., Caldarone, C.A., Redington, A., Van Arsdell, G.S.: National aeronautics and space administration "threat and error" model applied to pediatric cardiac surgery: error cycles precede $\sim$ 85% of patient deaths. J. Thorac. Cardiovasc. Surg. **149**, 496–507 (2015)
5. Gerstle, C.: Parallels in safety between aviation and healthcare. J. Pediatr. Surg. **53**, 875–878 (2018)
6. Ruskin, K.J., Stiegler, M.P., Park, K., Guffey, P., Kurup, V., Chidester, T.: Threat and error management for anesthesiologists: a predictive risk taxonomy. Curr. Opin. Anaesthesiol. **26**, 707–713 (2013)
7. Hudson, P., Verschuur, W.L.G., Parker, D., Lawton, R., van der Graaf, G.: Bending the rules: managing violation in the workplace. In: Invited Keynote Address, Society of Petroleum Engineers International Conference on Health, Safety and Environment in Oil and Gas Exploration (1998)
8. Klinect, J., Murray, P., Merritt, A., Helmreich, R.L.: Line operation safety audits (LOSA): definition and operating characteristics. In: Proceedings of the 12th International Symposium on Aviation Psychology, pp. 663–668. The Ohio State University, Dayton (2003)

9. Helmreich, R.L., Merritt, A.C.: Culture at Work: National, Organizational and Professional Influences. Ashgate, Aldershot (1998)
10. Helmreich, R.L., Merritt, A.C., Wilhelm, J.A.: The evolution of crew resource management in commercial aviation. Int. J. Aviat. Psychol. **9**, 19–32 (1999)
11. Merandi, J., Vannatta, K., Davis, J.T., Mcclead, R.E., Brilli, R., Bartman, T.: Safety II behaviour in a pediatric intensive care unit. Pediatrics **141**, e20180018 (2018)
12. Andriessen, J.H.T.H.: Safe behaviour and safety motivation. J. Occup. Accid. **1**, 363–376 (1978)
13. Vroom, V.: Work and Motivation. Wiley, New York (1964)
14. Dekker, S.W.A.: Just culture: who gets to draw the line? Cogn. Technol. Work **11**, 177–185 (2008)
15. Marx, D.: Patient safety and the "just culture": a primer for health care executives. Columbia University, New York (2001)
16. Ruitenberg, B.: Court case against Dutch controllers. Controller **41**, 22–25 (2002)
17. Eurocontrol Performance Review Commission: Report on Legal and Cultural Issues in Relation to ATM Safety Occurrence Reporting in Europe. Performance Review Unit, Brussels (2006)
18. Dekker, S.W.A.: Just Culture: Balancing Safety and Accountability. CRC Press, London (2016)
19. Ferguson, J., Fakelmann, R.: The culture factor. Front. Health Serv. Manag. **22**, 33–40 (2005)
20. Dekker, S.W.A.: Failure to adapt or adaptations that fail: contrasting models on procedures and safety. Appl. Ergon. **34**, 233–238 (2003)
21. Klein, G.A.: A recognition-primed decision (RPD) model of rapid decision making. In: Klein, G.A., Orasanu, J., Calderwood, R., Zsambok, C. (eds.) Decision Making in Action: Models and Methods, pp. 138–147. Ablex, NJ (1993)
22. Sanne, J.: Creating safety in air traffic control. Doctoral Dissertation, Arkiv (1999)
23. Hudson, P.T.W., Verschuur, W.L.G.: Why People Offshore Bend the Rules. Report for SIPM, Centre for Safety Science, Leiden University (1995)
24. Hilyer, B., Veasey, A., Oldfield, K., McCormick, L.: Effective Safety and Health Training. CRC Press, Boca Raton (2000)
25. Barnett, J.D.: Safety Management Handbook: CCH Safety Professional Series, vol. 2. Health and Human Resources, Chicago (2000)
26. Cullen, E.T., Fein, A.H.: Tell Me a Story: Why Stories are Essential to Effective Training. NIOSH, Cincinnati (2005)
27. Garguilo, T.L.: Making Stories: A Practical Guide to Organisational Leaders and Human Response Specialists. Quorum Books, Westport (2002)
28. Burke, M.J., Sarpy, S.A., Smith-Crowe, K., et al.: Relative effectiveness of worker safety and health training methods. Am. J. Public Health **96**, 315–324 (2006)
29. Herbet, S.: From spy to okay guy: trust and validity in fieldwork with the police. Geogr. Rev. **91**, 304–310 (2001)
30. Spielholz, P., Clark, R., Sjostrom, T.: Fatality narratives: an effective way to convey hazard information. Prof. Saf. **52**, 22–25 (2007)
31. Bliss, M.M., Dalto, J.: Storytelling in safety training. Prof. Saf. **63**, 34–35 (2018)
32. Sanne, J.M.: Incident reporting or storytelling? Competing schemes in a safety-critical and hazardous work setting. Saf. Sci. **46**, 1205–1222 (2008)
33. Weick, K.E.: Sensemaking in Organisations. Sage, Thousand Oaks (1995)
34. Smith, A.F., Goodwin, D., Mort M., Pope, A.: Take-off and landing in anaesthesia: negotiating the human-machine collective. In: Proceedings from the Society for the Social Studies of Science, Milwaukee (2003)

35. Frank, A.W.: The standpoint of storyteller. Qual. Health Res. **10**, 354–365 (2000)
36. Bruner, J.: Acts of Meaning. Harvard University Press, Cambridge (1990)
37. I.C.A.O. Circular 240-AN/144: Human Factors Digest No. 7, Investigation of Human Factors in Accidents and Incidents (1993)
38. Lesyk, J.J.: The nine mental skills of successful athletes. In: Annual Conference of the Association for the Advancement of Applied Sport Psychology, Hyannis (1998)
39. Erciş, S.: Comparison of mental skills of elite and non-elite athletes. J. Educ. Training Stud. **6**, 72–75 (2018)
40. Fontani, G., Migliorini, S., Benoccie, R., Facchini, A., Casini, M., Corradeschi, F.: Effect of mental imagery on the development of skilled motor actions. Percept. Mot. Skills **105**, 803–826 (2007)
41. National Research Council: Opportunities in Neuroscience for Future Army Applications. National Academies Press, Washington (2009)
42. Davis, J.M., Bailey, S.P.: Possible mechanisms of central nervous system fatigue during exercise. Med. Sci. Sports Exerc. **29**, 45–57 (1997)
43. Davis, J.M.: Nutrition, neurotransmitters and CNS fatigue. In: Maughan, R.J. (ed.) Nutrition in Sport, pp. 171–183. Blackwell Science, Malden (2000)
44. Stetz, M.C., Thomas, M.L., Russo, M.B., Stetz, T.A., Wildzunas, R.M., McDonald, J.J., Romano, J.A.: Stress, mental health, and cognition: a brief review of relationships and countermeasures. Aviat. Space Environ. Med. **78**, 252–260 (2007)
45. Meichenbaum, D.: Stress Innoculation Training. Pergamon Press, New York (1985)
46. Driskell, J.E., Johnston, J.H.: Stress exposure training. In: Cannon-Bowers, J.A., Salas, E. (eds.) Making Decisions Under Stress: Implications for Individual and Team Training, pp. 191–217. American Psychological Association, Washington D.C. (1998)
47. Beitz, J.M.: Unleashing the power of memory: the mighty mnemonic. Nurse Educ. **22**, 25–29 (1997)
48. O'Daniel, M., Rosenstein, A.H.: Professional communication and team collaboration. In: Hughes, R.G. (ed.) Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Agency for Healthcare Research and Quality, Rockville (2008)
49. Albert, A., Hallowel, M.R.: Revamping occupational safety and health training: integrating andragogical principles for the adult learner. Australas. J. Constr. Econ. Build. **13**, 128–140 (2013)
50. Juárez-Carrillo, P.M., Liebman, A.K., Reyes, I.A.C., Ninco Sánchez, Y.V., Keifer, M.C.: Applying learning theory to safety and health training for hispanic immigrant dairy workers. Health Promot. Pract. **18**, 505–515 (2017)
51. Knowles, M.S.: The Modern Practice of Adult Education: Andragogy Versus Pedagogy. Association Press, Oxford (1970)
52. Reece, I., Walker, S.: Teaching, Training and Learning: A Practical Guide. Business Education Publishers Ltd., Houghton Le Spring (2016)
53. Blair, E., Seo, D.C.: Safety training making the connection to high performance. Prof. Saf. **52**, 42–48 (2007)
54. Lombardo, M., Eichinger, R.: The Career Architect Development Planner. Lominger Limited (1996)
55. Miller, K.L.: Objective-Based Safety Training. CRC Press, Boca Raton (1998)
56. Lee, S.Y., Bates, P.R., Murray, P.S., Martin, W.L.: An exploratory study on the post-implementation of threat and error management training in Australian general aviation. Int. J. Training Res. **15**, 136–147 (2017)
57. Stewart, J.: Managing for World-Class Safety. Wiley, New York (2002)

58. Helmreich, R.L., Klinect, J.R., Wilhelm, J.A.: Models of threat, error, and CRM in flight operations. In: Proceedings of the Tenth International Symposium on Aviation Psychology, pp. 677–682. Ohio State University, Columbus (1999)
59. Khoshkhoo, R., Jahangirian, A., Sharafbafi, F.: Analysis of fleet type impact on the threats and errors of an airline using line operations safety audit (LOSA). Aviation **22**, 31–39 (2018)

Theoretical Advances in Human Error and Performance

Relating the Learning Styles, Dependency, and Working Memory Capacity to Performance Effectiveness in Collaborative Problem Solving

Kaveh Sheikhrezaei and Craig Harvey(✉)

Louisiana State University, Baton Rouge, LA, USA
harvey@lsu.edu

Abstract. Although much research has concentrated on the forecast of efficient team performance and the variables that might detract or promote team efficiency, little research reviewed has assessed the multitude of individual characteristics their impact on collaborative problem solving (CPS). Much of the research investigates only a single individual characteristic and its effect on group performance. This research proposes to explore three individual attributes (interpersonal dependency, individual working memory capacity, and preferred learning style) on performance effectiveness in CPS. A wide range of fields including healthcare and the military has explored CPS; however, the bulk of teamwork research to date has dealt with behavioral coordination on a single feature. This study will explore the association between team-member attributes and CPS skills. Noteworthy interactions might be observed to demonstrate that there are mixtures of traits more (or less) productive than anticipated, indicating further evidence of how group composition influences group performance.

Keywords: Human factors · Collaborative problem solving · Individual differences factors · Performance effectiveness

1 Introduction and Literature Review

As the use of groups has increased, research studies concentrating on the forecast of efficient team performance and the variables that might detract or promote team efficiency has actually increased [1–3]. Collaborative problem solving (CPS) is being used in a variety of group task environments (e.g. face-to-face) specifically for novel or non-routine tasks. Several research studies and reports indicate the importance of CPS [4–7].

Most complex problems require that teams work together to find solutions. The very principle of collaborative problem solving (CPS) is merging the individuals' knowledge to accomplish common goals. Lack of education and training in CPS provides an opportunity to recognize strategies to improve CPS. Researchers that have studied CPS have identified opportunities and challenges for the development of research on CPS [5]. The latest improvements in digital technologies can be employed to automate the CPS processes along with the detection and evaluation of different CPS competencies. If this is successful, digital technologies will allow investigators to

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 53–64, 2020.
https://doi.org/10.1007/978-3-030-20037-4_5

gather and evaluate very large sets of data in a varied range of tasks, settings, and populations. This would make significant progress in discriminating theories, testing hypotheses and developing an educational curriculum for CPS training and education. In the past, understanding of team members' communication was a challenge in evaluating collaboration [5].

In general, collaborative problem solving has two primary parts: the collaborative (e.g., social aspects or communication) and the cognitive aspects or knowledge (e.g., domain-specific problem-solving techniques) [8]. These two parts are frequently described as "task work" and "teamwork". The individual problem solving and collaborative problem solving are different from the social part in the context of a group task. Problem-solving requires team's exchange ideas, communicate, and problem solve with their team mates.

There are numerous and varied instances of collaborative problem-solving activities from casual class activities to large-scale official assessments of cooperation by online training systems [9]. There is likewise substantial research on the elements that impact the achievement of collaborative performance and collaborative learning [10].

Even with a growing quantity of organizations performing tasks using groups, little is understood how people included in a team impact intragroup procedures and results. The predominant theory of considering groups is the input– process – output design [11–14]. According this model, intragroup procedures and outputs are affected by the inputs integrate. According to Hackman [1], inputs are classified into the three groups: group level factors (e.g., team composition), individual-level factors (e.g., team-member qualities), and environmental-level elements (e.g., task attributes).

Early researchers (e.g., [15, 16]) assumed that team composition affected both team processes and outputs. Further, Senior and Swailes [17] have recognized team composition as a crucial aspect that affects team performance. The composition considers the personal characteristics of participants (e.g., ability, experience, and skill) as well as how they can possibly integrate to determine total efficiency results for the group. Regardless of the understanding of team composition elements value [18], few researchers have studied the result of non-demographic attributes on team processes and outcomes. For instance, self-report procedures of skill, knowledge, collectivism, experience, group size, and flexibility as composition variables have been used in two different research [19, 20]. The considerable relations among these composition elements and team process and efficiency measures have been shown by the researchers. Additionally, Yazici [21] found evidence that shows the value of learning style preferences to involve learners in different collaborative tasks and to design effective varied teams. Future research is needed to understand critical individual characteristics (e.g., learning style, dependency, and working memory capacity) of team members and their interactive effect upon team member performance.

1.1 Team Composition (Group Level Factors)

Team composition describes the general mix of attributes amongst individuals in a group, which is a component of at least two people who connect interdependently to attain a common goal [22]. As a result, team composition has actually been a popular subject. In theory, team composition research study goes to the heart of comprehending how individual characteristics integrate to form effective interdependent groups.

1.2 Team-Member Attributes (Individual-Level Factors)

Human performance can be influenced for many reasons (e.g. age, mental state, physical health, personal attitude, emotions, and cognitive biases). In this research, characteristics such as interpersonal dependency, individual working memory capacity, or preferred learning style are hypothesized to contribute considerably to the variation in group performance efficiency.

Learning Styles. Learning styles point to a number of competitive and controversial theories whose purpose is to recognize differences in the individuals' learning processes [23]. These theories suggest that all individuals can be categorized based on the style of "learning" though different theories offer different perspectives on how to define and classify them [23]. A common opinion is a difference in ways that people learn something [24]. Individualized learning styles have been considered since the 1970s [23] and has significantly impacted education in spite of the received criticism from some researchers [25].

Individuals have various learning styles characteristic preferences and strengths in their way of capturing and processing information. While some tend to the emphasize facts, algorithms, and data; others are more interested in mathematical models and theories. Some people respond more to graphical information forms, like diagrams, schematics, and pictures; some individuals are more comfortable with spoken and written explanations. Finally, while some have a preference in learning actively and interactively; others function more individually.

Kolb [26] in his empirical learning model indicates that learning is an interactive procedure containing four different modes of learning: (1) Active Experimentation (AE); (2) Concrete Experience (CE); (3) Reflective Observation (RO); and (4) Abstract Conceptualization (AC). Concrete and abstract make up one continuum while Reflective and Active make up another continuum. Depending on where an individual falls within each continuum, four specific styles are defined: the accommodative (AE/CE), the assimilative (RO/AC), the convergent (AC/AE), and the divergent (CE/RO).

Mumford and Honey [27] started using the Learning Style Inventory (LSI) introduced by Kolb, a readily presented and very first diagnostic instrument, for observing how individuals learn.

Given that the four classes are linked to a modified variation of Kolb's empirical learning cycle, the relations with Kolb's Learning Style Inventory (LSI) remain significant. Therefore, for instance, activists are known to be qualified for having experience; reflectors review experience; theorists make conclusions from their experience; and pragmatists for relying on practical actions (see Fig. 1).

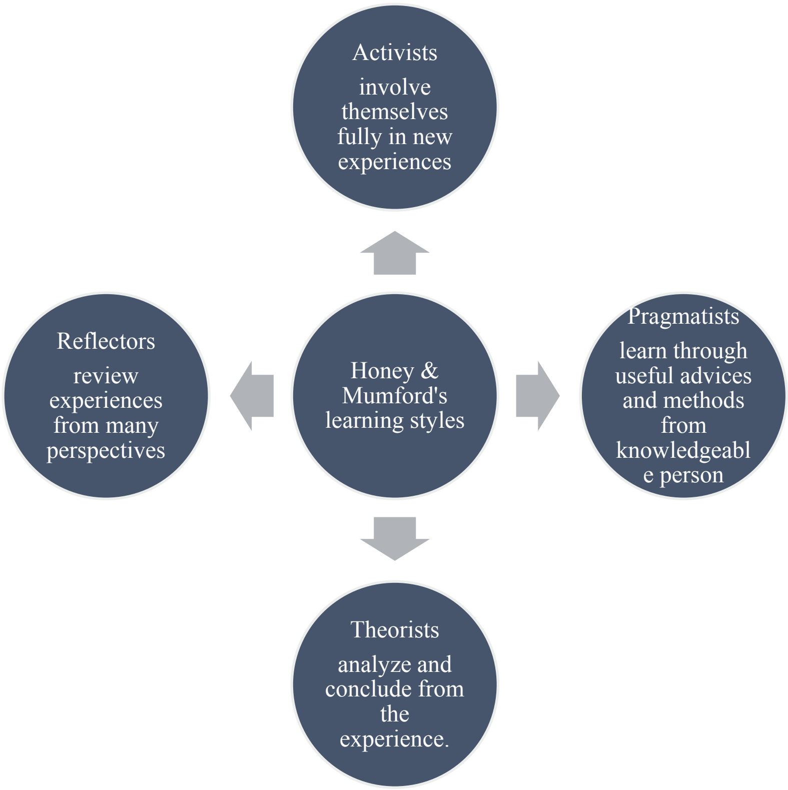


Fig. 1. Dimensions of honey and mumford's learning style

Table 1 shows the different researchers that have explored learning styles since the 80s.

Table 1. Different researchers about learning styles

Year	Researcher(s)	Measure
1985	Gregorc	Gregorc Mind Style Delineator (GSD)
1985	Myers-Briggs	Myers-Briggs Type Indicator (MBTI)
1989	Hermann	Brain Dominance Instrument (HBDI)
1996	Allinsom and Hayes	Cognitive Style Index (CSI)
1998	Entwistle	Approaches to Study Inventory (ASI)
1998	Riding	Cognitive Styles Analysis (CSA)
1998	Vermunt	Inventory of Learning Styles (ILS)
1999	Sterberg	Thinking Styles
1999	Kolb	Learning Style Inventory (LSI)
2000	Honey & Mumford	Learning Styles Questionnaire (LSQ)
2001	Apter	Motivational Style Profile (MSP)
2002	Jackson	Learning Styles Profiler (LSP)
2003	Dunn & Dunn	Productivity Environmental Preference Survey (PEPS)

Leaning Styles and Problem Solving. There are few investigations in the literature that assess the connection between learning styles and problem-solving abilities. Bhat [28] found some support for this when he concluded that learning styles have effects on the students' problem-solving ability and that among all learning styles assimilator had the better problem-solving ability.

More recently, a study by Aljaberi [29] determined significant differences between the students' learning styles in solving the mathematical problems. His research also shows the Activist-Reflector style the most preferred style and also the superior performance in mathematical problems.

Similarly, Sebastian [30] reported that student's level of difficulties for computational and conceptual might be influenced by his/her learning style in solving problems. The accommodator and assimilator students are expected to have average difficulty level with conceptually difficult problems; while a low and high difficulty level are expected of converger and diverger students respectively. For computationally difficult problems, convergers have a low difficulty level; while assimilators tend to have average to high difficulty level. Both the diverger and accommodator tend to have an average difficulty level. Others such as Sirin and Güzel [31], who used the Problem-solving Inventory [32] and the Learning Style Inventory [26] found that the students' learning style types are not related to their problem-solving abilities.

Conversely, it was observed that problem-solving abilities had a negative correlation with abstract conceptualization (AC) learning style and positive relationship with reflective-observation (RO) learning style. The students' problem-solving ability levels were perceived as poorer than expected [31].

Dependency. Individuals seek security, support, assurance, and guidance from outside themselves as the result of personal dependency. Another person, a social unit or a symbolic belief system are some examples from which individuals are given help and support. The desired support can be physical (reliance on caregiver), cognitive (affiliation between a learner and instructor) and/or emotional (dependence on someone else to ensure and love).

People vary in the quantity of convenience and assistance required from others. Some individuals are extremely dependent on those around them, while others operate more independent.

Various evaluation instruments have been established to evaluate levels of interpersonal dependency. Various measures of dependency have been established since the idea of dependency is of interest to scientists in widespread areas. In Table 2, different scales to measure the personal dependency are presented.

Table 2. Different scales to measure the personal dependency

Year	Researcher(s)	Scale (latest version)
1949	Blum	Blacky Test Oral Dependency Scale (BTODS)
1956	Kagan & Mussen	Thematic Apperception Test dependency scale (TAT)
1967	Masling et al.	Rorschach Oral Dependency scale (ROD)
1976	Blatt et al.	Depressive Experiences Questionnaire (DEQ)
1977	Hirschfelf et al.	Interpersonal Dependency Inventory (IDI)
1983	Beck et al.	Sociotropy-Autonomy Scale (SAS)
1987	Zimmerman & Coryell	Inventory to Diagnose Depression - Lifetime version (IDD-L)
1991	Morey	Personality Assessment Inventory (PAI)
1994	Paulhus	Balanced Inventory of Desirable Responding (BIDR)
1996	Beck et al.	Beck Depression Inventory - II (BDI-II)
2008	Ben-Porath & Tellegen	Minnesota Multiphasic Personality Inventory – the latest version Restructured Form (MMPI-2-RF)
2015	Millon et al.	Millon Clinical Multiaxial Inventory - Fourth Edition (MCMI-IV)

Dependency and Problem Solving. Some research has evaluated the association between dependency and problem-solving skills. Research studies performed in the previous 30 years on dependency reveals that it relates to problem-solving. Ronning et al. [33] discovered students with lower level of dependency significantly outperformed than students with higher level of dependency on problems. Although students with higher level of dependency may well benefit from thoroughly structured direction and specific goals. Similarly, a research study by Hagaa et al. [34] reveals that problem solving is likewise associated with dependency, however it keeps a substantial relation with depressive sign seriousness once reliance is statistically managed.

More recently, Wang et al. [35] stated a nonsignificant difference in solving simple and intermediate problems, but a significant effect in solving a complex problem. They indicated that independent students solved complex problems much better than dependent students.

Working Memory Capacity. Memory is essential to experiences and keeping information over time that affects future actions [36]. We might not be able to establish or learn a language, establish relationships, nor individuality handle problems, if we were unable to keep in mind previous occasions [37]. Frequently, memory is comprehended as an information processing system that is comprised of a sensory processor, short-term (or working) memory, and long-lasting memory [38].

As a brain system, working memory make available temporary storage and the required information to perform the complex cognitive tasks (e.g. language comprehension) [39]. Working memory can be defined as essential element in several practical tasks [40]. Since several everyday tasks include actively keeping information in mind, manipulating, and combining them in memory, working memory is an essential element to completing tasks. The amount of working memory (WM) capacity may determine how individuals perform different real-world cognitive tasks [41].

Over the past 30 years, numerous approaches have been proposed to study individual differences in working memory capacity (WMC). According to Conway et al. [42], perhaps a complex span paradigm is the best known and most common task to measure WMC. Several researchers on individual differences in WMC make this design solely by one or more complex tasks. Consequently, there is a very recent theory of how individual differences in WMC (perhaps very limited) can impact a complex span task class (e.g., [43–45]).

Working Memory and Problem Solving. Working memory maintains newly processed information to connect it to the newest input and also it holds the information to construct an overall representation of the problem. Therefore, Swanson and Beebe-Frankenberger [46] observed an association between the working memory and arithmetic problem solving ability for elementary school students.

Similarly, Barrouillet and Lépine [47] reported that both the efficiency and frequency of the retrieval strategy are influenced by the children's working memory capacities in simple mathematics problem-solving. Children with higher WMC completed retrieval tasks quicker. A study by Beilock and Carr [48] assessed the effects of working memory and pressure on students" ability to solve mathematical problems. In conditions where participants did not have any external pressure (e.g., time), they found individuals with low working memory (LWM) capacity solved the high-demand problems poorer. Nevertheless, in the condition where participants were under time pressure, the level of achievement for LWM was not decreased. However, time pressure impacts on the students with higher level of WM.

The work by Ashcraft and Krause [49] indicated that by increasing the number of steps in multistep problems, the reliance on working memory is increased. Similarly, when the need to retain intermediate values and goals is increased, working memory capacity becomes much more important. To solve mathematical problems, Wiley and Jarosz [50] have found an association between that students' performance and their working memory capacity. They also found the working memory capacity can improve the attention controlling, decrease distraction, and confine the problem solvers to search through a problem space. In analytic problem-solving contexts, the higher level of working memory capacity resulted usually better performance.

2 Proposed Model of Team Effectiveness

Figure 2 outlines an integrative model of attributes proposed to impact team performance for collaborative problem solving. Several research questions are proposed based on the individual moderating variables of learning style, dependency, and working memory capacity:

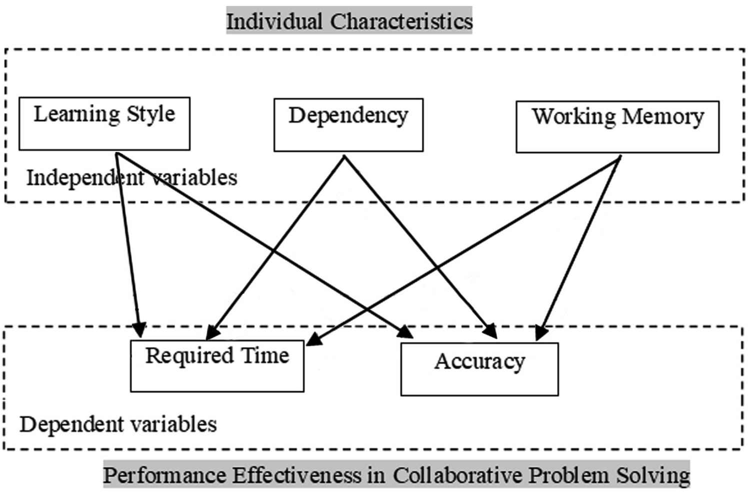


Fig. 2. Research model of individual learning style, dependency, and working memory on collaborative problem solving

1. Is there a difference in the pairs of individuals performance (amount of time required and accuracy (number of errors)) to complete a simple and a complex task?
2. Are the pairs of individuals' performances related to the composition of the groups?
3. How does performance vary based on the composition of the groups?

2.1 Summary of Hypotheses

Several hypotheses (Table 3) are proposed to answer these research questions looking at both main and interaction effects of the individual team member attributes.

Table 3. Summary of proposed hypotheses

	Hypotheses	Rationale
Main Effects	$H1_{a\text{-}b}$ The preferred Learning Styles has a significant effect on performance effectiveness in collaborative problem solving (amount of time required and accuracy)	Need to evaluate how learning styles and collaborative problem-solving abilities are connected. Research has not shown that there is evidence that one preference is better than another

(*continued*)

Table 3. (*continued*)

		Hypotheses	Rationale
		$H2_{a-b}$ The level of Personal Dependency has a significant effect on performance effectiveness in collaborative problem solving (amount of time required and accuracy)	This evaluates the association between dependency and collaborative problem-solving skills. The analysis of social science literature clearly shows that individual dependency was considered as negative [51] and [52] or positive [53] and [54] terms. On the one hand, individual reliance is equal to weakness and obstruction to develop an independent and mature individual. On the other hand, individual reliance is considered as a fundamental human inspiration to perform essential adaptive tasks
		$H3_{a-b}$ The level of Working Memory Capacity has a significant effect on performance effectiveness in collaborative problem solving (amount of time required and accuracy)	Working memory and problem-solving skills are related. The amount of working memory (WM) capacity may determine how individuals perform different real-world cognitive tasks [41]
Interaction Effects	First order interaction	$H4_{a-b}$, $H5_{a-b}$, and $H6_{a-b}$ There is no difference in Team Problem Solving Outcomes (amount of time required and accuracy) based on all possible pairs of independent variables (Learning Styles, Dependency and Working Memory Capacity)	Noteworthy interactions might be observed to demonstrate that there are mixtures of traits more (or less) productive than anticipated, providing proof that group composition influences group performance. Hence, some mixtures of individual traits may yield group performance differences. They might contribute considerably to the variation in group performance efficiency
	Second order interaction	$H7_{a-b}$ There is no difference in Team Problem Solving Outcomes (amount of time required and accuracy) based on all three independent variables (Learning Styles, Dependency and Working Memory Capacity)	

3 Conclusion and Justification

Teams solve many of our complex problems in society. As a result, it is important to understand how to improve team performance. This research proposes to understand critical individual characteristics (e.g., learning style, dependency, and working memory capacity) of team members and their interactive effect upon collaborative problem solving. With an understanding of the proposed attributes, guidance could be developed that could impact team performance based on elements such as team composition, team guidance toward a goal or even the means in which teams interact. While the literature on understanding teams is vast, our knowledge is still very limited at understanding the elements that contribute to team performance. This research hopes to close some of that gap in knowledge.

References

1. Hackman, J.R.: Groups that work and those that don't, (No. E10 H123). Jossey-Bass (1990)
2. Pieper, T.M., Klein, S.B., Jaskiewicz, P.: The impact of goal alignment on board existence and top management team composition: evidence from family-influenced businesses. J. Small Bus. Manag. **46**(3), 372–394 (2008)
3. National Research Council: Enhancing the effectiveness of team science. National Academies Press (2015)
4. D.C.P.S. Framework: Pisa 2015 Draft Collaborative Problem Solving Framework (2013)
5. Graesser, A.C., Fiore, S.M., Greiff, S., Andrews-Todd, J., Foltz, P.W., Hesse, F.W.: Advancing the science of collaborative problem solving. Psychol. Sci. Public Interes. **19**(2), 59–92 (2018)
6. Fiore, S.M., Graesser, A., Greiff, S.: Collaborative problem-solving education for the twenty-first-century workforce. Nat. Human Behav. **1** (2018)
7. Sottilare, R.A., Burke, C.S., Salas, E., Sinatra, A.M., Johnston, J.H., Gilbert, S.B.: Designing adaptive instruction for teams: a meta-analysis. Int. J. Artif. Intell. Educ. **28**(2), 225–264 (2018)
8. Fiore, S.M., Graesser, A., Greiff, S., Griffin, P., Gong, B., Kyllonen, P., Massey, C., O'Neil, H., Pellegrino, J., Rothman, R., Soulé, H.: Collaborative problem solving: considerations for the national assessment of educational progress (2017)
9. Griffin, P., Care, E. (eds.): Assessment and teaching of 21st century skills: Methods and approach. Springer, Heidelberg (2014)
10. Von Davier, A.A., Zhu, M., Kyllonen, P.C. (eds.): Innovative assessment of collaboration. Springer, Heidelberg (2017)
11. Hackman, J.R.: The design of work teams. In: Lorsch, J.W. (ed.) Handbook of Organizational Behavior, pp. 315–342. Prentice Hall, Englewood Cliffs (1987)
12. Guzzo, R.A., Shea, G.P.: Group performance and intergroup relations in organizations. In: Dunnette, M.D., Hough, L.H. (eds.) Handbook of Industrial and Organizational Psychology, 2nd edn., vol. 3, pp. 269–313. Consulting Psychologists Press, Palo Alto (1992)
13. McGrath, J.E.: Social Psychology: A brief introduction. Holt, Rinehart & Winston, New York (1964)
14. Gladstein, D.L.: Groups in context: a model of task group effectiveness. Adm. Sci. Q. **29**, 499–517 (1984)

15. Cattell, R.B.: Concepts and methods in the measurement of group syntality. Psychol. Rev. **55**, 48–63 (1948)
16. Haythorn, W.: The influence of individual members on the characteristics of small groups. J. Abnorm. Soc. Psychol. **48**, 276–284 (1953)
17. Senior, B., Swailes, S.: The dimensions of management team performance: a repertory grid study. Int. J. Prod. Perform. Manag. **53**(4), 317–333 (2004)
18. Sundstrom, E., De Meuse, K.P., Futrell, D.: Work teams: applications and effectiveness. Am. Psychol. **15**, 120–133 (1990)
19. Campion, M.A., Medsker, G.J., Higgs, A.C.: Relations between work group characteristics and effectiveness: implications for designing effective work groups. Pers. Psychol. **46**(4), 823–847 (1993)
20. Campion, M.A., Papper, E.M., Medsker, G.J.: Relations between work team characteristics and effectiveness: a replication and extension. Pers. Psychol. **49**, 429–452 (1996)
21. Julie Yazici, H.: A study of collaborative learning style and team learning performance. Education + training, **47**(3), 216–229 (2005)
22. Hackman, J.R., Wageman, R.: A theory of team coaching. Acad. Manag. Rev. **30**(2), 269–287 (2005)
23. Coffield, F., Moseley, D., Hall, E., Ecclestone, K.: Learning styles and pedagogy in post-16 learning: a systematic and critical review (2004)
24. Willingham, D.T., Hughes, E.M., Dobolyi, D.G.: The scientific status of learning styles theories. Teach. Psychol. **42**(3), 266–271 (2015)
25. Pashler, H., McDaniel, M., Rohrer, D., Bjork, R.: Learning styles: concepts and evidence. Psychol. Sci. Public Interes. **9**(3), 105–119 (2008)
26. Kolb, D.A.: Experiential Learning: Experience as the Source of Learning and Development. Prentice Hall Inc., Englewood Cliffs (1984)
27. Honey, P., Mumford, A.: Using your learning styles. Chartered Institute of Personnel and Development (1986)
28. Bhat, M.A.: The effect of learning styles on problem solving ability among high school students. Int. J. Adv. Soc. Sci. Humanit. **2**(7), 1–6 (2014)
29. Aljaberi, N.M.: University students' learning styles and their ability to solve mathematical problems. Int. J. Bus. Soc. Sci. **6**(4) (2015)
30. Sebastian, M.A.: Students' gender, learning style, and difficulties in solving problems in college algebra (2017)
31. Sirin, A., Güzel, A.: The relationship between learning styles and problem solving skills among college students. Educ. Sci.: Theory Pract. **6**(1), (2006)
32. Heppner, P.P., Petersen, C.H.: The development and implications of a personal problem-solving inventory. J. Couns. Psychol. **29**(1), 66 (1982)
33. Ronning, R.R., McCurdy, D., Ballinger, R.: Individual differences: a third component in problem-solving instruction. J. Res. Sci. Teach. **21**(1), 71–82 (1984)
34. Haaga, D.A., Fine, J.A., Terrill, D.R., Stewart, B.L., Beck, A.T.: Social problem-solving deficits, dependency, and depressive symptoms. Cogn. Ther. Res. **19**(2), 147–158 (1995)
35. Wang, L.J., Wang, X., Ren, M.Z.: Field-dependent-independent cognitive style in solving dynamics problems. Psychol. Rep. **92**(3), 867–880 (2003)
36. Sherwood, L.: Human Physiology: From Cells to Systems. Cengage Learning, Boston (2015)
37. Eysenck, M.W., Brysbaert, M.: Fundamentals of Cognition. Routledge, Abingdon (2018)
38. Baddeley, A.: Working Memory, Thought, and Action, vol. 45. OUP, Oxford (2007)
39. Baddeley, A.: Working memory. Science **255**(5044), 556–559 (1992)
40. Baddeley, A.D., Hitch, G.: Working memory. In: Psychology of Learning and Motivation, vol. 8, pp. 47–89. Academic press (1974)

41. Engle, R.W.: Working memory capacity as executive attention. Curr. Dir. Psychol. Sci. **11** (1), 19–23 (2002)
42. Conway, A.R., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., Engle, R.W.: Working memory span tasks: a methodological review and user's guide. Psychol. Bull. Rev. **12**(5), 769–786 (2005)
43. Barrouillet, P., Portrat, S., Camos, V.: On the law relating processing to storage in working memory. Psychol. Rev. **118**(2), 175 (2011)
44. Unsworth, N., Engle, R.W.: On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. Psychol. Bull. **133**(6), 1038 (2007)
45. Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., Greaves, M.: Modeling working memory: an interference model of complex span. Psychon. Bull. Rev. **19**(5), 779–819 (2012)
46. Swanson, H.L., Beebe-Frankenberger, M.: The relationship between working memory and mathematical problem solving in children at risk and not at risk for serious math difficulties. J. Educ. Psychol. **96**(3), 471 (2004)
47. Barrouillet, P., Lépine, R.: Working memory and children's use of retrieval to solve addition problems. J. Exp. Child Psychol. **91**(3), 183–204 (2005)
48. Beilock, S.L., Carr, T.H.: When high-powered people fail: working memory and "choking under pressure" in math. Psychol. Sci. **16**(2), 101–105 (2005)
49. Ashcraft, M.H., Krause, J.A.: Working memory, math performance, and math anxiety. Psychon. Bull. Rev. **14**(2), 243–248 (2007)
50. Wiley, J., Jarosz, A.F.: Working memory capacity, attentional focus, and problem solving. Curr. Dir. Psychol. Sci. **21**(4), 258–262 (2012)
51. Bornstein, R.F.: The Dependent Personality. Guilford Press, New York (1993)
52. Greenberg, J.R., Mitchell, S.A.: Object Relations in Psychoanalytic Theory. Harvard UP, Cambridge (1983)
53. Kelley, H.H.: Personal Relationships: Their Structures and Processes. Psychology Press, London (2013)
54. Folkman, S.: Stress: appraisal and coping. In: Encyclopedia of Behavioral Medicine, pp. 1913–1915. Springer, New York (2013)

Research on Human Error in Operation Task Under the Coupling of Time of Day and Stress

Xing Pan, Xianheng Zhao, Wenjin Zhang(✉), and Yuhang Jiang

School of Reliability and Systems Engineering, Beihang University, Beijing 100191, China
{panxing, zwjok}@buaa.edu.cn, 13020091921@163.com, forimba@outlook.com

Abstract. By investigating human errors occur in rocket's fuel-filling mission, several important performance shaping factors (PSFs) which have great influence on human error are revealed. It is also found that among all error modes, timing errors (including action too late and action too early) and omissions (including omitting entire step and omitting operation in step) occur most frequently. Timing errors and omissions are both considered as human errors that fail the mission. This paper selects time of day and stress as two PSFs to study the effects of PSFs on human error probability (HEP). This paper also explores whether there is a coupling between the two PSFs.

Keywords: Human reliability · Human error probability · Stress index · Time of day · Fuel filling simulation · Significant interaction

1 Introduction

Space launch missions have always been the focus. In the field of spaceflight, human factors are involved in design, implementation and maintenance stages. Even in the highly automated rocket launching stage, human factors play a role. It is found through research, that human error is the main cause of risk accidents in the field of space launch [1, 2]. In some other cases, human error may even lead to disastrous consequences, such as casualties or damage to key equipment. Therefore, in order to improve the reliability of the whole launch mission, it is necessary to consider fully the reliability of operators, that is, human reliability.

In fact, there is already a consideration of human reliability which aims at improving the safety and reliability in the field of spaceflight [3]. Human reliability analysis is a combination of qualitative and quantitative methods, which can conduct a comprehensive and systematic analysis of human errors [4]. The analysis can recognize human factors that cause negative results in a system, give a probability of human errors that may occur, and identify the effects of happened human errors.

As the executors of space launch mission, human beings have subjective initiative but also could be affected by other factors. The analysis of human reliability in space launch mission can provide a theoretical basis for the diagnosis, quantification and prevention of human error in the field of spaceflight, so it is of great significance.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 65–74, 2020.
https://doi.org/10.1007/978-3-030-20037-4_6

In this paper, fuel filling task in the process of rocket launching is selected as research object [5, 6], and human errors and improvement methods in the process of fuel filling are analyzed through designed simulation experiment [7].

2 Overview of Simulation Environment

2.1 Experimental Device

This experiment was mainly carried out by a computer with developed filling simulation program. The experimental device includes a computer with filling simulation program, an operating keyboard and a display screen, as shown in Fig. 1 below.

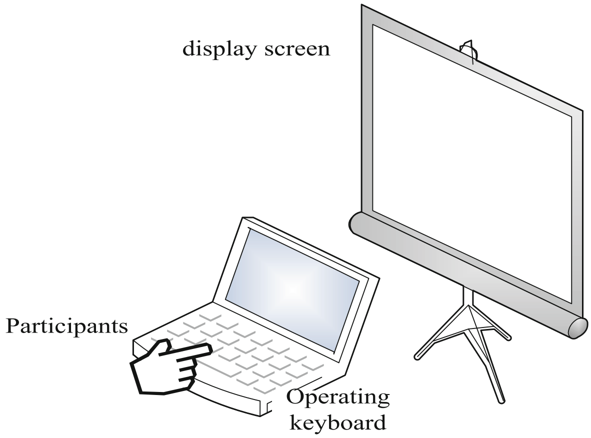


Fig. 1. Simulation experiment device

In the experiment, participants input the operation according to instructions on the keyboard and observe the visual information given on the display screen. Elements such as valves, pumps, vaporizers, fuel tanks and corresponding parameters in the fuel filling process are displayed graphically on the screen, and corresponding visual feedback is given for key input, thus revealing human-computer interaction.

The filling simulation program includes fast filling and normal filling simulation program in experiments. Two different stress levels are differed by different filling speeds, so that participants' stress can be distinguished.

The filling interface program is written in Python language. The keys input, the time and the filling status of corresponding time of participants could be recorded in background and output in the form of Excel table.

3 Design of the Experiment

The experimental design can be divided into two parts: the design of performance shaping factors (PSFs) as well as the selection of failure modes and corresponding explanations. In this experiment, the authors designed two performance shaping factor-two level to collect human error data for human error analysis [8].

The schematic diagram of the experimental design is given below (Fig. 2).

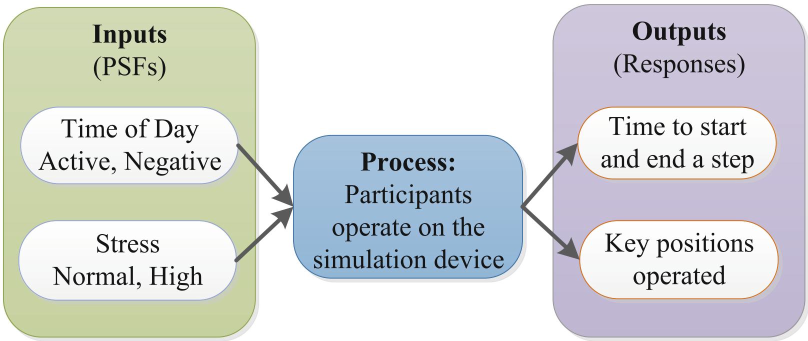


Fig. 2. Schematic diagram of experimental design

3.1 Performance Shaping Factors

Through data collection and the analysis of actual rocket launching site, two PSFs, time of day and stress, which are the most prominent factors affecting human errors, are extracted. Each factor was designed at two levels.

Time of Day. This PSF refers to the state of operator's body during the experiment, that is, whether it meets the requirements of normal work [9–11]. In this study time of day (TOD) is defined under two levels: active and negative. According to Melanie's study [12], operation time has a great influence on the efficiency and quality of work for operators. In this experiment, 24:00 pm was chosen as the negative level of time of day, while 10:00 am was chosen as the active level.

Stress. Stress refers to the degree of psychological tension caused by internal or external factors when operators participate in the experiment. [13] In this experiment, two different time pressure levels were selected, and the stress index (SI) was used to represent time pressure. Stress index (SI) is defined as the ratio of time required (Tr) to time available (Ta). So there is an equation SI = Tr/Ta. [14] Time required refers to the time each participant needs to complete the operation process as quickly as possible without time constraints. Time available refers to the time that the participant is allowed to use in the operation process. In this experiment, two levels of time pressure were given: normal level when SI = 1 and high SI = 1.5. The time available in operation process was given according to a large number of operation experiments conducted by previous experimental designers and had converted into the speed of filling. On the basis of the filling speed of SI = 1, the filling speed was changed into SI = 1.5. In addition, because of a large amount of training and long-time interval between two operations, the experience from the previous experiments can be ignored.

3.2 Human Error Modes

In the operation of real rocket launching site, there are many human error modes (Fig. 3) [15, 16]. In this paper, two error modes, timing errors (including action too early and action too late) and omissions (including omitting entire step and omitting operation in step), were selected to study. Detailed modes and explanations are shown in Table 1.

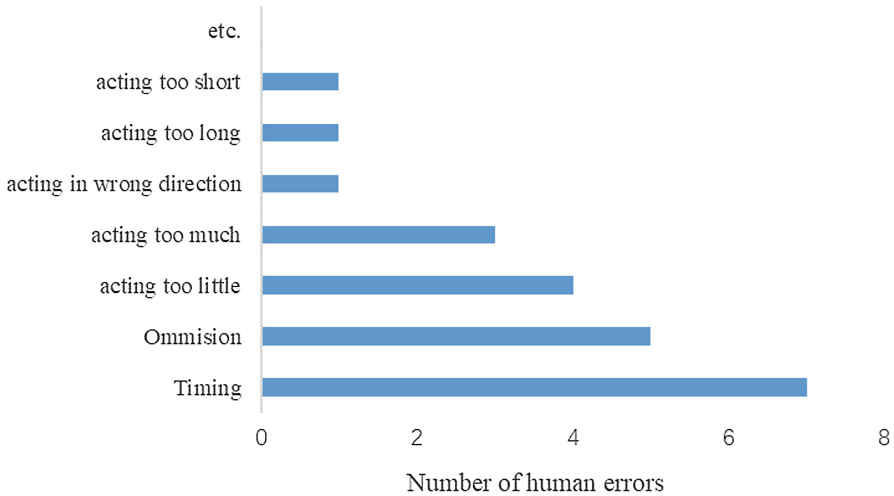


Fig. 3. Statistics of external error modes of launch sites

For the operation of whole filling task, as far as one step was concerned, this experiment considered that the maximum number of times that an error (sub)mode may occur is one. There are 10 steps in the experiment, the authors could thus get the times that an error (sub)mode occurs in an experimental operation, as shown in Table 2.

Table 1. Human error modes and criteria

Error mode	Error submode	Error criterion
Timing errors	Action too early	Taking next step before the specified time
	Action too late	Taking next step after the specified time
Omissions	Omit entire step	None of the operations in the step are done
	Omit operation in step	One or more operations in a step have not been done

Table 2. Human error's possible occurrences in every step

Step	Task stage	Possible occurrences			
		Timing errors		Ommisions	
		Action too early	Action too late	Omit entire step	Omit operation in step
1	TASK STAGE 1	–	–	–	1
2	TASK STAGE 2	1	1	1	1
3	TASK STAGE 3	1	1	1	1
4	TASK STAGE 4	1	1	1	–
5	TASK STAGE 5	1	1	1	–
6	TASK STAGE 6	1	1	1	1
7	TASK STAGE 7	1	1	1	1
8	TASK STAGE 8	1	1	1	1
9	TASK STAGE 9	1	1	1	1
10	TASK STAGE 10	1	1	1	1
Total		9	9	9	8
		18		17	

4 Data Collection and Primary Data Process

4.1 Participants

This experiment recruited 20 undergraduate participants who had not been exposed to relevant experimental operations before. They all have good daily routine. In order to rationalize time of day, this experiment required them to ensure normal work and rest a few days before the experiment. All of them received adequate operational training before formal execution. In order to reduce the impact of individual differences, 20 participants were randomly divided into two groups on average and human errors were counted separately in the two groups.

Assuming that a level combination of two PSFs is regarded as one scenario, then each participant should participate in the operations of all four scenarios (Table 3), which means that each participant should complete four experimental operations under different level combinations of two PSFs.

Table 3. Scenarios needed to be attended

Time of day	Stress	Scenario
Active	SI = 1	S1
Active	SI = 1.5	S2
Negative	SI = 1	S3
Negative	SI = 1.5	S4

4.2 Data Collection and HEP Calculation

4.2.1 Data Collection

According to the Excel table output from the experiment and the given error criteria, the number of human errors per person can be obtained, the total number of errors in every level combination of the two PSFs can also be counted. The data acquisition table is shown in Table 4.

Table 4. Data acquisition table

State of the factors	Time of day	Active		Negative	
	Stress	SI = 1	SI = 1.5	SI = 1.0	SI = 1.5
Scenario		S1	S2	S3	S4
Number of timing errors	Action too early	(2,1)	(2,0)	(1,0)	(5,3)
	Action too late	(12,14)	(18,25)	(22,23)	(35,33)
	Total	(14,15)	(20,25)	(23,23)	(40,36)
Number of omissions	Omit entire step	(1,0)	(0,0)	(0,0)	(1,0)
	Omit operation in step	(0,1)	(1,0)	(2,0)	(1,3)
	Total	(1,1)	(1,0)	(2,0)	(2,3)
Number of human error	Sum of errors	(15,16)	(21,25)	(25,23)	(42,39)
	Total	31	46	48	81
	Average	15.5	23	24	40.5

HEP Calculation. During the whole fuel filling operation, any kind of error mode should be considered as human error that failed filling task operation. Therefore, the logical relationship diagram of errors can be obtained (Fig. 4).

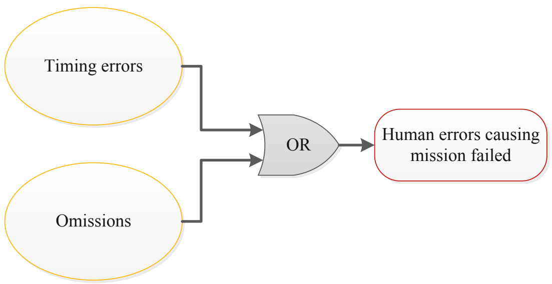


Fig. 4. Logical relationship diagram of errors

In the experiment, the authors can count the number of human errors of each participant. Assuming that N_i human errors occurred in a scenario on a participant, then the authors know the human error probability (HEP_i) of that participant [17].

$$HEP_i = \frac{N_i}{18 + 17} = \frac{N_i}{35} \quad (1)$$

Assuming that there are k participants in the experiment, the authors can get the total human error probability (HEP_t) in a scenario.

$$HEP_t = \frac{\sum_1^k N_i}{k \times 35} \quad (2)$$

According to Eq. (2), the authors can get the probability of the two error modes and HEP_t table in every scenario (Table 5).

Table 5. HEP table

Time of day	Stress	Scenario	Probability of timing errors	Probability of omissions	HEP_t
Active	SI = 1	S1	0.081	0.006	0.044
	SI = 1.5	S2	0.125	0.003	0.066
Negative	SI = 1	S3	0.128	0.006	0.069
	SI = 1.5	S4	0.211	0.014	0.116

5 Data Analysis and Discussion

The effect of two PSFs of the experiment can be analyzed on the basis of the statistics above. In this paper, Analysis of Variance (ANOVA) method [9] was used to analyze human error.

5.1 ANOVA Considering Interaction Effect

In order to study whether there is interaction between time of day and stress, [12] this paper used variance analysis method considering interaction to analyze the data (Table 6).

Table 6. Variance analysis table of interaction

Category	SS	df	MS	F	P-value	F crit*
Stress	288	1	288	76.8	0.00093	7.71
Time of day	388	1	388	90.13	0.00069	7.71
Stress*TOD	40.5	1	40.5	10.8	0.030	7.71

* $\alpha = 0.05$

According to Table 6, it can be seen that the interaction between stress and time of day is significant ($F(1,4) = 10.8$, $P < 0.05$), while the main effects of stress ($F(1,4) = 76.8$, $P < 0.05$) and time of day ($F(1,4) = 90.13$, $P < 0.05$) are also significant, indicating that both stress and time of day have effects on human error.

5.2 Simple Effect Analysis

In order to clarify the relationship between each PSF and human error, the authors analyzed the influence of one PSF on human error at a certain level of another PSF, and got the table of variance analysis as follows.

According to Table 7, when SI = 1 ($F(1,2) = 57.8$, $P < 0.05$), different levels of time of day have a significant effect on human error probability; when the pressure becomes 1.5 ($F(1,2) = 49$, $P < 0.05$), different levels of time of day also have a great effect on human error.

Table 7. Simple effect ANOVA table

Category	SS	df	MS	F	P-value	F crit*
Time of day						
SI = 1	72.25	1	72.25	57.8	0.017	18.51
SI = 1.5	306.25	1	306.25	49	0.020	18.51
Stress						
Active	56.25	1	56.25	12.24	0.068	18.51
Negative	272.25	1	272.25	83.77	0.012	18.51

* $\alpha = 0.05$

When the level of time of day is negative ($F(1,2) = 83.77$, $P < 0.05$) different stress levels have a significant effect on human error. But when the level of time of day changes to be active, according to the analysis of variance ($F(1,2) = 12.24$, $P > 0.05$), it can be said that when the stress level changes from SI = 1 to SI = 1.5, the number of human error does not increase significantly. That is, the change of stress level has no significant effect on HEP when the level of time of day is active. According to the significant interaction between time and stress revealed in 5.1, it can be concluded that different time levels can change the effect of stress on human error.

The fitting chart of human error probability with two PSF-two level is shown in Fig. 5. It can be seen that when the level combination of two PSFs is active-SI = 1, HEP is the lowest, indicating the best level combination of two PSFs. However, when the combination is negative-SI = 1.5, human error probability is the highest, showing that this is the worst level combination of PSFs.

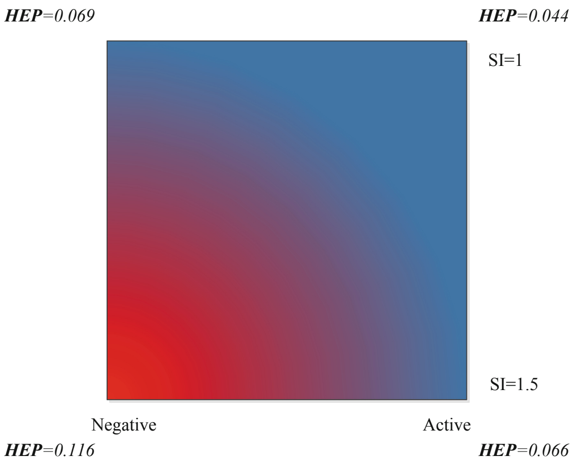


Fig. 5. Fitting chart of HEP at different level combinations

6 Conclusion

In this paper, a simulation experiment was designed for studying the effect of two PSFs —time of day and stress on rocket fuel filling mission. Through this simulation filling experiment, the authors analyzed the influence of time of day and stress on human error.

As for the relationship between time and stress, the authors found out that there is a significant interaction between this two PSFs throughout the interactive analysis of variance. The setting of time level can change the effect of stress level on human error probability. Therefore, when the operator's stress is difficult to adjust, the effect of stress on HEP can be reduced by changing operator's operating time. Through this experiment, the optimal combination of factor levels to minimize HEP can be determined, that is, when the stress is normal and the time is active there will be the least human errors.

Through one-way ANOVA, the authors discovered that when time level was active, human errors occurred less than when the time level was negative. It was also found that operator performed better when the stress level was normal than when the stress level was high.

Besides, it can be observed that the number of timing errors especially action too late is much more than omissions, which means that timing error is a more significant error mode. Thus, measures ought to be taken to reduce timing errors in particular.

In the experiment, the authors only introduced two PSFs and two kinds of human errors and considered that timing errors appeared too often while omissions only occurred occasionally. Therefore, the design of the experiment should be modified and a bit more PSFs and error modes should be introduced to get a better and more convincing result in human error analysis.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grants No. 71171008/No. 71571004, the National Key Research and Development Program of China, the Special Scientific Research Project for Civil Aircraft of the Ministry of Industry and Information Technology.

References

1. David, M.: A critical review of human performance reliability predictive methods. IEEE Trans. Reliab. **R-22**, 116–123 (1973)
2. Groth, K.M., Mosleh, A.: A data-informed PIF hierarchy for model-based human reliability analysis. Reliab. Eng. Syst. Saf. **108**, 154–174 (2012)
3. Calhoun, J., Savoie, C., Randolph-Gips, M., Bozkurt, I.: Human Reliability analysis in spaceflight applications. Qual. Reliab. Eng. Int. **29**, 869–882 (2013)
4. Barry, K.: Human error identification techniques for risk assessment of high-risk systems-Part 2 towards a framework approach. Appl. Ergon. **29**, 299–318 (1998)
5. Osipov, V.V., Daigle, M.J., Muratov, C.B., Foygel, M., Smelyanskiy, V., Watson, M.D.: Dynamical model of rocket propellant loading with liquid hydrogen. J. Spacecraft Rockets **48**, 987–998 (2011)
6. Charlie, G., Sriram, N., Matthew, J.D., Walter, H., Robelt, J., Barbara, B.: Applying model-based diagnosis to a rapid propellant loading system. In: 20th International Workshop on Principles of Diagnosis, Stockholm (2009)
7. Yoshikawa, H., Wu, W.: An experimental study on estimating human error probability (HEP) parameters for PSA/HRA by using human model simulation. Ergonomics **42**, 1588–1595 (1999)
8. Musharraf, M., Bradbury-Squires, D., Khan, F., Veitch, B., MacKinnon, S., Imtiaz, S.: A virtual experimental technique for data collection for a Bayesian network approach to human reliability analysis. Reliab. Eng. Syst. Saf. **132**, 1–8 (2014)
9. Loudoun, R., Allan, C.: The effect of time of day on injury patterns amongst adolescents in Australia. Appl. Ergon. **39**, 663–670 (2008)
10. Galy, E., Melan, C., Cariou, M.: Investigation of task performance variations according to task requirements and alertness across the 24-h day in shift workers. Ergonomics **51**, 1338–1351 (2008)
11. Edwards, B.J., Lindsay, K., Waterhouse, J.: Effect of time of day on the accuracy and consistency of the badminton serve. Ergonomics **48**, 1488–1498 (2005)
12. Ruger, M., Gordijn, M.C., Beersma, D.G., de Vries, B., Daan, S.: Time-of-day-dependent effects of bright light exposure on human psychophysiology: comparison of daytime and nighttime exposure. Am. J. Physiol. Regul. Integr. Comp. Physiol. **290**, R1413–1420 (2006)
13. Foy, H.J., Chapman, P.: Mental workload is reflected in driver behaviour, physiology, eye movements and prefrontal cortex activation. Appl. Ergon. **73**, 90–99 (2018)
14. Wang, L., He, X., Chen, Y.: Quantitative relationship model between workload and time pressure under different flight operation tasks. Int. J. Ind. Ergon. **54**, 93–102 (2016)
15. Hollnagel, E., Kaarstad, M., Lee, H.C.: Error mode prediction. Ergonomics **42**, 1457–1471 (1999)
16. William, B.R., Senior, M., Sandra, H.R.: Analysis and classification of human error. IEEE Trans. Syst. Man Cybern. **SMC-13**, 539–549 (1983)
17. Kim, Y., Park, J., Jung, W., Jang, I., Hyun Seong, P.: A statistical approach to estimating effects of performance shaping factors on human error probabilities of soft controls. Reliab. Eng. Syst. Saf. **142**, 378–387 (2015)

Organizational Risk Dynamics Archetypes for Unmanned Aerial System Maintenance and Human Error Shaping Factors

Yi Lu[1(✉)], Huayan Huangfu[1], Shuguang Zhang[2], and Shan Fu[1]

[1] Department of Automation, School of Electronic and Electrical Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China
luyil@sjtu.edu.cn
[2] School of Transportation Science and Engineering, Beihang University, Beijing 100191, China

Abstract. As revealed by the mishap causal factor statistics, human errors pose more threats to the safe operation of Unmanned Aerial System (UAS). Moreover, the number of human error induced maintenance accident has risen to a comparable level as the accidents due to flight crew error, but little prior research on the causality analysis can be found, especially consider the organizational context of human performance shaping factors. Based on the System Dynamics approach, this study proposed hierarchical risk archetypes that model the interactions of organizational, human and physical system factors leading to maintenance accident of large UASs. The archetypes help to clarify why technical reliability improvement measures, career training and accident investigation always fail to gain expected safety benefits. As organizational risk assessment tools, more detailed quantitative SD model can be developed based on those archetypes to evaluate potential safety policy and management decisions in the field of large UAS maintenance.

Keywords: Large UAS maintenance · Risk archetype · Human factors · System dynamics model

1 Introduction

With the wide application of the large Unmanned Aerial System (UAS) constructed by the Unmanned aerial Vehicle (UAV), ground control station, data links and recovery & launch system (with the maximum take-off weight above 599 kg, referring to the Category III specified by US DOD and FAA), their safety problems have attracted more attentions around the global: (1) the limited field of pilot view and spatial cognition have induced a large number of accidents, the ratio of this causal factor is 50% higher than in manned aircraft; (2) the high-strength tasks and lower maintenance cost have intensified the human factors in maintenance to be one of the most common causes of UAS accidents [1–3]. According to the statistics of the US Office of the Secretary of Defense (OSD) in 2003, the average Class A mishap rate per 100,000 h of the military large UASs was an order of magnitude higher than the manned aircraft [4]. From 2004 to 2006, 20% of Class A mishaps (i.e., causing the total loss valued above

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 75–87, 2020.
https://doi.org/10.1007/978-3-030-20037-4_7

$100 million or causalities) of the US Air Force (USAF) can be attributed to the MQ-1 Predator fleet which had 21 mishaps in total and 17 vehicles completely destroyed. Moreover, in the single year of 2015, although the number of the MQ-1 Class A mishaps continued a decreasing tendency from 12 (2010) to 7, it comprises about 57% of the total Class A mishaps of the USAF and a growth acceleration of the scale can be seen.

Meanwhile, due to the inherent advantages on low-cost compared to manned aircrafts, the civil large UASs have vast potentials for future development, such as coast guarding, geology exploring, filed investigate, etc. According to statistics since 2010, the size of global civil UAS market has grown to $100 billion and was growing year by year and the UAS regulator and industry were considering the integration of large UAS into national airspace [5, 6]. However, at present the safety level of civil UASs cannot satisfy the airworthiness requirement of 1 Class A mishap per 100,000 h, which challenges the UAS safety engineering in the future. Since 1990s, some researchers and organizations carried out statistical analyses over aviation accident risk factors and have obtained an overall trend: with the increasing of operation frequency and the accumulation of operation time, the reliability of technical system is continuously improved, which induces the accidents caused by human and organizational factors to become more and more obvious. Facing this, some theories on risk mechanisms involving non-technical factors were raised and has been applied on the large UAS accident analysis and safety improvement. For example, based on the Swiss Cheese Model (SCM), Wiegman and Shappell proposed the Human Factors Analysis and Classification System (HFACS) to analyze the aviation accident causality using the event-chain model as a framework [7]. This approach has gained a positive initial benefit on decreasing the aviation accident rate but its effects waned after a later on. In fact, accident analyses only based on general categorized errors and traditional safety philosophy were often superficial or structured, and the further effectiveness may experience a continuous descending trend with time. Most such risk identification methods were in hindsight. They confused the importance of factors and neglected the dynamic processes of risk transferring, especially in human and organizational levels [8, 9]. Due to lacking of safety evaluation techniques in medium and long term, in most cases the preventive measures only solved problems partially and also increased final costs. More importantly, the number of human error induced maintenance accident has risen to a comparable level as the accidents due to flight crew error, but little prior research on the causality analysis can be found, especially consider the organizational context of human performance shaping factors.

Properly understanding UAS maintenance safety risks in a systematic and dynamic way requires first understanding how and why this social-technical system migrate towards states of increasing risk. While individual UAS maintenance accidents usually have unique features at the surface, further accident investigation often reveals common organizational patterns that led to an increase in risk. By identifying these safety patterns, or risk archetypes, UAS maintenance organizations can better understand past accidents, monitor risk, and decrease the likelihood of future accidents. General organizational archetypes have been described by various authors in safety fields such as system dynamics. In risk analysis, these archetypes can help UAS maintenance organizations to understand the risk spectrum in different levels. In accident prevention, the archetypes can be used to develop dynamic models that predict the benefits of safety efforts in medium-and-long term.

2 Brief Overview of System Dynamics

System dynamics is grounded on the theory of nonlinear dynamics and feedback control, but also draws on cognitive and social psychology, organization theory, economics, and other social sciences to analyze complex system behavior [10]. It provides a framework for dealing with dynamic complexity. In the field of system safety, system dynamics has been used as an important supplement to analyze organizational accidents and proposed safety policy in the field of aviation, astronautics and chemical industries [11–13]. Especially, in a view of social-technical system, organizational accidents are increasingly being studied using system dynamics approach. It helps to model the risk interactions of organization safety with conceptual description, causation analysis and time-domain simulation tools.

The risk archetypes are constructed from three basic building blocks: the reinforcing loop, the balancing loop, and the delay.

2.1 Reinforcing Loop

It refers to a particular behavior that encourages similar behavior in the future and it corresponds to a positive feedback loop in control theory. As Fig. 1(a) shows, an increase of State 1 causes a positive consequence in State 2, as indicated by "+", which then causes an increase in State 1. For an example of positive consequences, the increase of training investment can improve the maintainer skills and on-the-job performance. The reinforcing loop can also be applied to negative consequences.

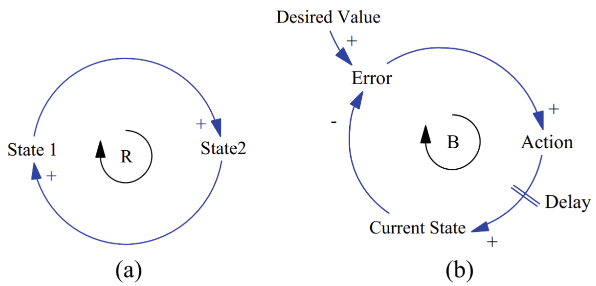


Fig. 1. Reinforcing loop (a); balancing loop and delay (b).

2.2 Balancing Loop

It exists when a particular behavior attempt to move from a current state to seek balance. It corresponds to a negative feedback loop in control theory, as Fig. 1(b) shows. The driving force in the loop is the size of gap between the goal and current value. For example, facing the gap between actual technical system reliability (it is limited by national industrial level) and its goal (it is required in design specification), the design modification and maintenance procedure revision will be promoted, which help to reduce such gap.

2.3 Delay

It is used to model the time the actions need to take effect and may result in unstable system behavior. It is indicated by a double line as shown on the balancing loop in Fig. 1(b). Caused by the delays, actions are deemed unsuccessful prematurely to achieve expected results. For example, due to maintainer experience needs time to accumulate, maintenance organizations always obtain a delayed training benefit. Delays can occur within both balancing and reinforcing loops.

2.4 Modeling Process

In this study, the critical risk factors embedded roots in large UAS development-operating-maintenance (DOM) processes. The data supporting risk analysis include:

1. Engineering assumptions grounded in practical experience and accident investigation related to DOM processes flaws.
2. Organization behavior modes and safety features proposed in literatures reviews, such as accident and risk models.
3. Accessible safety data, such as human factors in DOM processes identified in accident statistics.

Based on the above information, the intended modeling hierarchies and critical variables are initially defined. Then the Causal Loop Diagrams (CLDs) are developed to draw the critical interactions involving the underlying UAS accident causation, which form the risk dynamics archetypes. The reinforcing and balancing feedback loop (R/B) structure based conceptual model is used to describe the dynamic influences of organization, human resource and technical system factors on UAS catastrophic accident.

3 Data Sources and Analysis Methods

3.1 Data Sources

The UAS have a characteristics of broad spectrum of type and operation field. The difference exists in size, weight, range, flight speed, payload and mission performance. The early operation of UAS was implemented by military planners to carry out reconnaissance and/or attack missions. Beware of UAS's convenience and low-cost characteristics, the use of UAS was rapidly expanding to more civil areas, such as disaster relief, environmental conservation, filmmaking or cargo transports. Under such background, civil UASs increased dramatically and they now have outnumbered military UAVs vastly, with estimates of over several millions per year. However, the widespread UASs threaten airspace security in numerous ways, including unintentional collisions with population and other manned aircrafts. Indeed, in terms of quantity, most members in civil drone family are quadcopter types or short scale fixed wing designs operated under direct visual line of sight, which have relatively low impact energy and limited remote range. The most significant risks to public safety or mishap loss come from the operations of those large UASs with bigger size/weight, long-endurance, high speed and payload features and the majority are military types even derived civil types with similar configurations, such as the MQ-1 Predator series. They remain as an overarching concern for most Aviation Authorities worldwide when considering their potential integration with manned airspace in the future.

Based on those open sources, many causation analyses of large UAS accident were implemented [14–17]. For example, Tvarynas et al. found the frequency of human factors mishaps of US military large UAS fleet was increasing, with data on UAS mishaps during fiscal years 1994–2003 [1]. In the UAS type spectrum, the MQ-1 Predator UAS of the General Atomics Aeronautical Systems, Inc. is a medium altitude, long-endurance vehicle and the largest current generation UAS in service with the U.S. military. Predators entered the U.S. Air Force (USAF) inventory in 1995, and it flight hours increased rapidly, expanding more than 100 times in the two decades from 1996 to 2017. Nullmeyer et al. used the USAF MQ-1 Predator Class A mishaps as a case study and derived flight crew training measures [15]. Consequently, MQ-1 Class A mishaps attribute to human error (flight crew and maintainer) decreased despite increasing numbers of mishaps overall. Especially, USAF Sustainment Center generates mishap investigation reports for typical UAS for every Class A mishap by fiscal year and by UAS type and provides results at varying levels of granularity.

3.2 Framework of Risk Analysis

As indicated by literatures and high-risk organization accident investigation reports, human factors in aviation are a complicated concept including human physiology, psychology (perception, cognition, memory, social interaction, error, etc.), work place design, environmental conditions, human-machine interface and anthropometrics. It can also be divided into individual factors and group cooperation factors (i.e., Crew Resource Management, CRM). Based on the Swiss Cheese Model (SCM), the Human Factors Analysis and Classification System (HFACS) has built a bridge between theory and practice by accommodating human factors in aviation in a more systematic way [7]. The model considers all aspects of human errors, including the conditions of operators and organizational failures. It divides aviation accident related human factor into four categories which form a hierarchical structure, namely Unsafe Acts, Preconditions for Unsafe Acts, Unsafe Supervisions, and Organizational Influences. In this study, to identified UAS maintenance safety related risk factors, the HFACS level is adopted to determine the source of factors and the logic sequences beneath them, as Table 1 shows. Moreover, we also consider the risks of technical system level in this framework.

In this table, the category item *3.4 Mission Maintainer Experience* represents the knowledge, skill and attitudes of maintainer which determine the proficiency of maintainer when implementing required tasks, e.g., replace system components following specified intervals and technical procedures. The category item *4.1 Mistakes*

Table 1. Factors identification framework for UAS maintenance risk analysis.

No.	UAS maintenance risk factors	HFACS levels
1	1.1 Mission tempo pressure 1.2 Mission cancellation 1.3 Class-a mishap rate 1.4 Actual flight sorties	Organizational Influence
2	2.1 Scheduled total mission duration 2.2 Gaps between scheduled and actual UAS number 2.3 Maintainer employ requirement 2.4 Maintainer trainer employ requirement 2.5 Revealed system failures 2.6 System design modification 2.7 Maintainer trainer population 2.8 Accident investigation report	Unsafe supervision
3	3.1 Maintenance procedure modification 3.2 Gap of maintainer population 3.3 Maintainer total experience 3.4 Mission maintainer experience (KSA) 3.5 Known system failures	Preconditions for unsafe acts
4	4.1 Mistakes (M) 4.2 Skill-based errors (E) 4.3 Procedure violations (V)	Unsafe acts
5	5.1 System operation risk 5.2 Adverse system interactions	System malfunctions
6	6.1 Scheduled mission duration of single system 6.2 Gaps between required and actual system reliability 6.3 System failures 6.4 System operation risk	Component failures

include two types, rule-based mistakes and knowledge based mistakes. And the category item *4.2 Skill-based Errors* (perception, memory failure, fault-based errors) and *4.3 Procedure violations* (pattern, scenario-based violations) also have their components.

This trend shows: although the probability-based risk theories truly have the initial effects on risk elimination, their component-failure based view-point still embedded roots in static and linear safety philosophy and can hardly address accident causality involving interactive complexity. Of the more than 100 Class-A mishaps occurring during the period of fiscal years 1996–2017, 65.4% involved the en-route phase and 57 mishaps involved human error factors. It should be noticed that, the majority of human factors-related problems should be attributed to maintenance error and the critical technical system related mishaps involved the propulsion system failures. In fact, with the traditional statistics method of accident causal factors, it is hard to distinguish the different failure causes rooting in the flawed system design and/or inadequate maintenance activities. For most accident investigation report, accident was often classified as belonging to both "propulsion system failure" and "maintenance error". In this term the causality revealed by the accident investigation was ignored. In fact, human error causal factor often cannot be explained by the single point failure. According to Marais et al. (2006), accident analysis based only on categorized errors and traditional safety philosophy is superficial or structured for future preventive measures [9]. In order to analyze the dynamic interactions and impact of organizational risk factors, the system dynamics method is introduced to analyze the UAS maintaining risk mechanisms in terms of feedbacks.

4 The Archetypes

This section proposes and discusses the UAS maintenance risk dynamics archetypes in a view of hierarchical framework.

4.1 STAMP-Based UAS Safety Risk Analysis

Based on the Systems-Theoretic Accident Model and Process (STAMP) and STAMP-based hazard analysis technique (STPA) raised by Leveson, accident is a property emerging from the interactions within the social-technical system components rather than a sequence of events linked by static cause and effect factors [18, 19]. Safety is reformulated as a control problem rather than a reliability problem. Unlike the traditional methods using event-chain as basic element for risk analysis, STAMP/STPA approach uses the hierarchic safety control structure as a guide to identify the control flaws resulting in adverse interactions between safety actors which violate specified safety constraints. The derived safety control measures help to handle the identified unsafe scenarios of complex systems in a systematic way. With the view of the safety control structure (SCS) derived from the STAMP causality model, the large UAS safety risk related development-operation-maintenance (DOM) processes can be described as shown in Fig. 2.

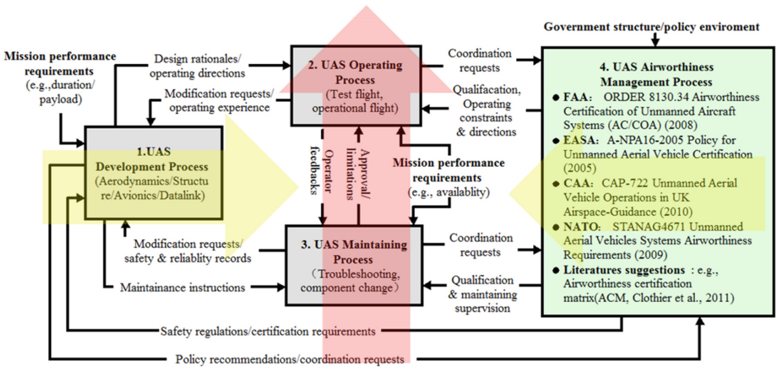


Fig. 2. Safety control structure (SCS) of large UAS development-operating-maintenance (DOM) processes

In this structure, the risk effects of the large UAS operating process and the UAS maintaining process on UAS safety are regarded as the local workplace factors, as the red arrow shows. Meanwhile, the operational UAS is generated by the UAS development process and is supervised to operate normally within specified safety constraints by the UAS airworthiness management process. The risk effects of these upper factors are represented in terms of the interactions between those processes, shown by the yellow arrows in Fig. 2.

4.2 Emergency and Organizational Level

On the top level of large UAS maintenance risk dynamics archetypes, the Emergency Level archetype (EL) represents the indicators of UAS safety and mission availability (e.g., mission abort and cancellation caused by maintenance issues). In this level, some risk factors form the decision base or activity goal for organization management so the factors in Organizational Level archetype (OL) are involved in one frame for better representation of their interactions, as shown in Fig. 3.

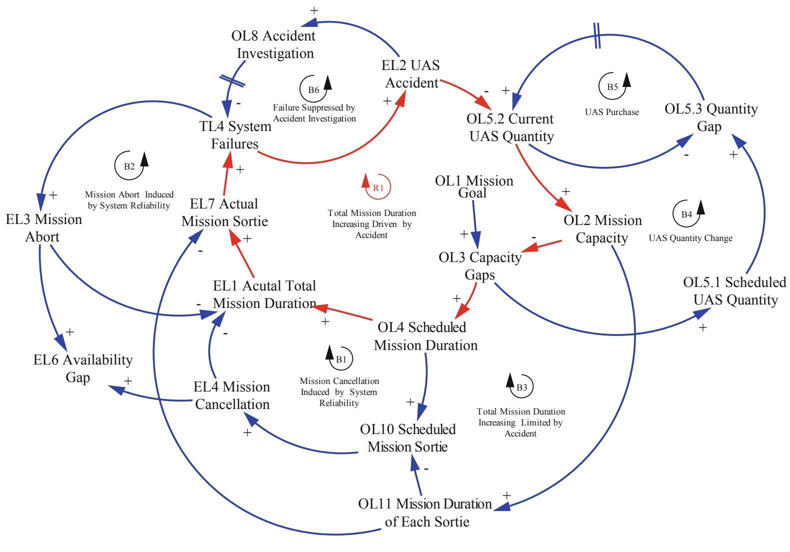


Fig. 3. UAS maintenance risk archetype in the emergency and organizational levels

In this figure, six balancing loops (B_1–B_6) and one reinforcing loop (R_1) are identified. They can be categorized into three groups as following:

1. $R_{1\text{-OL}}$ (EL_1-EL_2-OL_2), $B_{4\text{-OL}}$(OL_2-$OL_{5.1}$-$OL_{5.2}$) and $B_{5\text{-OL}}$($OL_{5.2}$-$OL_{5.3}$)

In this group, the critical node variables are OL_2 and $OL_{5.2}$. The balancing loop B4 and B5 describe the organizational activity that purchases new UAS to compensate the mission capacity affected by large UAS mishaps. Meanwhile, the organization also increase the scheduled mission duration (OL_4) to compensate the mission capacity under reduced UAS quantity. However, due to the increasing of OL_4, the causal link

$OL_4 \xrightarrow{+} EL_2$ may suppress the expected causal link $OL_4 \xrightarrow{+} OL_2$. The reinforcing loop R_1 explains such vicious spiral phenomenon which often appeared in the USAF MQ-1 Predator UAS operation history and were revealed by some accident investigation reports.

2. $B_{1\text{-}OL}$ (EL_4-EL_2-OL_4), $B_{2\text{-}OL}$ (TL_4-EL_3-EL_1) and $B_{3\text{-}OL}$ (OL_{11}-EL_4-EL_7-OL_2)

In this group, the critical node variables are EL_1, TL_4 and OL_{11}. Under the increasing of scheduled mission sorties (OL_{10}), the system reliability problems aggravate the effects of balancing loop B_1 and B_2 which induce the gaps between actual and scheduled mission sorties. Moreover, facing the increasing accidents (EL_2) the organization reduces mission duration of each sortie (loop B_3), which overlay on loop B_1 and possess adverse effects on actual total mission duration (EL_1). It reveals the underlying cause that how UAS accidents influenced the availability of large UAS fleet in a long-term vision.

3. $B_{6\text{-}OL}$ (OL_8-TL_4-EL_2)

This balancing loop describes the source of hindsight safety efforts raised by organizational accident investigation which plays an important role in connecting the communication between UAS development and maintenance processes.

4.3 Human Level

In this level, the risk interactions between organizational and human level are focused. This archetype modelled their causations in two aspects: (1) the effects of risk factors on maintainer on-the-job experience; (2) the effects of maintainer experience on technical system reliability. As shown in Fig. 4, three balancing loops (B_1–B_3) are identified. They can be categorized into two groups as following:

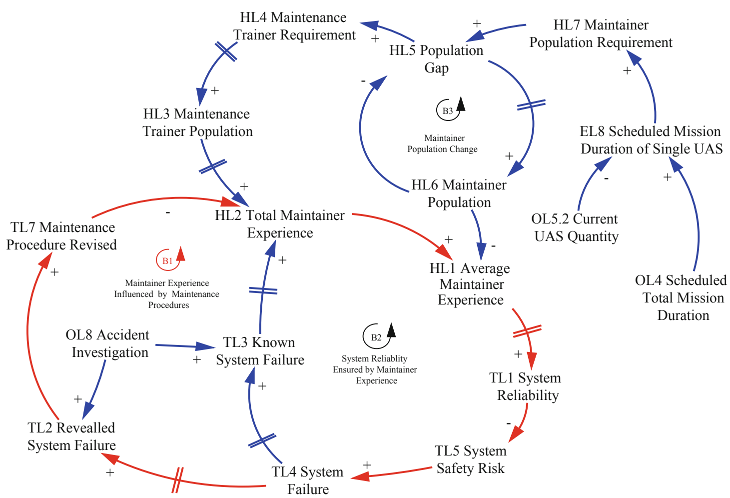


Fig. 4. UAS maintenance risk archetype in the human level

1. $B_{1\text{-HL}}$ (HL_1-TL_7-HL_2) and $B_{2\text{-HL}}$ (TL_4-HL_3-TL_1)

In this group, the critical node variables are TL_4 and HL_2. The two loops have similar model structure: the accident investigation can enforce the modifications of maintenance procedure (TL_7). In long-term view, the system reliability can be improved due to the operation of loop B_2, but such procedure modification can also reduce the familiarity of maintainer on current tasks. The loop dominance determines the effect of risk factors in the Human Level on system reliability.

2. $B_{3\text{-HL}}$ (HL_5-HL_6)

This group induced a delay to model the variation process of maintainer population in a simplified way. As the goal of the balancing loop B_4, the maintainer population requirement (HL_7) is determined from scheduled mission duration of single UAS (i.e., mission tempos specified in most literatures). Meanwhile, the population gap enforces the organization to employ more maintenance trainer (HL_3) to ensure the maintainer proficiency. Whether the maintainer can get adequate training resource determines the average maintainer experience which is also an important indicator to evaluate a UAS fleet's training investment. Meanwhile, due the multiple delay links between those variables, the input-output relation characterizing the causation between mission requirement and maintainer experience is non-linear, which always influences the organizational decision-making on safety issues.

4.4 Technical System Level

In this level, the risk interactions among technical systems are focused. As shown in Fig. 5, two balancing loops (B_1–B_4) and one reinforcing loop (R_1) are identified.

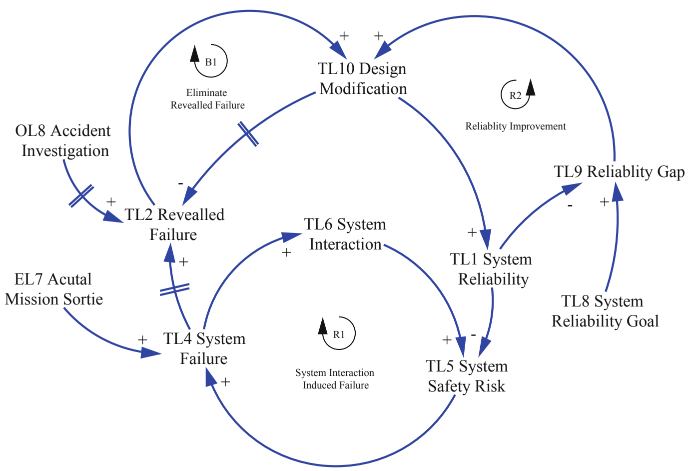


Fig. 5. UAS maintenance risk archetype in the technical system level.

In this level, the critical node variables are TL_4, TL_5 and TL_{10}. Similar to the risk dynamics in the Human Level, the accident investigation can also enforce the modifications of system design (TL_{10}), which initializes the operations of two balancing loops (B_1 and B_2): (1) system reliability improvement (e.g., system redundancy added and/or component quality enhanced), regarding the reliability gap revealed in missions; (2) system failure mitigation (e.g., unexpected system behavior control to reduce improper system design features). To characterize the system reliability of large UASs, some conceptual indicators have been used, for example, the Mean Time Between Failure (MTBF). Regarding the enhanced system reliability, the revealed system failures begin to reduce and the effects of accident investigation on UAS safety improvement become less prominent. This is the reason why the classical organizational behavior "trial-and-error" always fails when operating a UAS for a long time. Without systematic and dynamics view on organizational safety risk mechanism, the organizations only relaying on regular experience accumulated in routine operations often cannot handle their safety situation when new risks appear, such as maintainer population gap, high mission tempo and new system technical features.

5 Conclusion

This study introduces organizational risk dynamics archetypes to describe the maintenance safety risk mechanisms of large UASs. Based on the HFACS framework for factor identification and STAMP model for risk causation analysis, the proposed archetypes integrate the risk factors involving organizational, human and technical dimensions rather than identifying static accidental factors in textual way, when regarding the operation safety of large UASs is the emergency property of a social-technical system which performs the development-operation-maintenance processes of large UASs. The risk archetypes explain:

1. The variation trends of loop domination determine the risk level of UAS maintenance organization. Increasing scheduled mission sorties and durations are always the instinctive response of the organization to cope with the UAS mishaps, considering the low operation cost features of UAS compared to manned aircrafts.
2. The maintenance procedure modification derived from UAS accident investigation may introduce adverse effects on maintainer mission experience and it is the reason why such hindsight may suppress safety benefits of the training investments. The similar risk archetype also exists in the technical system level: with the reduced system failures that can be revealed in accident, the improvements on system reliability expected by system design modification always encountered bottleneck.
3. Due the multiple delay links between mission requirement and maintainer experience, the input-output relation to characterize their causation is non-linear. It influences the organizational decision-making in the aspect of human resource and forms the human error shaping factors affecting UAS maintenance safety.

Acknowledgments. This research was sponsored by the National Science Foundation of China (No. 61803263). The first author also thanks Professor Karen B Marais for helpful advice during the process of this research.

References

1. Tvaryanas, A., Thompson, W., Constable, S.: Human factors in remotely piloted aircraft operations: HFACS analysis of 221 mishaps over 10 years. Aviat. Space Environ. Med. **77** (7), 724–732 (2006)
2. ICAO: Unmanned aircraft systems (UAV), 1st edn. ICAO (2011)
3. Lu, Y.: Feedback process-based UAV system accident analysis and system safety dynamics investigation. Ph.D. thesis. Beihang University, Beijing (2015). in Chinese
4. Schaefer, R.: Unmanned aerial vehicle reliability study, OSD UAV reliability study. Technical report, Office of the Secretary of Defense, Washington, DC (2003)
5. Li, W.-J.: Unmanned Aerial Vehicle Operation Management. Beihang University Press, Beijing (2011). (in Chinese)
6. Ramalingam, K., Kalawsky, R., Noonan, C.: Integration of unmanned aircraft system (UAS) in non-segregated airspace: a complex system of systems problem. In: Proceeding of the 2011 IEEE International Systems Conference, pp. 1—8. IEEE Press, New York (2011)
7. Wiegmann, D.A., Shappell, S.A.: A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System, pp. 45–56. Ashgate, Burlington (2003)
8. Wolstenholme, E.F.: Toward the definition and use of a core set of archetypal structures in system dynamics. Syst. Dyn. Rev. **19**(1), 7–26 (2003)
9. Marais, K.B., Saleh, J.H., Leveson, N.G.: Archetypes for organizational safety. Saf. Sci. **44** (7), 565–582 (2006)
10. Sterman, J.D.: Business Dynamics: Systems Thinking and Modeling for A Complex World. Irwin/Mac-Graw Hill, Boston (2000)
11. Lu, Y., Zhang, S.-G., Hao, L., Huangfu, H.-Y., Sheng, H.: System dynamics modeling of the safety evolution of blended-wing-body subscale demonstrator flight testing. Saf. Sci. **89**, 219–230 (2016)
12. Dulac, N., Owens, B., Leveson, N.G.: Demonstration of a new dynamic approach to risk analysis for NASA's constellation program. MIT CSRL Final Report to the NASA ESMD Associate Administrator (2007)
13. Bouloiz, H., Garbolino, E., Tkiouat, M., Guarnieri, F.: A system dynamics model of behavioral analysis of safety conditions in a chemical storage unit. Saf. Sci. **58**(1), 32–40 (2013)
14. Williams, K.W.: A summary of unmanned aircraft accident/incident data: human factors implications. Technical report, Civil Aerospace Medical Institute, FAA, Oklahoma City, OK (2004)
15. Nullmeryer, R.T., Herz, R., Montijo, G.A.: Training interventions to reduce air force predator mishaps. In: 15th International Symposium on Aviation Psychology, Dayton, OH (2009)
16. Nullmeryer, R.T., Herz, R., Montijo, G.A., Leonik, R.: Birds of prey: training solutions to human factors issues. In: 10th the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), Dayton, OH (2007)

17. DOD: Report to congress on future unmanned aircraft systems training, operation, and sustainability. Technical report, Under Secretary of Defense for Acquisition, Technology and Logistics, Department of Defense, Washington, DC (2012)
18. Leveson, N.G.: Engineering a Safer World. MIT Press, Cambridge (2012)
19. Leveson, N.G.: A new accident model for engineering safer systems. Saf. Sci. **42**(4), 237–270 (2004)

Time-Reliability Correlation for the Human Reliability Analysis of a Digitalized Main Control Room

Wondea Jung(✉) and Jinkyun Park

Korea Atomic Energy Research Institute,
111, Daedeok-Daero 989, Yuseong-gu, Daejeon 34057, Korea
{wdjung, kshpjk}@kaeri.re.kr

Abstract. This paper introduces a method for the simulator data collection and analysis that we have developed to generate a time-reliability correlation (TRC) for the human reliability analysis (HRA) of an advanced control room with digitalized human system interfaces. We provide guidance on the definition of TRC and the timing points to be collected, as well as the processing of collected data. An application study was carried out using a set of simulator records collected from a reference plant to confirm the suitability of the proposed method and to generate a preliminary TRC for the HRA of the reference plant.

Keywords: Human reliability analysis · Time-reliability correlation · Simulator data

1 Introduction

Numerous human reliability analysis (HRA) methods including THERP (technique for human error rate prediction) [1] apply the concept of time-reliability correlation (TRC) to evaluate diagnostic human error probability (HEP). The TRCs of those methods were developed to support the HRA of a nuclear power plant (NPP) that installed analog type human system interfaces (HSIs), which were produced based on expert judgment or limited simulator data around 30 $\sim$ 40 years ago in the U.S. The operating environment of a new NPP, however, has been considerably changed by applying digital technologies to the design of control room HSIs. Therefore, in order to support the HRA for a new NPP, it is necessary to generate a new TRC that reflects the design and operational characteristics of an advanced main control room (MCR) with digitalized HSIs.

The TRC of THERP has been used for over 30 years to evaluate the diagnosis HEP in the probabilistic safety assessments (PSAs) performed in Korea. However, the present operators' task performance of NPPs in Korea might be different from the performance in the U.S. at the time when the THERP TRC was developed. The design and operation of NPPs have continued to improve, and their safety capability has been enhanced by adding new safety features or modifying existing safety systems. From the viewpoint of HSIs design, which is closely related to human performance during

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 88–94, 2020.
https://doi.org/10.1007/978-3-030-20037-4_8

accident progression, most NPPs in Korea have installed the latest interfaces when compared to most NPPs in the U.S. of about 40 years ago. In fact, Korea has a new NPP with fully digitalized HSIs that is in operation.

Therefore, the question has naturally been raised in Korea as to whether the TRC of THERP could be used for the HRA of the new NPP without validating its suitability [2]. To answer this question, it is necessary to develop a new TRC that reflects the design and operational characteristics of digitalized HSIs including a computerized procedure system (CPS), soft control (SC), and other influencing factors such as level of experience and training. This paper summarizes an approach of simulator data collection and analysis that we developed to generate a TRC for the HRA of a reference plant with fully digitalized HSIs. We propose guidance on the definition of TRC and the timing points to be collected, as well as the processing of collected data. An application study was carried out using a set of Audio-Video records secured from a full-scope simulator in the reference plant to confirm the suitability of the proposed method and to generate a preliminary TRC for the HRA of the reference plant.

2 TRCs Used in HRA

2.1 Review of Previous TRC Studies

The risks of NPPs are clearly correlated to the response reliability of the crew operating the plant during abnormal events or accidents. A few studies before the 1970s scrutinized the correlation of response times and human reliability that would be used to develop similar time-dependent models for hardware. Based on this, a few studies were conducted in the 1980s to collect the crew response time from nuclear plant simulators for various events in order to produce a TRC that can evaluate the HEP for event diagnosis [3].

Among several previous studies regarding TRC, three can be selected as being representative for TRC in HRA history. The first one was the TRC generated by Hall et al. [4] as part of the Operator Action Tree (OAT)/TRC method. Their curve was developed with the consensus of three experts in the operations, reliability, and human factors of a NPP. An error probability of 0.015 was assigned to 30 min and 0.0001 to 100 min based on rough estimates of industry-wide experience. The OAT/TRC method also suggested a limit to time dependent human reliability, closing the TRC at 0.0001 or 0.00001.

The second was the TRC of THERP developed by Swain and Guttmann [1]. In THERP, available time is used to produce the probability that a crew will not diagnose an abnormal event. The diagnosis failure probability decreases as the time from the event occurring increases and is assumed to be lognormal over time. The THERP TRC was derived from expert judgment considerations that were guided by some simulator data undertaken by General Physics and Oak Ridge.

The last one was the TRC of the HCR/ORE (human cognitive reliability/operator reliability experiments) method [5] developed by EPRI. The HCR model [6], which was the original version of HCR/ORE, provided three basic types of TRCs to quantify crew failure probability as a function of time, that is, one each for skill-, rule-, and knowledge-based behaviors. Although the HCR model has been widely used in PSAs, the question of its validity remained to be answered. EPRI performed a simulator experimental program called ORE to examine the validity of and modify the HCR model by using data from the simulators of NPPs. Finally, the HCR/ORE method provides a TRC represented by a normalized lognormal function with two parameters: the median and logarithmic standard deviation of crew response times.

2.2 TRC Used in K-HRA Method

In Korea, the K-HRA method [7] is used as a standardized method for the HRA of nuclear PSA. In K-HRA, it is assumed that the human error probability can be assessed by analyzing the diagnosis and execution errors separately. Here, diagnosis means a set of cognitive activities such as event perception, situation diagnosis, and response planning, and execution is the implementation of a planned response. Quantifications of diagnosis and execution error probabilities for a human failure event (HFE) can be conducted using the following equations:

$$\mathrm{HEP(HFE)} = \mathrm{HEP(D)} + \mathrm{HEP(E)}. \quad (1)$$

$$\mathrm{HEP(D)} = \mathrm{nominal\ \text{-}\ HEP(D)} \times \Pi \mathrm{Wi(PSFi)}. \quad (2)$$

$$\mathrm{HEP(E)} = \Sigma \mathrm{i}\,[\mathrm{nominal\ \text{-}\ HEP(Ei)} \times \mathrm{HEP(Ri)}]. \quad (3)$$

where, nominal-HEP(D) = f (available time for diagnosis),
nominal-HEP(Ei) = f (task type(i), stress level(i)),
HEP(Ri) = f (available time(i), HSI(i), supervisor recovery(i)).

In K-HRA, the HEP of a given HFE can be produced by summing the diagnosis HEP and the execution HEP. As shown in Eq. (2), the diagnosis HEP, HEP(D), is determined by a function of the available time (AT) of a required task and the weighting factors of the performance shaping factors (PSF). The nominal HEP of a diagnosis error (nominal-HEP (D)) can also be derived using the TRC from the THERP Handbook [1]. Thus, to calculate the HEP of an HFE using the K-HRA method, the AT of a relevant action should be determined. The AT of an action can be defined as the time duration from the occurrence of the related cue(s) to the maximum point in time when the operators should recognize the necessity of the action to successfully complete it. Accordingly, to estimate the AT of an action, three time points should be defined: maximum allowable time, cue time, and execution time for an action related to an HFE [7].

3 Method of Data Collection to Generate a TRC

To support the HRA of NPPs in Korea, we have collected simulator data to generate plant specific HRA data using HuREX (Human Reliability data Extraction) [8]. HuREX was developed as a framework for simulator data collection and analysis to generate HRA data such as HEP, performance time (PT), and correlations between PSFs and HEPs. As mentioned in Sect. 2, K-HRA uses the TRC of THERP to estimate nominal-HEP (D), which has become one of the technical issues whose validity should be proven. Among the various HRA data generated by HuREX, part of the PT data can be used to produce a plant specific TRC. Therefore, we now propose a method that collects and analyzes time data from simulators to develop a TRC.

The data obtained from simulator experiments include time required to detect and diagnose an event and formulate an appropriate response from the occurrence of the event. Then the time data are sorted with ascending order of time. Afterwards, the TRC (in other words, non-response probability) can be derived by the following equation:

$$\Pr(\mathrm{TRCi}) = \Pr(\text{response time} > \mathrm{ti}) = 1 - \mathrm{i}/(\mathrm{n}+1),\ \mathrm{i} = 1,\ 2,\ 3,\ \ldots,\ \mathrm{n}. \quad (4)$$

where, i is the i'th data point, ti is the i'th time in the ascending order of response time, and n is the total number of samples.

To generate a TRC, we first need to collect time data taken to diagnose a simulated event. Time data means the time purely taken by a crew to diagnose (signal perception, situation diagnosis, and response planning) a simulated event, referred as "DiagTime" in Fig. 1. To extract the time data from simulator experiments, we need to design simulation scenarios, including a few abnormal and/or emergency events, and collect the following time data for each event:

- Event time: Occurrence time of an event
- Cue time: Occurrence time of the first cue triggered by the occurred event
- Response initiation time: Time of starting operators' response to the occurred event

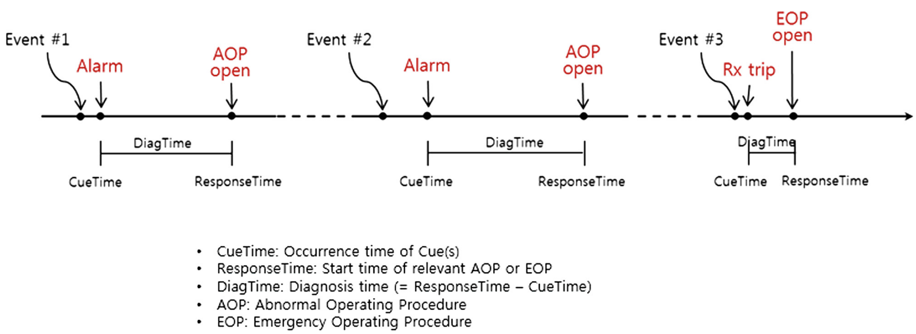


Fig. 1. Data points for extracting diagnostic times in a simulated scenario.

To collect the above diagnosis time information from simulator simulations, we developed a guideline for collecting and analyzing time information from simulator experiments. There are two ways to collect simulator data: (1) obtain records from physical equipment such as a simulator itself or a camcorder, and (2) record related performance data from an observer. All the information regarding the plant's operational status and crew's responses during simulation is recorded in various types of simulator logs: injected malfunctions, generated alarms, values and trends of key operational parameters, all manipulations performed by operators, state changes of components, and so on. Time data can be collected through analyzing the simulator logs regarding alarms, plant parameters, and operator responses that are recorded during simulations. In addition, we can confirm all the information from the operator responses during simulation from video records, specifically communication between operators, movement line of each operator, and manipulation of a specific component at a certain time point.

4 An Application Study

4.1 Data Collection

An application study was performed to show the feasibility of the proposed method. For a case study, we applied the proposed method to collect simulator data from the full-scope training simulator of a reference NPP in Korea. Data collection was carried out during the regular training programs in 2017 that require for the MCR operators to work at the reference plant. A total of 12 operating crews participated in the data collection, and simulator logs and records were secured for eight simulated scenarios, including 18 abnormal and emergency events. The simulated events and data used for the case study are summarized in Table 1.

Table 1. Simulated events and data points.

Event type	Scenario	# of data
Abnormal event	Leak of turbine control oil	6
	Flow control valve of RCP-01A sealing fails to open	6
	A VCT divert valve fails to open	6
	Turbine trip due to fire in a main transformer	6
	A valve in a letdown line fails to close	7
	Leak in a steam generator (SG) tube	7
	Turbine trip due to rupture in a condenser orifice	4
	Spurious auxiliary feedwater actuation signal (AFAS)	8
	Break of a normal drain valve in a feedwater re-heater	8
	Leak from a letdown line	7
	Earthquake (less than OBE)	7
Emergency event	RCP seal break (LOCA)	5
	Station blackout (SBO)	6
	Steam generator tube rupture (SGTR)	7
	Loss of offsite power (LOOP)	10
	Loss of all feedwater (LOAF) with reactor trip failure	9
	Loss of coolant from SI-652 internal break (ISLOCA)	12
	POSRV stuck open (LOCA)	7

4.2 Data Analysis

Of the total 128 data points, a TRC (as shown in Fig. 2) was generated using 125 data points after excluding three outliers. Two different TRCs were also produced, each for abnormal events (69 data points) and diagnosis events (56 data points), respectively.

The TRC of abnormal events shows a similar pattern with the TRC of Fig. 2, but the TRC of emergency events is different from the TRC shown in Fig. 2 or the TRC of abnormal events. The data shows that there is not much difference in response times for emergency events. We interpret it is because the way of emergency responses using EOPs after reactor trips is more formal than that in abnormal situations. Also, the number of data points from emergency events is smaller than that from abnormal events. According to basic statistics, the average diagnosis time for the emergency events was 9.1 min, which was larger than the average time for the abnormal cases, 4.6 min. However, the standard deviations were almost the same in both the emergency and the abnormal events, which were 2.9 and 2.8 min respectively.

Next, we compared the TRC obtained from the case study with the TRC of THERP. It shows that the probability of diagnosis failure estimated by the TRC of the reference plant is higher than the median of THERP TRC in the time interval from 1 to 10 min. However, after 10 min, it is almost similar to the THERP median. The probabilities of diagnostic failure in the TRC of THERP are defined as 1.0, 0.1, and 0.01 when the available times for diagnosis are 1, 10, and 20 min, respectively. However, even within 10 min, the TRC is smaller than the upper limit of the THERP TRC.

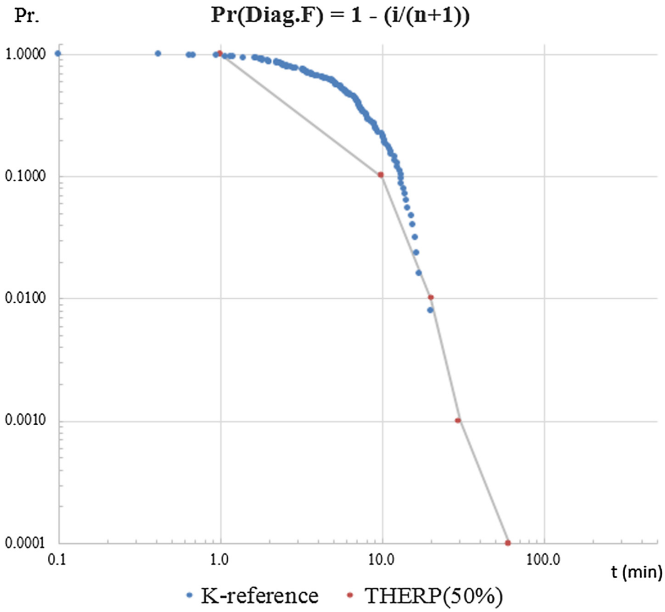


Fig. 2. TRC generated from the case study.

In the case study, all the times taken to diagnose an event were less than 25 min. It was harder than expected to collect timing data that were longer than 25 min due to the constraints of the simulator training program in the reference plant. Accordingly, it seems challenging to derive a TRC after 25 min from simulator data in the near future. However, from the TRC shown in Fig. 2, we can expect that the probability of diagnosis failure after 25 min would be lower than the median of the THERP TRC.

5 Conclusion

TRC is still an important technical basis when using a time dependent HRA method such as THERP, ASEP-HRA, or K-HRA method. The K-HRA method uses the TRC of THERP that was developed almost 40 years ago in the USA. Therefore, it is necessary to verify the suitability of THERP TRC by using plant specific operating experience or simulator data. To this end, we developed a method for collecting and analyzing simulator data to generate a TRC, which includes guidance on the definition of TRC and timing points to be collected, and the processing of collected data. An application study was carried out using a set of simulator records to confirm the suitability of the proposed method and to generate a preliminary TRC from the reference plant with digitalized HSIs. It showed that the probability of diagnosis HEP estimated by the TRC of the reference plant is higher than the median of THERP TRC in the time interval between 1 and 10 min. However, even within 10 min, the TRC was smaller than the upper limit of the THERP TRC. After 10 min, it was almost similar to the THERP's median.

As a result, the proposed method seems to be applicable to HRAs as a reasonably shaped TRC was successfully generated from the simulator data. As the diagnostic HEP estimated from the TRC herein is compatible with and more realistic than the existing TRCs, it is strongly expected that the proposed method will be a good starting point to enhance the quality of HRA results through the provision of a technical basis for estimating HEP from simulation data.

Acknowledgments. This research was supported by a Nuclear Research & Development Program of the National Research Foundation of Korea (NRF) grant, funded by the Ministry of Science and ICT (MSIT) [Grant Code: 2017M2A8A 4015291].

References

1. Swain, A.D., Guttmann, H.E.: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, NUREG/CR-1278. US NRC, Washington, DC (1983)
2. Jung, W., Kim, Y., Park, J.: Method for Developing a Diagnosis TRC based on Simulator Data in Nuclear Power Plants, KAERI/TR-7254/2018. Korea Atomic Energy Research Institute (KAERI), Daejeon, Republic of Korea (2018)
3. Dougherty, E.M., Fragola, J.R.: Human Reliability Analysis – A Systems Engineering Approach with Nuclear Power Plant Applications. Wiley, New York (1988)
4. Hall, R.E., Fragola, J.R., Wreathall, J.: Post Event Human Decision Errors: Operator Action Trees/Time Reliability Correlation, NUREG/CR-3010, US NRC (1982)
5. Spurgin, A.J., et al.: Operator Reliability Experiments using Power Plant Simulators, NP-6937, vol. 1. Electric Power Research Institute (EPRI), Palo Alto (1990)
6. Hannaman, G.W., Spurgin, A.J., Lukic, Y.D.: Human Cognitive Reliability Model for PRA Analysis, NUS-4531. NUS Corporation (1984)
7. Jung, W., Kang, D.I., Kim, J.: Development of a Standard Method for HRA of Nuclear Power Plants - Level I PSA Full Power Internal HRA, KAERI/TR-2961/2005. Korea Atomic Energy Research Institute (KAERI), Daejeon, Republic of Korea (2005)
8. Jung, W., Park, J., Kim, S., Choi, S.Y., Kim, Y., Yang, J.E.: Handbook for Analysis of Human Error Probability Based on Simulator Data of Nuclear Power Plants, KAERI/TR-6649/2016. Korea Atomic Energy Research Institute (KAERI), Daejeon, Republic of Korea (2017)

Human Performance in the Simulated Multiple Asset Routing Testbed (SMART): An Individual Differences Approach

Noelle Brown(✉), Joseph Coyne, Ciara Sibley, and Cyrus Foroughi

US Naval Research Laboratory, 4555 Overlook Ave SW, Washington, DC, USA
{Noelle.Brown, Joseph.Coyne, Ciara.Sibley, Cyrus.Foroughi}@nrl.navy.mil

Abstract. We examined performance for a university population and military population in the Simulated Multiple Asset Routing Testbed (SMART). SMART is a testbed designed to be similar to the types of tasks future unmanned vehicle operators will perform. Specifically, participants were required to optimize route selections for unmanned aerial vehicles. Their goal was to obtain a maximum number of points, given the likelihood of finding their targets. Participants showed superior performance when provided with detailed icons that contained information relevant to the route selections, compared to a condition where this information was provided in a table format instead. Although the detailed icons improved performance, we found that working memory capacity (WMC) and numeracy were predictive of accuracy in SMART. These results suggest that individual differences in WMC and the ability to understand and manipulate numbers may play an important role in tasks that require one to weight and optimize multiple outcomes. We discuss the implications of these findings for tool design and job selection and training.

Keywords: Human performance · Individual differences · Human-systems integration

1 Introduction

The Department of Defense (DoD) continues to invest in artificial intelligence and automation technologies that are rapidly changing how missions are conducted. Even newer technologies such as unmanned aerial vehicles (UAVs) are quickly changing. DoD is pushing to change how UAVs are operated: from direct user control to the supervision of multiple autonomous systems. This push will mean that UAV operators will no longer directly control a single aspect of a vehicle (e.g., payload operator), but rather they will be tasked with overseeing multiple vehicles and deciding how to best allocate various assets to meet their mission objectives. This may change the types of skills these operators need to perform in fundamental ways. For example, psychomotor ability, used as selection criteria in manned aviation, may no longer have any predictive validity.

It is likely that future unmanned system operators will perform more complex decision making tasks, such as how to best allocate multiple unmanned assets.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 95–105, 2020.
https://doi.org/10.1007/978-3-030-20037-4_9

However, the majority of the basic research in decision making involves a choice between two alternatives. Similarly, many existing theories of decision making are based on paradigms limited to simple gambles and other contrived tasks [1]. Thus, many existing theories fail to accurately account for decisions made in more complex settings. In contrast, decisions involving route planning for multiple vehicles and multiple objectives can involve choices across a large number of alternatives. Any increase in the number of vehicles and objectives increases the number of alternatives exponentially. Thus, it is important to understand the demands of these tasks and their impact on human performance.

Coyne, Foroughi, Brown and Sibley [2] began investigating how individuals develop UAV mission plans within their Simulated Multiple Asset Routing Testbed (SMART). In SMART, participants are trained and tested on how well they develop route plans which involve assigning UAVs to search for targets of varying values and probabilities of being found. The authors hypothesized that developing a plan which optimized over multiple factors would tax working memory capacity (WMC), due to the number of UAVs (3) and targets (6) involved. Research has shown that individuals with higher WMC are more likely to engage in rational decision making and are less likely to be susceptible to biases that influence their decisions [3]. However, Coyne et al. did not find any correlation between WMC, as measured by the Operation span task, and performance in SMART [2]. Additionally, the study included multiple trials in which participants were provided with a risky, sub-optimal route suggestion. Accepting these suggestions should have decreased performance, however when participants were provided with these initial route suggestions their performance on the task improved. The authors suggested that a lack of variability between route alternatives and a failure to comprehend the task may have influenced the results.

The present study seeks to further examine performance within SMART by addressing several of the limitations of the first study and expanding the types of measures used to predict performance within SMART. With these changes, the authors maintain that WMC should be correlated with performance in SMART. Additionally, a measure of numeracy was included because planning optimal routes involved assessing probabilities and expected values. Furthermore, research has shown that highly numerate individuals are less susceptible to framing effects than those lower in ability [4]. As such, the authors also predicted that numeracy would be positively correlated with performance within SMART. SMART was also adapted to include two types of icons to determine if there was a benefit to route planning by presenting information graphically as opposed to a tabular format. Similarly, the scenarios used within SMART were designed to have greater variability between alternatives to broaden the variety of route plans. Last, this study expanded the population beyond undergraduate students to include Navy and Marine Corps Officers currently attending flight school that may closely resemble that of real UAV operators. These changes to the SMART paradigm were expected to increase the variability in performance and provide a better test our hypotheses.

2 Method

This research complied with the American Psychological Association Code of Ethics and was approved by both the Navy and University's Institutional Review Boards.

2.1 Participants

Seventy-eight students from George Mason University participated in exchange for partial course credit. One-hundred eighteen student Naval aviators enrolled in flight school at the Naval Air Station Pensacola, FL volunteered to participate in this study.

2.2 Design and Procedure

After signing an informed consent form, participants completed a series of tasks and surveys in the following order: The shortened Automated Operation Span (Ospan) [5, 6], SMART [2], demographic survey, Risk Propensity Scale [7], Barret Impulsiveness Scale [8] and Numeracy Scale [9]. Each task is discussed in more detail below.

Shortened Automated Operation Span (Ospan). This is a shortened version of the original Automated Operation Span task designed to assess WMC [4]. In this task, participants were required to solve a two-part math equation (e.g., $(2 \times 3) + 1$) and then were presented with a solution which they judged as true or false. After each judgment, participants were presented a letter (e.g., K) to be recalled at the end of each set. The processing and memory tasks alternated 4–6 times depending on the set size (i.e., alternated five times for set size five). At the end of each set participants recalled the to-be-remembered letters in the order in which they were presented. Each set size was randomly presented twice and scores ranged from 0 to 30.

SMART. SMART [2] was designed as a multi-route optimization problem where participants weighed risks and rewards across multiple targets presented at different geographic locations within a map. The SMART interface consisted of two main regions (Fig. 1). The top portion of the screen was a map that showed the UAV and target locations. Once participants assigned targets to a UAV, the map showed a route from the UAV to the target, and any following target(s) if assigned.

There were two versions of the map icons presented in separate blocks and counterbalanced across participants. One version had simple icons, which only included the target name in the map (Fig. 2A). Any other information about the target was included in the information table. The other version had detailed icons which included the point values, a circular graphic that represented the targets' search areas (larger circles represented larger search areas), a bar that indicated the probability of finding the targets, target deadlines, and the target names (Fig. 2B). This information was duplicated in the table, which was identical for both icon conditions.

An information table was presented in the bottom of the screen along with counters that tracked the current trial number and running score. The column labeled "Objectives", provided target names, point values, deadline (how long it would take to search for the targets), and the distances of the targets' search areas. The remaining columns included information specific to each UAV's mission. This portion of the table was

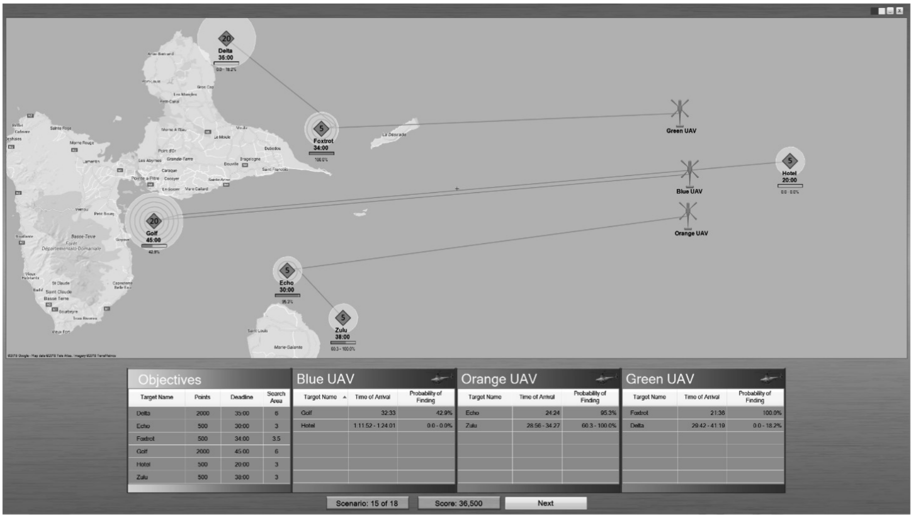
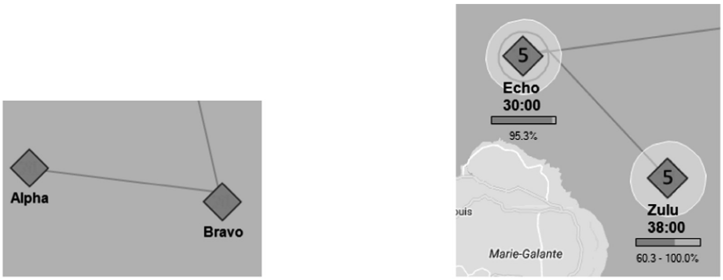


Fig. 1. SMART display after the user assigned targets to each UAV. The information table shows the information provided about each target (O*bjectives*) and each UAV's mission information (*Target Name, Time of Arrival, Probability of Finding*). The map shows the route plans for each UAV to each target. The above map shows detailed icons which contained additional information unique to each UAV' mission plan (*deadline, search area shown as white circles around the target, target point value, and probability of finding the target represented both numerically and graphically*).



A) B)

Fig. 2. (A) Simple icons or (B) detailed icons. Simple icons only showed the target name and route plan in the map. They lack the visual information about each target's point value, deadline, search area, and the probability of finding the target. For example, in the detailed icon (B) Echo's point value is 500, its deadline is 30:00, the search area is shown as a white circle under the target, and there is a 95.3% chance of finding the target before its deadline.

blank until participants assigned targets to the UAV. At that point, the table was populated with the target name, time of arrival and the estimated probability of finding the target.

At the outset of SMART, participants completed an automated self-paced training. Following the training, participants completed 18 self-paced experiment trials. On each trial, participants began by locating the UAV and target locations in the map. They were to use this information to assign the targets to each UAV. There were several ways to do this, but the most efficient was to attempt to assign two targets to each UAV, starting with targets that were closet to the UAVs. In each trial there were two high value targets (2000 points) and four low value targets (500 points). The likelihood of a given UAV finding a target varied based upon the distance of the UAV to the target, the size of the search area, and the amount of the search that could be completed prior to the deadline. The searchable area varied throughout the task such that the targets closest to each UAV varied from (0%–100%). The order in which participants searched for targets impacted the likelihood of finding them. The probability of finding a second order target was presented as a range based upon both the minimal and maximum length search of the previous target and the time to travel from the first to second target. For each UAV, participants were often forced to choose between a single guaranteed low value target or a chance to find two low value targets with no guarantee for either. As an example, participants may have a 100% chance of finding a 500 point target (e.g., Alpha) if searched for first and 0% chance for the second 500 point target (e.g., Bravo). Alternatively, if the order was reversed and participants search first for Bravo and second for Alpha, they may have a 66% chance of finding Alpha and a 50% chance of finding Bravo.

Participants were able to assign targets to UAVs and to switch the order in which they were searched as many times as they wanted. Once participants were satisfied with their route selections they clicked “Next” to end the trial. Participants were told that they were competing against a risk averse computer opponent. After each trial they received feedback on their trial score and overall score as well as the computer opponent’s scores.

Demographic Survey. The demographic survey was conducted on the computer and was a self-report survey.

Barratt Impulsiveness Scale (BIS). Responses to the BIS were collected on the computer. The BIS is a 30 item questionnaire designed to assess personality and behavioral tendencies toward impulsiveness and non-impulsiveness [7].

Risk Propensity Scale. Risk Propensity Scale is a 7-item self-report instrument designed to measure general risk-taking tendencies [8].

Numeracy Scale. Administered on the computer and adapted from Lipkus, Samsa and Rimer [9]. This is a self-report instrument designed to measure how challenging people find basic probabilities and mathematical concepts. Eleven items were presented and ordered by difficulty beginning with the easiest problem.

3 Results

Results from the Barret Impulsivity Scale and Risk Propensity Scale were expected to relate to people' willingness to risk searching for the high value targets (2000 points) first, even in cases where it was not deemed worth it to do so. However, these results are beyond the scope of this paper and are not discussed further.

3.1 Ospan

One pilot did not complete Ospan due to a computer error. We report partial Ospan scores [5] and did not include participants with accuracy at less than 70% on the math. The scores from the pilots (M = 24.0, SD = 5.2) were significantly higher than the university sample (M = 19.4, SD = 6.7), $F(1, 191) = 27.47$, $p < .001$, $\eta_p^2 = .13$.

3.2 Numeracy

Six university participants and one pilot did not complete the numeracy scale due to a computer error. The pilots showed a greater ability to manipulate and understand numbers and probabilities (M = 9.3, SD = 1.7), compared to the university students (M = 7.9, SD = 2.0), $F(1, 188) = 29.33$, $p < .001$, $\eta_p^2 = .14$.

3.3 Smart

SMART route plans were scored by expected value (EV) which was calculated by multiplying the probability of finding each target P(χ) by the target's point value (χ) and summing these values across all targets (see Eq. 1). For second order targets which showed a range of search probabilities (e.g., 20–40%) the average of the min and max was used (e.g., 30%).

$$E[X] = \sum [x_i \cdot P(x_i)]. \tag{1}$$

For each trial, we also calculated the maximum EV (max EV). This was calculated as the maximum number of points that could be obtained with the greatest probability for each trial. Max EV was determined a priori. Route plans in SMART were calculated as a proportion correct score by dividing the participant's EV for that trial by the max EV for the trial and ranged from .0–1.0. If participants optimized their routes, their EV was equal to the max EV resulting in a proportion correct score of 1.0. Poor route plans resulted in lower proportion correct scores.

Table 1 shows an example of the max EV for the optimal route plan and the EV if the targets were assigned to the same UAV, but searched in the reverse order. The effect of search order on the probability of finding a target is also demonstrated in Table 1. It was possible for participants to make optimal route selections for all, some or none of the three UAVs. Thus, EV varied by the participant's final route selection.

Table 1. Example of the optimal route plan and resulting max EV (*top row*) versus an alternative plan (*bottom row*) for a single trial.

Target search order	Blue UAV		Orange UAV		Green UAV		EV
	Point value	*Probability*	*Point value*	*Probability*	*Point value*	*Probability*	
Target A then B	2000	43%	500	100%	500	42%	2150
	500	48%	2000	0%	500	68%	
Target B then A	500	100%	2000	19%	500	100%	1590
	2000	9%	500	0%	500	6%	

Route plan scores and response latencies (sec) are presented in Table 2. We ran a mixed factorial ANOVA with Icon Type (Simple, Detailed) as the within-Subject's variable and Group (military, university) as the between-Subject's variable to determine whether providing detailed information in the icons was beneficial to route planning. However, the only significant difference was between Group, $F(1, 183) = 26.50$, $p < .001$, $\eta_p^2 = .13$. There was no main effect of Icon Type on route plan selections $F(1, 183) = .62$, $p = .432$, $\eta_p^2 = .00$, nor was there a significant interaction, $F(1, 183) = .92$, $p = .338$, $\eta_p^2 = .01$. The results suggest that the only effect on route planning ability was group.

Because overall accuracy was so high ($M = .89$, $SD = .06$), we also ran a mixed factorial ANOVA on median response times for Simple and Detailed icons as a more sensitive measure of performance. The results showed an effect of Icon Type, $F(1, 183) = 15.73$, $p < .001$, $\eta_p^2 = .08$, no effect of Group $F(1, 183) = .35$, $p = .558$, $\eta_p^2 = .00$, and the interaction was not significant, $F(1, 183) = .02$, $p = .894$. $\eta_p^2 = .00$. The results showed that Detailed Icons improved the ability to make route selections by reducing the time to make decisions compared to Simple Icons.

Table 2. Proportion correct scores for route plans and median response latencies (sec) in SMART. Standard deviations are shown in parentheses.

Group	Route plan scores		Response latencies	
	Simple	*Detailed*	*Simple*	*Detailed*
Pilots	.90 (.07)	.91 (.06)	76.9 (30.0)	68.5 (22.6)
University Students	.86 (.08)	.86 (.07)	75.1 (34.9)	66.1 (22.9)

Together the results suggest the pilot sample was superior in all of the tasks. This finding was not surprising because the pilots represent a sample pre-screened and selected on a set of cognitive and behavioral performance measures (Aviation Selection Test Battery, ASTB) [10]. Thus, their sample provided a very restricted range of performance that tends to be negatively skewed on most cognitive measures. Importantly, providing people with detailed icons did improve performance. The effect was found in faster route decisions for Detailed Icons than simple Icons but the effect did not extend to accuracy. The null findings for accuracy were likely due to the near ceiling performance in route selection accuracy.

Table 3. Bivariate correlations for performance in each task. Correlations significant at $\alpha = .01$ are marked with two asterisks (**).

Bivariate correlations		
Task	SMART	Ospan
Ospan	.27**	
Numeracy	.43**	.27**

3.4 Individual Differences

Although the military sample outperformed the university sample on all of the tasks, we were most interested in the contributions of Ospan and numeracy on the ability to optimize route selections in SMART across the groups. The results showed significant relationships between Ospan, numeracy and SMART (Table 3).

To further investigate these effects we ran a multiple regression on SMART route selection scores with Ospan score (only participants who maintained processing accuracy at 70% or greater) and numeracy entered simultaneously as predictors. The results are reported in Table 3 and Fig. 3. The multiple regression model significantly accounted for 21% of the variance in SMART route selection score. Ospan and numeracy contributed significantly to the model suggesting that WMC and an ability to manipulate numbers both play a role in optimizing route selections in SMART. The results further suggest that numeracy has a stronger effect on route optimization in SMART (Table 4).

Table 4. Multiple regression model on SMART Performance. Values significant at $\alpha = .05$ are marked with a single asterisk (*) and values significant at $\alpha = .001$ are marked with three (***). $R^2 = .21$, $F(2, 185) = 24.31$, $p < .001$.

Multiple regression model				
Variable	β	Semipartial *r*	*t*-value	R^2
Ospan	.14	.14	2.09*	.21***
Numeracy	.40	.38	5.86***	

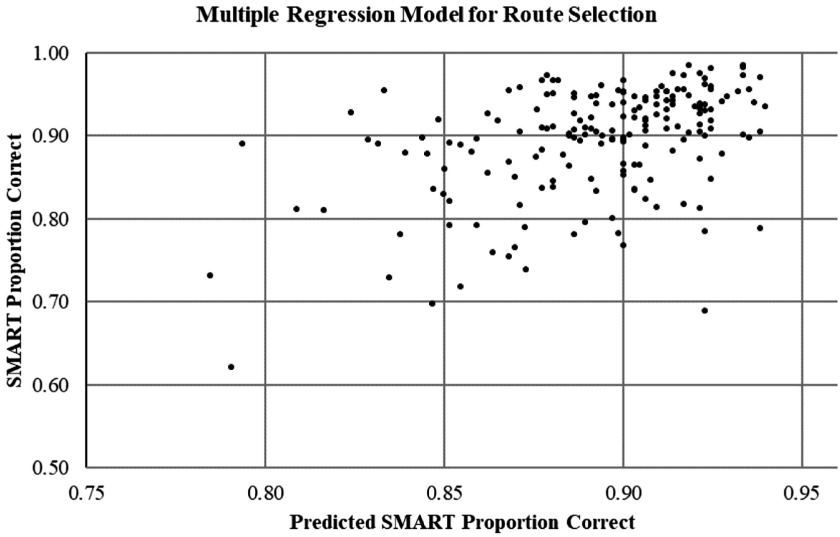


Fig. 3. Linear multiple regression model results for Predicted SMART Scores (*x-axis*) plotted against Actual SMART Scores (*y-axis*). $R^2 = .21$, $F(2, 185) = 24.31$, $p < .001$.

4 Discussion

We investigated the effect of type of icon and individual differences in the Simulated Multiple Asset Route Testbed (SMART) for a group of military pilots and university students. The results showed a clear distinction in cognitive abilities of the two samples. We also found a benefit of the detailed icons in processing speed and contributions of both WMC and numeracy to route planning ability.

It was not surprising that the military group consistently outperformed the university group. The pilots were screened and selected based on their performance on the ASTB. The ASTB is a battery of cognitive, behavioral and personality measures designed to select candidates for pilot and flight officer positions [10]. Unlike the university students, the military pilots represent a group of participants with a very restricted range of performance abilities that are typically negatively skewed. Notably, we expected this group to be representative of future UAV operators. The Air Force selects UAV operators from current pilot candidates. Thus, it is reasonable to expect that the Navy, Marine Corps and Coast Guard may also select future UAV operators from their set of pilot candidates as well.

We predicted that providing detailed information in the icons would benefit performance for both accuracy and response latency compared to simple icons; however, the effect of icon type was only found for response latency. We suspect that the effect of icons did not impact accuracy because performance was near ceiling. Alternatively, the effect of icon type on accuracy may not exist because the icons do not provide any information beyond that provided in the information table. It is noteworthy that the detailed icons did speed up processing without influencing the quality of their route plans. The military pilots made decisions on average about 8.4 s faster with detailed icons and the university students made decisions bout 9 s faster. That is equivalent to an 11% reduction in decision time. This reduction may be critical for time sensitive tasks. Further, these results suggest that decision criteria provided in a graphic format may improve the design of decision aids.

Our investigation on the role of individual differences in route planning showed contributions of both WMC and numeracy. All three measures, Ospan scores, numeracy and SMART scores were correlated with each other. Additionally, numeracy and Ospan scores were predictive of route planning accuracy. These findings are significant because they suggest that numeracy and WMC are abilities likely to influence UAV operator performance. As such, UAV operator selection may be improved if future operators were additionally screened on these aptitudes. In comparison, the findings also suggest that reducing the working memory load and reliance on numeracy may also benefit performance.

Overall, the results suggest that UAV operators who are similar in ability to current military pilots are likely to perform well on UAV tasks that involve route optimization. Importantly, the design of the interface can improve their decisions further by reducing the amount of time required to make their decisions. This can have a significant impact for the military where decisions are often time critical. The findings also showed that numeracy and WMC are critical to route planning. We feel that this has fundamental implications for both UAV operator selection and task design. First, selecting UAV operators based on these criteria may improve their success. Second, UAV route planning tasks and displays should be designed to minimize the influence of numeracy and WMC. These can be achieved through human-systems integration design techniques and leveraging artificial intelligence to help reduce the manual processing of the operator.

One limitation of the current study is that it is unclear whether the benefit of the detailed icons was due to a better understanding of graphical information than tabular or if it reduced the number of times people looked back and forth between the table and map. Understanding this difference can further design improvements. For instance, if graphical information is more readily understood, including the same information in tabular format may be a waste of space. Future research should also continue to explore the contributions of numeracy and working memory to determine whether their contributions to route planning can be mitigated with improvements to the displays.

Acknowledgments. This work was funded by the Office of Naval Research. Specifically, the authors would like to thank Dr. Jeffrey Morrison, Program Officer for the Command Decision Making (CDM) program for his continued support.

References

1. Stewart, N.: Decision by sampling: the role of the decision environment in risky choice. Q. J. Exp. Psychol. **62**, 1041–1062 (2009)
2. Coyne, J.T, Foroughi, C.K., Brown, N.L., Sibley, C.M.: Evaluating decision making in a multi-objective route planning task. In: Proceedings of the Human Factors and Ergonomic Society Annual Meeting, vol. 62, pp. 217—221 (2018)

3. Fletcher, J.M., Marks, A.D.G., Hine, D.W.: Working memory capacity and cognitive styles in decision making. Pers. Individ. Differ. **50**, 1136–1141 (2011)
4. Peters, E., Västfjäll, D., Slovic, P., Mertz, C.K., Mazzocco, K., Dickert, S.: Numeracy and decision making. Psychol. Sci. **17**, 407–413 (2006)
5. Oswald, F.L., McAbee, S.T., Redick, T.S., Hambrick, D.Z.: The development of a short domain-general measure of working memory capacity. Behav. Res. **47**, 1343–1355 (2015)
6. Foster, J.L., Shipstead, Z., Harrison, T.L., Hicks, K.L., Redick, T.S., Engle, R.W.: Shortened complex span tasks can reliably measure working memory capacity. Mem. Cognit. **43**, 226–236 (2015)
7. Patton, J.H., Stanford, M.S., Barrett, E.S.: Factor structure of the Barratt impulsiveness scale. J. Clin. Psychol. **51**, 768–774 (1995)
8. Meertens, R.M., Lion, R.: Measuring an individual's tendency to take risks: the risk propensity scale. J. Appl. Soc. Psychol. **38**, 1506–1520 (2008)
9. Lipkus, I.M., Samsa, G., Rimer, B.K.: Performance on a numeracy scale among highly educated samples. Med. Decis. Making **21**, 37–44 (2001)
10. Aviation Selection Test Battery (ASTB). https://www.med.navy.mil/sites/nmotc/nami/pages/astboverview.aspx

Shared Spatial Situation Awareness as a Team Performance Indicator in Collaborative Spatial Orientation Task

Baptiste Prebot[1(✉)], Jean-Marc Salotti[1,2], Coralie Vennin[1], and Bernard Claverie[1]

[1] ENSC, IMS Laboratory, Bordeaux INP, Bordeaux University, CNRS, 109 Avenue Roul, 33400 Talence, France
{baptiste.prebot, jean-marc.salotti, coralie.vennin, bernard.claverie}@ensc.fr
[2] Auctus, Inria, Talence, France

Abstract. The present study investigates the link between time taken by a team to perform a spatial orientation task and the evaluation of spatial shared situational awareness (SSA). Paired in teams, volunteers have to collaborate to send a vehicle to a specific location on a computer simulation as quickly as they can. The roles and information they have to reach that goal are different. Every 45 s participants are asked to mark on their map the location they believe the vehicle to be. Along with its real position, these marks are used to objectively evaluate spatial SSA. First results allow us to divide participants into three groups in accordance with Endsley's distinction of Shared SA evaluation. Interestingly, fastest teams were not the ones with the most accurate and shared spatial representation of the situation. Potential use of such indicators in team training is outlined.

Keywords: Collaboration · Shared situation awareness · Spatial awareness · Team performance

1 Introduction

We address the problem of team shared situation awareness assessment in complex organizations with complex decision making processes. Applications can be found for examples in the context of command and control in the military domain or in the context of scientific exploration on far planetary surfaces. The assessment of shared situation awareness (SSA) is usually carried out a posteriori (e.g., after an accident) and in general by means of a qualitative analysis. An important difficulty is to be able to make a quantitative measurement of variables that are correlated to the exactness of the situation awareness and to the similarity of the shared representation. It is proposed here to analyze the spatial component of situation awareness, which can be directly inferred from the marks that participants can put on a map to indicate a belief of position. Depending on the quality of the collaboration, the marks can be more or less exact and similar for all participants of the same team. An experiment has been carried out in a virtual environment to test the SSA and to examine the links between the

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 106–115, 2020.
https://doi.org/10.1007/978-3-030-20037-4_10

spatial SSA and task performance. The paper is organized as follows. Section 2, the method is presented and the experiment is described. Section 3, the main results are given and discussed. In conclusion, some perspectives are proposed.

2 Method

2.1 Main Concepts

Collaboration is seen in this study as the integrative form of cooperation, as defined by Schmidt [1], where agents are engaged in subtasks of the same main task, while having different but complementary skills. This idea of complementarity and role division under a common goal is also found in Salas, Dickinson, Converse and Tannenbaum [2] definition of a team: "*a distinguishable set of two or more people who interact dynamically, interdependently, and adaptively toward a common and valued goal/object/mission, who have each been assigned specific roles or functions to perform, and who have a limited life span of membership" (pp. 126–127).*

From search and rescue missions to space exploration or military operations, the experimental conditions of two separated operators, one having direct impact on the environment and the other having more information, is a common occurrence. In such collaborative environments, the shared ability to guide one another and localize objects in unknown terrains, in real time and possibly in degraded situation is critical for the overall team performance. To do so efficiently, teammates must share the most common and accurate representation of the situation as possible.

Already a widely studied subject in the field of human factors, situation awareness (SA) has gained even more interest in the rise of team cognition and performance evaluation. Since the 1980s numerous models have been proposed [3, 4]. Endsley formally defines it as "*... the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in near future" (p. 36)* [5].

Extending the concept to teams and collaborative work Shared SA can be defined as *"the degree to which team members possess the same SA on shared SA requirements" (p. 48)* [6]. As Shared SA can be seen as a matter of both knowledge and coordination, two levels of measurement are distinguished [7]. The degree of *accuracy* of an individual's SA and the *similarity* of two individual's SA are both needed in order to assess the Shared SA between two team members. The evaluation of SA accuracy is what most evaluation techniques are focused on. For each participant of the observed experiment, the understanding of the situation is compared to the true state of the environment at the time of evaluation, leading to the assessment of SA as being either right or wrong. This binary evaluation requires a ground truth to which the SA has to be compared. The evaluation of the similarity of SAs is based on the direct comparison of situation awareness elements, which are relevant to both of them. The assessed SA is either shared or not on these specific elements. Combining evaluations of each participant lead to SA being assessed as either right or wrong and shared or not, defining three possible SSA states: different, both correct, both incorrect [6].

When guiding someone remotely, the current position of the person being guided in the field can be defined as a necessary shared knowledge element [8], a must-shared information for the correct accomplishment of the collaborative task.

In this study the focus is put on a spatial orientation task [9]. Thus, working on the sharing of spatial representation of the situation, we consider here spatial situation awareness [10] as a part of SA, a restriction of the one's global understanding of the situation to only some elements relative to the position in space of the teammate or of oneself.

The overall goal of this study is to investigate how spatial Shared SA similarity and accuracy relate to the team performance (here the time spent to accomplish the task) and in what extent they can themself be used as objective quantitative indicators of team's performance.

The research questions can be summarized as follow:

- Can Spatial SSA Accuracy and Similarity be used as quantitative team performance metrics?
- Can Spatial SSA Accuracy and Similarity be used to identify teams profile in regard of their performance to the collaborative task?

2.2 Description of the Experiment

62 participants (38 female and 24 male) ranging from 18 to 43 years old (M = 21.6) took part in this study. They were recruited around the local campus to perform a collaborative orientation task. Paired up in teams, volunteers have to collaborate to send a vehicle (a rover) to a specific location on a computer simulation of a Martian environment as quickly as they can. To do so, they are assigned two specific roles. One person, the astronaut (Astro), drives the rover in the virtual environment and has a map of that environment, while the second person, the captain (Capcom) in charge of guiding him, only has access to the map, with the target (a white rock) location marked. They had no previous knowledge of the environment and were separated so as to be able to communicate only orally.

2.3 Material

A A4-size paper map representing the simulated environment was given to each participant (Fig. 1). No scale was indicated nor any grid or coordinate system. Only the starting point of the rover and its orientation were stipulated on each map. Maps were given oriented in the same way to each participant. The simulation is Unity based, developed internally, and was displayed on a 24-inch monitor.

2.4 Design and Procedure

Before the experiment, each teammate was assigned the role either of Astro or Capcom. Then they were brought to their work station and given instructions depending on their role. The map was handed to them and the starting position of the rover and initial orientation was indicated on it. The Astro task was to navigate the rover (first person

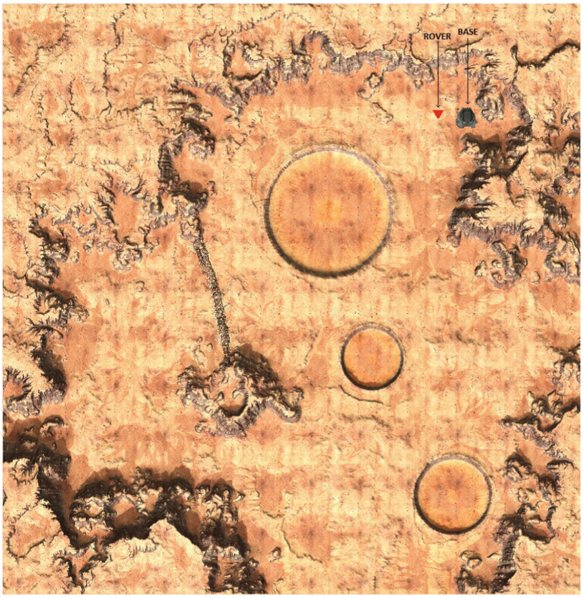


Fig. 1. Example of the map given to the Astro

view of the environment, as he was the driver), following directions given by CapCom, in order to find a specific rock. They were told that the goal was to find the rock as quickly as possible. During the whole experiment they were allowed to communicate only orally (Fig. 2). CapCom was instructed not to directly give the Astro the rock position (not allowed to say "the rock is on the top left corner of the map") but to orient him in real-time. Every 45 s the simulation was paused and the subjects were asked to mark down on their map where they estimate the rover to be (Position Evaluation Point). Teammates were not allowed to communicate during this phase of the experiment. If after 15 min, participants did not find the rock, it was considered a failure and the simulation was stopped. After the experiment, they answered a questionnaire composed of background questions and subjective ratings and feedbacks on the evaluation of their collaboration.

2.5 Measures

Three types of quantitative team performance metrics have been recorded. Two spatial metrics are extrapolated from the marked positions on the map. Spatial SSA accuracy is objectively measured by comparing the position of the marks on each map with the exact position that is registered by the simulation. By measuring the distance between the positions marked by teammates for a same position evaluation point we evaluated the similarity of spatial SSA. The distance is expressed in Unity measurement metric (1 unit = 0.083% of the map). Finally, the time taken (in seconds) to complete the task has been recorded and serves as the Time performance metric.

Fig. 2. Experimental set-up

Using Likert scales, participants' subjective evaluation of the team performance, their own and their teammate performance, their quality of sharing of SA and the perceived difficulty of the task have been collected through a post-experiment questionnaire.

3 Results

Among the 31 tested pairs, 3 crashed the vehicle. These 3 experiments have not been considered for situation awareness assessment. Among the remaining 28 pairs, 8 did not find the rock under 15 min. As they failed to complete the task, they have been grouped in a separate category. Though it is interesting to look at the cause of failure, we propose to focus in this paper only on the 20 successful teams. The mean time to complete the task is 460.5 s (min = 274 s; max = 648 s, SD = 102.56). SSA similarity and SSA accuracy have been calculated. The results are presented and analyzed in the next paragraphs.

3.1 SSA Similarity

The 42 participants generated 386 Position Evaluation Points. Coordinates of each point have been logged. Each Astro point was paired with the corresponding Capcom point. For each pair, the distance between the two points is calculated. The shorter the distance between the two points, the more similar, and thus shared, the representation of the spatial position is between the teammates. Then for each team, distances have been averaged across the number of points that have been marked on their maps.

3.2 SSA Accuracy

As stated earlier, the real position of the rover has been registered by the simulation all along the experiment. Each Position Evaluation Point coordinates have been compared to the real position of the rover recorded by the simulation at that time.

The distance between the two points is calculated. Averaging the distances by subjects shows a difference between Astro and CapCom SSA accuracy. Astros mean SA accuracy is lower (Mean Astro = 97.67, SD = 107.02) than CapComs mean accuracy (Mean CapCom = 141.06, SD = 137.03), meaning that Astros have generally a more accurate representation of their spatial location than CapComs. This is coherent since Astro is the one whose position is evaluated and the one navigating the simulation, thus more inclined to have a more accurate evaluation of his own position in the environment. For a team level evaluation, the accuracy distances for each point and each teammate have been averaged. We obtain a single average accuracy value for each team. As for SSA Similarity, the lower the mean value of the distance between those points is, the more accurate the SA is. Perfect accuracy would be 0. Note that the more accurate the team is, the more similar it is supposed to be (Table 1).

Table 1. Statistics of SSA similarity, accuracy and time measurements.

	Similarity	Accuracy	Time (s)
Mean value	94,87	103,62	457
Standard deviation	71,40	65,28	104.14
Min	4,54	46,16	274
Max	339,52	318,87	648

3.3 Group Analysis

As seen in Fig. 3, two teams can clearly be identified from their Shared SA Accuracy and Similarity performance. Team N°16 (G16, top right corner of the plot) possess mean Similarity and Accuracy scores both more than two times over the average values (Similarity = 201.08, Accuracy = 238.17). Both teammates are being incorrect in different ways, thus having an inaccurate and really different representation of the situation. Team N°10 (G10, bottom right) is the less accurate one, with a mean Accuracy of 318.87, more than three times the average value, while interestingly the mean Similarity is under the average value (81.39). It reflects teammates sharing a common but erroneous representation of the situation.

Surprisingly, Time performance of these two teams is average. A further analysis of their collaboration would be required.

These two specific cases cover two of the three possible Shared SA states as described by Endsley [6], with inaccurate and either similar or not representation of the situation.

The following analysis is focusing on the remaining 18 teams as having an overall accurate and similar SSA but showing a great variability in Time performance.

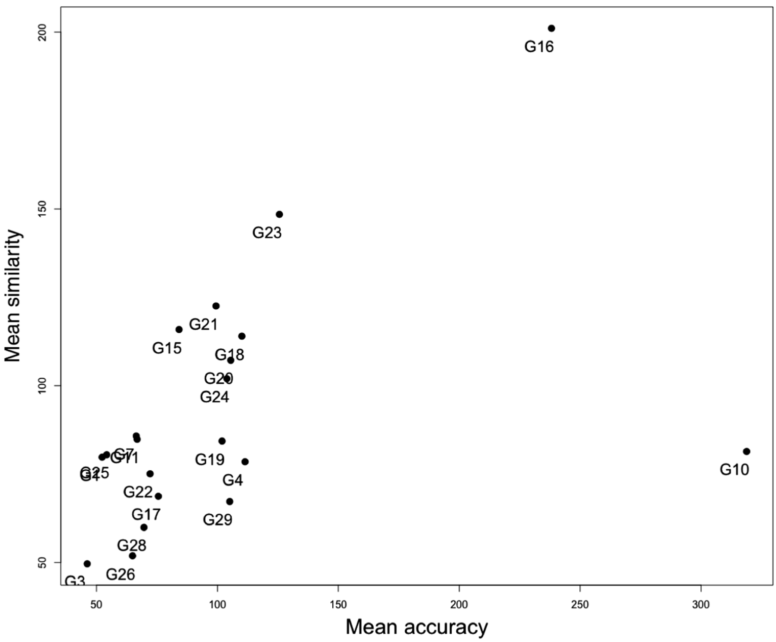


Fig. 3. Teams plotted by similarity and accuracy SSA performance

3.4 K-means Analysis

In order to find possible clusters of teams based on their SSA accuracy, SSA similarity and Time performance, a K-means clustering method was used, implemented in R package of Morisette and Charier [11], based on Hartigan and Wong algorithm [12]. The algorithm classifies teams according to the similarity criterion and determines for each class a referent vector. For each class, the sum of the square distances between each team vector and the referent vector represents the inter-individual variability. Then, the sum of the square distances between each pair of referent vectors describes the variability between each class. The number of teams assigned to each class is also defined.

From the K-means analysis, 2 groups can be characterized, both composed of 9 teams. Descriptive statistics of each identified cluster are presented in Table 2.

Results show that the first cluster identified has a better overall Shared SA performance. Teams in Cluster 1 have position evaluations that are closer to the reality. The average calculated for the accuracy variable is equal to 70.6 (SD = 9.9) while it is equal to 104.9 (SD = 11) for the second cluster. The standard deviations for the Accuracy and Similarity variables indicate that the evaluation performance is more consistent in cluster 1 than in cluster 2. The results for the time variable show that the average duration of the task is better in the first cluster. The teams took on average 381 s (SD = 77.2) to find the rock whereas the second cluster took on average 545 s (SD = 57). Standard deviations also indicate that cluster 1 has a bigger inter-teams variability. Indeed, the minimum duration is 274 s and the maximum is 462 s. In the second cluster the minimum duration is 476 s when the maximum duration is 648 s. As can be seen from the boxplots of Fig. 4, the confidence intervals measured for the Time variable indicate that the temporal performances are significantly different between the clusters 1 and 2. The same can be said for the SA accuracy.

Table 2. Statistical features of the two clusters of teams obtained with k-means.

	Cluster 1	Cluster 2
Mean similarity	63.1 (13.8)	105.2 (24,9)
Mean accuracy	70.6 (9.9)	104.4 (11)
Mean time	381.8 (77.2)	545.4 (57)
Sum of squares	49964.9	31936.5
min (mean accuracy)	46.1	84.1
max (mean accuracy)	75.5	125.6
min (mean similarity)	49.6	67.2
max (mean similarity)	85.7	148.4
min (mean time)	274	476
max (mean time)	462	648
N	9	9

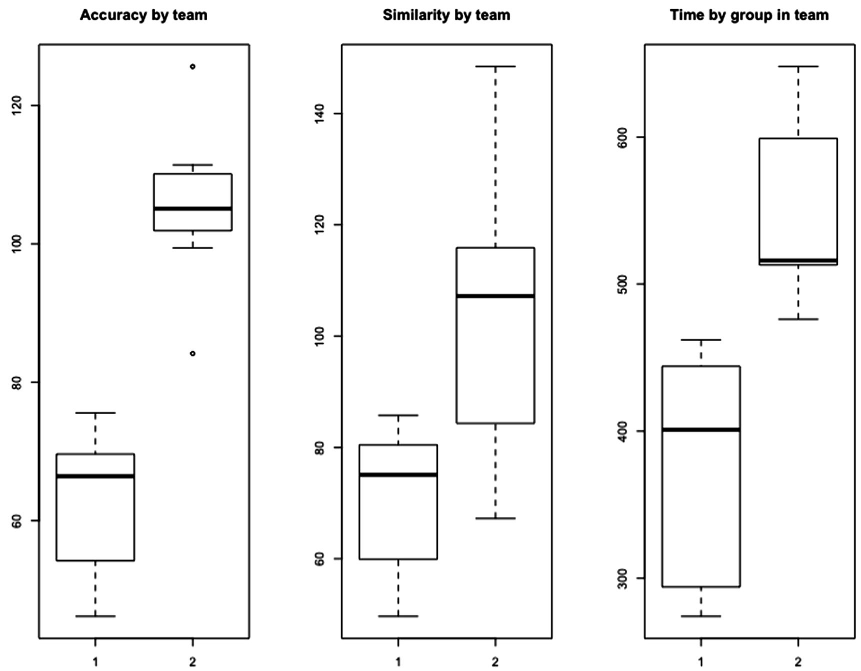


Fig. 4. Boxplot comparison of SSA accuracy, similarity and time performances by team.

Based on this observation, a relationship can be assumed between mean SA accuracy and Time performance. Pearson's correlation test allowed us to verify this relationship. The result confirm a positive linear relation between Shared SA Accuracy and Time performance ($r = 0.76$; $p < .001$). The smaller the mean SSA Accuracy distance is over the experiment, the smaller, thus better the time performance will be.

4 Discussion

Several results have be presented regarding the link between spatial Shared Situation awareness evaluation and time performance during a collaborative orientation task. First, the two specific cases of "inaccurate and different", and "inaccurate but similar" Shared SA have been identified.

When focusing our analysis on the task completed teams in terms of Shared SA, two groups emerged with a correlation between SSA and time performance. Although the overall spatial SSA difference alone is not enough to explain the inter-teams differences in Time performance, results reveal that spatial SSA Accuracy is positively correlated to it. This correlation implies that spatial SSA Accuracy may be used as a quantitative linear indicator to anticipate team's time performance.

However, Spatial SSA Accuracy might not be the only variable to take into account in order to explain Time performance. Previous work [13] already examined the use of common experience, workload similarity or communication distances, as predictors of shared situation awareness in teams. Complementary, integrating in our model answers from the post-experiment questionnaire could help identify if other factors, like individual perceived task difficulty and perceived team or teammate's performance, can be used to predict their performance. This will probably be the center of future work.

In sum, using distances (spatiality) as an evaluation metric of Shared SA, allowed us to quantify more precisely the situation awareness, giving levels to in the end qualify teams performance.

The findings presented here provide preliminary evidence that it is possible, in a collaborative spatial orientation task, to profile teams and anticipate, to some extent, their expected performance. Finally, in the domain of astronautics, this method can be used for training and testing astronauts' behavioral competency, especially situation awareness optimization and communication efficiency, which are in the list of competencies required by NASA [14]. Other applications could include defining an optimal performance matching the team collaboration profile or tailoring training simulations and exercises goals to each team.

Future work will also include analysis of the intra-teams temporal evolution of Spatial SSA metrics during the experiment.

Acknowledgments. Part of this study has been financed by the Direction Générale de l'Armement (DGA). We would like to particularly thank Laetitia Calice, Caroline Cavel, Mateo Mahaut and Adrien Leduque, ENSC students who have been highly involved in the experiment.

References

1. Schmidt, K.: Cooperative work: a conceptual framework. Distrib. Decis. Mak. Cogn. Model Coop. Work **21**, 75–110 (1991)
2. Salas, E., Dickinson, T.L., Converse, S.A., Tannenbaum, S.I.: Toward an understanding of team performance and training. Teams Train. Perform. **12**, 3–29 (1992)
3. Stanton, N.A., Salmon, P.M., Walker, G.H., Salas, E., Hancock, P.A.: State-of-science: situation awareness in individuals, teams and systems. Ergonomics **60**(4), 449–466 (2017)
4. Salmon, P.M., Stanton, N.A., Walker, G.H., Jenkins, D.P.: What really is going on? review of situation awareness models for individuals and teams. Theor. Issues Ergon. Sci. **9**(4), 297–323 (2008)
5. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors J. Hum. Factors Ergon. Soc. **37**(1), 32–64 (1995)
6. Endsley, M.R., Jones, W.M.: A model of inter and intra team situation awareness: implications for design, training and measurement. In: New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments Santa, Monica, pp. 46–67 (2001)
7. Saner, L.D., Bolstad, C.A., Gonzalez, C., Cuevas, H.M.: Measuring and predicting shared situation awareness in teams. J. Cogn. Eng. Decis. Mak. **3**(3), 280–308 (2009)
8. Wickens, C.D.: Situation awareness: review of Mica Endsley's 1995 articles on situation awareness theory and measurement. Hum. Factors **50**(3), 397–403 (2008)
9. Johansson, B.J.E., Hellgren, C., Oskarsson, P.-A., Svensson, J.: Supporting situation awareness on the move-the role of technology for spatial orientation in the field. In: Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management ISCRAM, Baden-Baden, Germany, May 2013, pp. 442–451 (2013)
10. Klippel, A., Hirtle, S., Davies, C.: You-are-here maps: creating spatial awareness through map-like representations. Spat. Cogn. Comput. **10**(2–3), 83–93 (2010)
11. Morissette, L., Chartier, S.: The k-means clustering technique: general considerations and implementation in mathematica. Tutor. Quant. Methods Psychol. **9**(1), 15–24 (2016)
12. Hartigan, J.A., Wong, M.A.: Algorithm AS 136: a k-means clustering algorithm. Appl. Stat. **28**(1), 100–108 (2006)
13. Bolstad, C.A., Foltz, P., Franzke, M., Cuevas, H.M., Rosenstein, M., Costello, A.M.: Predicting situation awareness from team communications, Pearson Knowledge Technologies, pp. 789–793 (2007)
14. Bessone, L., et al.: International space station human behavior and performance competency model volume II mission operations directorate, vol. 2, April 2008

Human Error Considerations in Design

A Risk Analysis Method for Human-Machine System Based on Risk Scenario and Interaction Interface Analysis

Xing Pan, Yuhang Jiang, Wenjin Zhang(✉), and Xianheng Zhao

School of Reliability and Systems Engineering,
Beihang University, Beijing, China
{panxing, zwjok, zhaoyou}@buaa.edu.cn,
forimba@outlook.com

Abstract. This paper proposes a risk analysis method for human-machine interaction system (HMIS). There are five correlated risk factors being analyzed in the HMIS: human conditions, human decision processes, human behaviors, machine and environment. With the considering of the system hazards in three types of risk scenario, we put forward an evaluation method on human-machine interaction risk for the HMIS, which provides a formalized table and contains the interaction interfaces risk evaluation procedure. In the method, we use a fuzzy method to quantify the human-machine interaction risk. Finally, a case study of the unmanned aerial vehicle (UAV) control system is conducted to illustrate the effectiveness of the method, and some advices for human-machine design are given.

Keywords: Interaction interfaces · Risk scenario · Human-machine system

1 Introduction

Human-Machine Interaction (HMI) has formed when the engineering system appears, which aimed to achieve communication between human and devices. As the continuous improvement of the automation in industry, the interaction between the human, machine and environment becomes more complicated. However, until the emergence of complex devices such as computers and traditional methods only for mechanical devices can be fully utilized in HMIS, the human-machine interface problem has become a hot topic of research. Statistics show that about 60–80% of recent industrial accidents are caused by errors or potential risks of interaction between people and equipment, especially in the nuclear industry and aviation.

Now, the problem of HMI becomes the hot topic of research. Miller evaluates the HMI process by the questionnaire based on the experiments [1]. Bemmel studied the HMI problem of the disease classification model, which used the doctor's decision process of disease classification as the research object, and proposed a systematic approach to achieve the analysis of HMI [2]. Bonney studied human factors in the field of architecture and engineering design, to find the impact of HMI process on manufacturing, using and maintenance costs [3]. Miyake conducts research on the impact of

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 119–130, 2020.
https://doi.org/10.1007/978-3-030-20037-4_11

common factors in HMI [4]. Magalhaes studied the impact of HMI in transport environment [5]. Lliu used the Simulink toolbox in Matlab to simulate complex human-machine-environment factors in a flying pilot system and give a quantitative analysis [6]. Jin proposes human-machine-environment risk-coupling concept for complex HMIS [7]. Taking into account the interaction between the factors, the overall risk analysis of the system can more reliably describe the system risk. HMI risk scenarios collect potential risks that may affect the HMI process and trigger system incidents. Therefore, it is important to consider these scenarios when conducting an overall risk analysis.

This paper takes the human-machine interface as the research object, conducts risk assessment, expands the research scope of traditional risk analysis methods, and provides innovation for HMI risk analysis. The paper is structured as follows, Sect. 1 introduces the theoretical basis of HMI. Section 2 studies human-machine interaction interfaces and risk scenarios. Section 3 gives a risk analysis approach in risk scenario. Section 4 presents a case study. Section 5 contains the discussion and conclusions of the paper.

2 HMI Interface and Risk Scenario Analysis

2.1 HMI Interface Analysis

The HMIS is composed of personnel, equipment and environment. As shown in Fig. 1, The interactions in the system are divided into three categories, human-machine interaction, human- environment interaction and machine- environment interaction [8]. This classification method is simple enough, but obviously does not completely clarify the relationship between the three. In order to further clarify this interaction relationship, the HMIS is reclassified according to the information flow in Fig. 2.

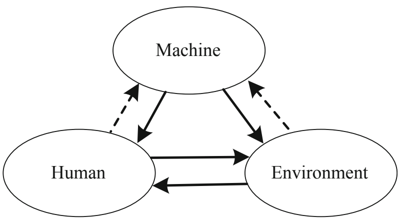


Fig. 1. Human-machine interaction system

In this method, human factors are divided into three categories, human condition, human behavior and decision processes, which can understand the interaction between human-machine-environment clearly. The information interface is the flow of information from the machine to the human condition. The operation interface is the flow of information from the human behavior to the machine. The environment-human interface is the flow of information from the environment to the human condition. The environment-machine interface is the impact of the environment on the machine. The

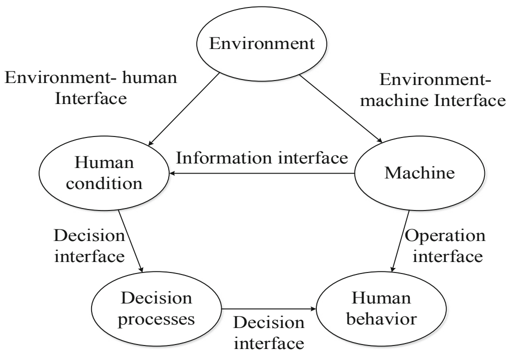


Fig. 2. Division of human-machine interaction interface

decision interface is the flow of information from the human decision-making process of the human condition to the human behavior.

According to the human-machine interaction interface, the risk factors in the system are composed of five types: human behavior, human decision, human condition, equipment factor and environmental factors. Due to the one-way flow characteristics, the order of the interaction interfaces can be determined. Through the flow of information between different interfaces, it is possible to sort out the risk factors in the human-computer interaction system and their corresponding interface relationships. When the risk factor corresponds to the risk interface, fill "√" in the form. As shown in Table 1, for example, there is a task "XXX". The equipment factors in system are monitor and LED, which belong to equipment-to-human information transfer; we fill "√" in the "Information interface". The environment factors are noise and vibration; we fill "√" in the "Environment- Human interface".

Table 1. HMIS interaction interface and risk factor analysis table. (Sample)

Task	XXX	Date		Written by		
Task detail						
System composition		Decision interface	Environment-human interface	Information interface	Operation interface	Environment-machine interface
Equipment factors	Monitor			√		
	LED			√		
Environment factors	Noise		√			
	Vibration		√			

2.2 Risk Scenario Analysis

A single risk factor may have little impact on the accident. However, through the interaction with other factors, a series of risk factors may cause accidents. The collection of these risk factors is called risk scenario. In HMIS, there are three types of accidents: casualties, equipment losses, and mission failures [9, 10]. Any event that causes three accidents is called initiating event, and any factor that may cause three types of accidents is called risk factor. Although the risk analysis method based on the initial event has a process of analyzing the system, when the risk analysis of complex HMIS is carried out, it is difficult to enumerate the unordered events. Therefore, the risk analysis method based on risk scenarios has higher practical value. In this paper, based on different types of risk factors and different interaction interfaces, risk scenarios are divided into three categories:

- The first type of risk scenario refers to the risk scenario formed during the interaction of risk factors. The interaction process involved is mainly the interaction between the risk factors of equipment, environment and human condition.
- The second type of risk scenario refers to the risk scenario formed by the human-machine interaction process, involving the human condition and the human behavior, without the environmental factors. This kind of risk scenario is mainly formed by the interaction of human condition factors and decision process factors.
- The third type of risk scenario refers to the risk scenario formed by the joint action of human-machine-environment, mainly the environmental factors and equipment systems and the impact of people's state on the factors affecting decision-making, leading to risk scenarios of decision-making mistakes.

After determining the risk scenarios and which HMIS interaction interface the risk scenarios belong to, according to the HMIS interaction interface and risk factor analysis table, a set of risk factors for the risk scenario can be obtained. This method is simpler and more flexible than the method based on enumeration analysis of the originating event.

3 Risk Scenario Analysis Based on Fuzzy Evaluation Method

After determining the risk factor set of the risk scenario, a risk scenario analysis is needed. The risk assessment is mainly divided into two parts, risk scenario probability level analysis and the risk consequence level assessment.

3.1 Risk Scenario Probability Level Analysis

The risk scenario of the HMIS is a collection of risk factors in the system, therefore, the probability of evaluating the risk scenario can be based on the probability of occurrence of the risk factor.

Probabilistic quantification of risk factors is an important process in risk analysis, mainly determined by experts giving the probability of occurrence of factors. According to the probability of occurrence, there are four levels, (Rare, Normal, Common, Frequent), corresponding to four segments (1.0–1.9, 2.0–2.9, 3.0–3.9, 4.0–4.9). The risk factor set of risk scenario A is $\{a_1, a_2 \ldots a_m\}$, experts need to evaluate these risk factors. The evaluation result of risk scenario A is $\frac{1}{m}\sum a_i$, and risk scenario probability level can be determined.

3.2 Risk Consequence Level Assessment

Different risk scenario will lead to different accidents, the risk consequence level can be judged by the level of accident hazard that is triggered.

System accidents are classified into five levels according to the consequences of accidents, (Good, Not bad, Normal, Not well, Terrible), the classification of the accident level is based on three system accidents, casualties, equipment losses, and mission failures, as shown in Table 2.

Table 2. Risk consequence level assessment table

Level	Evaluation criteria
Good	Individual non-critical parts fail, personnel discomfort, secondary task failure
Not bad	More non-critical parts fail, personnel can recover damage, multiple secondary tasks fail
Normal	Key component reparability failure, personnel can not recover minor damage, individual major tasks fail
Not well	Key component failure irreparable failure, personnel irrecoverable moderate damage, multiple major tasks failed
Terrible	Failure of multiple key components leads to equipment damage, heavy casualties, and core mission failures

3.3 Risk Scenario Fuzzy Evaluation

After determining the probability level and the risk consequence level, a comprehensive evaluation of risk scenarios was carried out. Various indicators used to measure the performance of the evaluated object have a degree of uncertainty. This uncertainty may lead the quantification difficultly. To make the impact of uncertainty as small as possible. A triangular fuzzy numbers method is used. This method can transform the fuzzy evaluation to a certainly value, it can deal with the contradiction that evaluated object cannot be measured precisely [11–13].

In HMIS risk analysis, we use set (good, not bad, normal, not well, terrible) as the risk consequence level, and set (rare, normal, common, and frequent) as the probability level. According to the triangular fuzzy numbers theory, the triangular fuzzy numbers can be given as Tables 3 and 4.

Table 3. Risk consequence level triangular fuzzy numbers table

Level	Evaluations	Triangular fuzzy numbers
1	Good	(0, 0, 0.25)
2	Not bad	(0, 0.25, 0.5)
3	Normal	(0.25, 0.5, 0.75)
4	Not well	(0.5, 0.75, 1)
5	Terrible	(0.75, 1, 1)

Table 4. Risk consequence level assessment table

Level	Evaluations	Triangular fuzzy numbers
1	Rare	(0, 0, 0.33)
2	Normal	(0, 0.33, 0.66)
3	Common	(0.33, 0.66, 1)
4	Frequent	(0.66, 1, 1)

In the Sect. 2, there are three risk scenarios and the factors of each scenario, mark as X_{ij}. If there are K experts participating in the evaluation, the evaluation result mark as $W_{ij} = \left\{W^1, W^2, \cdots, W^n, \cdots, W^k\right\}$,$W^n$ means the risk consequence level of the n^{th} expert on the X_{ij}. The probability level marks as $P_{ij} = \left\{P^1, P^2, \cdots, P^n, \cdots, P^k\right\}$,$P^n$ means the probability level given by the n^{th} expert on the X_{ij}. After performing fuzzy operations on X_{ij} evaluation result matrix and probability matrix, we can obtain an evaluation result of X_{ij}, mark as A_{ij}.

$$A_{ij} = \frac{1}{k} W_{ij} \cdot P_{ij} = \frac{1}{k}\left(W^1 \otimes P^1 \oplus W^2 \otimes P^2 \oplus \cdots \oplus W^k \otimes P^k\right) \quad (1)$$

In fact, A_{ij} is still a triangular fuzzy numbers, it needs de-fuzzy. To remove the fuzzy number in triangular fuzzy numbers (a, b, c), we can use the formula (2).

$$e = \frac{a + 2b + c}{4} \quad (2)$$

In formula (2), e is the evaluation parameters of risk scenarios. risk scenarios can be assessed by comparing the evaluation parameter.

4 Case Study

In this section, the UAV system is used as a case. Unmanned aerial vehicle (UAV) is a typical HMIS. The system environment factors are complex and changeable. Taking UAV control system as a case has strong representativeness.

4.1 Interaction Interface Analysis

The human-machine interaction interface analysis of UAV started from three types of risk factors, equipment factors, environment factors and human factors.

- The equipment components of the UAV system include functional modules, security modules and auxiliary modules. Function modules include LCD, central computer, keyboard, mouse, communication equipment, radar, indicator lights, ship access equipment, etc. The safety module includes insulation, console box, safety belt, warning light, short-circuit indicating device, etc. Auxiliary modules include electronic components, power systems, emergency power supplies, ship access equipment, seats, consoles, console fixtures, charging equipment, and more.

- The environment factors of shipborne UAV is special because the ocean is complex and changing, in this case, the main environmental issues are follows. The thermal environment includes relative humidity, temperature difference between day and night, heat radiations, air flow, the sound environment, the vibration environment and the light environment. Other environment includes air quality, oxygen concentration and external invasion.
- The operators involved in this system include two people in front of the console and one person in charge of the portable controller. In the operator condition, limited by the operating range and the irregular task time, there is a certain degree of physical fatigue, psychological fatigue, and mental fatigue. The impact of decision mainly focused on working memory, perceived time and attention resources. There are many categories of human behavior, but most of them are simple repetitive behaviors, such as task behavior including keyboard input, mouse click, and trackball slide, check behavior and maintains behavior.

After the equipment factors analysis, environment factors analysis, and human factors analysis, we can complete the table with the factors in the interaction processes as Tables 9 and 10 in Appendix.

4.2 Risk Scenario Analysis

According to the analysis in Sect. 2, the risk scenarios in UAV system are categorized into three types.

- The first kind of risk scenario. In terms of dangerous environmental factors, because of the detection at sea, there may be external invaders like birds, which may lead to the crash of UAVs.
- The second kind of risk scenario. When people operate the equipment, because the working area of people is small, and they will sit for a long time, it is easy to cause muscle fatigue. When the equipment feeds back, because a large amount of information is transmitted at the same time, human attention resources are insufficient, thus affecting decision-making judgment.
- The third kind of risk scenario. Vibration is the most obvious environmental impact of a ship as an operating platform. In this system, the operator will be in the vibration environment for a long time, resulting in a series of effects.

After the risk scenarios analysis, the risk scenarios in UAV system are viewed in Table 5.

Table 5. Risk scenarios in UAV system

Type	No.	Description of hazardous scenes
I	1a	UAV in a bird collision danger zone with a high speed
II	2a	Natural tendency behavior when feeling muscle ache
	2b	Operators perform self-test operations which occupy excessive data transmission resources when notice abnormal data in mission-critical areas
	2c	A large amount of information is transmitted at the same time with the insufficient attention
	2d	The information frequency fluctuates frequently which cause a change in attention allocation
	2e	Approaching rest time, the operator has psychological cues to relax and prefers conservative choices
	2f	Inertia thinking when the non-critical alarm information appears repeatedly
III	3a	A vibration environment which cause human fatigue
	3b	A slight relative displacement between the operator and the information display interface which increase the degree of visual fatigue
	3c	Seat belt may pressure to the operators and speed fatigue

4.3 Fuzzy Evaluation of Risk Scenarios

In this section, six experts are invited to assess the factors in risk scenarios. Taking the first risk scenario "UAV in a bird collision danger zone with a high speed" as an example. This risk scenario belongs to the environment-equipment and environment-human interface.

In risk scenario probability level analysis, according to the interface and factor interaction table, the risk factor set of the risk scenario is (UAV function, alarm function, external intrusion, pressure, fear), one of experts score is (1.0, 1.0, 1.0, 1.5, 1.5). The result of this risk scenario probability is 1.2, and this expert's probability assessment of this risk scenario is considered as rare. In risk consequence level analysis, "UAV in a bird collision danger zone with a high speed" may leads to equipment damage, heavy casualties, and core mission failures. Thus, this risk consequence level is considered as terrible.

In UAV system, risk scenario probability level table and risk consequence level table can be obtained by the same method, as shown in Tables 6 and 7.

Table 6. Risk scenario probability level table.

Experts		1	2	3	4	5	6
Hazard	1a	Normal	Rare	Normal	Rare	Rare	Rare
	2a	Normal	Normal	Normal	Normal	Normal	Normal
	2b	Rare	Rare	Rare	Rare	Rare	Rare
	2c	common	common	common	common	frequent	common
	2d	Rare	Normal	Normal	Rare	Normal	Rare
	2e	frequent	common	common	common	common	common
	2f	Normal	Rare	Rare	Rare	Rare	Normal
	3a	frequent	frequent	frequent	frequent	common	frequent
	3b	Normal	common	Normal	Normal	common	common
	3c	Rare	Rare	Rare	Rare	Rare	Rare

Table 7. Risk consequence level table.

Experts		1	2	3	4	5	6
Hazard	1a	Terrible	Terrible	Not well	Terrible	Terrible	Terrible
	2a	Not bad	Normal	Normal	Normal	Normal	Not bad
	2b	Not bad	Not bad	Not bad	Not bad	Good	Good
	2c	Not bad	Not bad	Not bad	Not bad	Not bad	Not bad
	2d	Not bad	Not bad	Not bad	Not bad	Not bad	Not bad
	2e	Not bad	Normal	Normal	Normal	Not bad	Not bad
	2f	Not bad	Not bad	Normal	Not bad	Not bad	Not bad
	3a	Normal	Normal	Normal	Normal	Normal	Normal
	3b	Normal	Normal	Not well	Normal	Normal	Normal
	3c	Normal	Not well	Normal	Normal	Not well	Not well

Taking the first risk scenario "UAV in a bird collision danger zone with a high speed" as an example, risk scenario probability level is P_{1a}, risk consequence level is W_{1a}, as shown in formula (3) and (4).

$$P_{1a} = \{(0,0,0.33),(0,0.33,0.66),\ (0,0,0.33),(0,0.33,0.66),(0,0,0.33),(0,0,0.33)\} \tag{3}$$

$$W_{1a} = \{(0.75,1,1),\ (0.75,1,1),\ (0.5,0.75,1),\ (0.75,1,1),\ (0.75,1,1),(0.75,1,1)\} \tag{4}$$

The evaluation result A_{1a} is calculated with formula (1)

$$A_{1a} = \frac{1}{6}W_{1a}\cdot P_{1a} = \frac{1}{6}\left(W^1 \otimes P^1 \oplus W^2 \otimes P^2 \oplus \cdots \oplus W^k \otimes P^k\right) = (0,0.165,0.5) \tag{5}$$

After de-fuzzification of A_{1a}, the evaluation parameters e_{1a} of risk scenarios X_{1a} is 0.2075. Similarly, the evaluation parameters of other risk scenarios can be calculated, as shown in Table 8.

Table 8. Risk scenarios evaluation parameters table.

Risk evaluation					
No.	1a	2a	2b	2c	2d
Result	0.2075	0.1912	0.0275	0.2125	0.0825
No.	2e	2f	3a	3b	3c
Result	0.2785	0.0887	0.4775	0.2925	0.0515

By analysing the cases of UAV control systems, it can be seen that the evaluation parameters of No. 3a, 3b, and 2e are relatively high, these risk scenarios related to these accidents require more attention. For example, No. 3a is the risk scenario, which may cause human fatigue in vibration environment, we may use the shock absorber in the work space to reduce the impact. No. 3b is the risk scenario that may increase the degree of visual fatigue, we can design a more comfortable system interface to decrease the relative displacement. In risk scenario No. 2e, there should be a training for this approaching rest time situation.

5 Conclusions

This paper establishes a risk analysis method based on risk scenarios for HMIS. Combining with multi-disciplinary knowledge, the factors and interaction interfaces of the whole system are classified, and the types of risk scenarios are sorted to obtain relevant projects for expert evaluation. Fuzzy evaluation method is used to reduce uncertainty and ensure results that are more accurate. Through the case study of the UAV control system, it can be seen that the method can comprehensively and systematically analyze the risk scenarios and has certain application value. However, this method still has some shortcomings. For example, the assessment of risk consequence level and the probability level are highly dependent on expert experience. In addition, there is a lack of systematic methods to assess factors. Future research will focus on introducing appropriate qualitative methods to make this method more systematic.

Acknowledgements. This work is supported by the National Natural Science Foundation of China under Grants No. 71171008/No. 71571004, the National Key Research and Development Program of China, the Special Scientific Research Project for Civil Aircraft of the Ministry of Industry and Information Technology.

Appendix

Table 9. Interaction interface and risk factor analysis table in UAV system

Task	UAV control	Date			Written by		
Task detail		The task time is during the daytime and is in the early stage of long-term task execution. It does not use a portable control device. It only uses the console and no multi-machine coordination tasks is involved					
System composition			Decision interface	Environment-human interface	Information interface	Operation interface	Environment-machine interface
Equipment factors	Function modules	UAV					√
		Radar			√		
		Monitor			√		
		LED			√		
		Communication equipment			√	√	√
		Alarm			√		√
	Security modules	Seat belt				√	√
		LRU					
		Leak-proof				√	
		Console fixation				√	
	Support modules	Seats				√	

Table 10. Interaction interface and risk factor analysis table in UAV system (continued)

System composition			Decision interface	Environment-human interface	Information interface	Operation interface	Environment-machine interface
		Emergency power				√	
Environment factors	Thermal	Temperature		√			
		Coldness		√			
		Humidity		√			
		Airflow		√			
		Radiation		√			√
	Sound	Noise		√			√
	Vibration	Vibration		√			√
	Light	Light condition		√			
		Excess stimulation		√			
	Space	Workspace					
		Stimulation					
	Task	Object	√				
		Task time		√			
	Sport	Acceleration		√			
		Relative velocity					
	Others	Air quality		√			
		Oxygen		√			
		External invasion		√			√
Decision processes	Working memory	Capacity	√		√		
		Allocation	√		√		
	Perceived time		√		√		
	Attention	Capacity	√		√		
		Allocation	√		√		
Human behavior	Task behavior					√	
	Check behavior					√	
	Maintains behavior					√	
Human condition	Psychological	Visual fatigue	√		√		
		Muscle fatigue	√		√	√	
		Mental fatigue	√		√		
	Physiological	Pressure	√	√	√	√	
		Fear	√	√			

References

1. Miller, L.H.: A study in man-machine interaction. In: National Computer Conference, June 1977
2. Bemmel, J.H.V., Gelsema, E.S., Hasman, A., Talmon, J.L., Duisterhout, J.S.: Man-machine interaction in disease pattern classification (1983)

3. Bonney, M., Case, K., Porter, M.: Applications of SAMMIE and the Development of Man Modelling (1989)
4. Miyake, Y.: Co-creation in man-machine interaction. In: IEEE International Workshop on Robot & Human Interactive Communication (2003)
5. Magalhaes, L.P., Gimeno, L., Barther, U., Bittner, H.: Data base for man-machine interaction in a transport environment, pp. 495–498 (1980)
6. Balomenos, T., Raouzaiou, A., Ioannou, S., Drosopoulos, A., Karpouzis, K., Kollias, S.: Emotion analysis in man-machine interaction systems, vol. 3361, pp. 318–328 (2005)
7. Liu, C.C., Yin, T.W., Shan, F.U.: Man-machine-environment complex flight system simulation and human factors quantitative research methods **30**, 58–63 (2013)
8. Teichner, W.H., Olson, D.E.: A preliminary theory of the effects of task and environmental factors on human performance. Hum. Factors **13**, 295–344 (1971)
9. Jin, T., Yang, W., Wang, X., Zhao, T.: Hazard analysis based on human-machine-environment coupling (2011)
10. Lin, L., Chen, Y., Le, Z.: A military aircraft safety risk evaluation method based on "human-machine-environment system". In: International Conference on Future Information Technology & Management Engineering. (2010)
11. Chou, C.C.: The canonical representation of multiplication operation on triangular fuzzy numbers. Comput. Math Appl. **45**, 1601–1610 (2003)
12. Chang, S.C.: Fuzzy production inventory for fuzzy product quantity with triangular fuzzy number. Fuzzy Sets Syst. **107**, 37–57 (1999)
13. Facchinetti, G., Ricci, R.G.: A characterization of a general class of ranking functions on triangular fuzzy numbers. Fuzzy Sets Syst. **146**, 297–312 (2004)

Human Factors Engineering Analysis for Human System Interface of Severe Accident Management Support System Based on Human Factors Engineering Program

Sungheon Lee and Jonghyun Kim(✉)

Department of Nuclear Engineering, Chosun University, 309 Pilmun-daero, Dong-gu, Gwangju 501-709, Republic of Korea
jonghyun.kim@chosun.ac.kr

Abstract. This study aims at performing Operating Experience Review (OER) and Functional Requirement Analysis (FRA) for human-system interface design of Severe Accident Management Support System (SAMSS) through Human Factors Engineering (HFE) program of NUREG-0711. The OER reviews human and organizational issues in Fukushima nuclear power plant (NPP) accident that is a severe accident. It also reviews the operator support system for severe accident management to identify design requirements for the SAMSS. The FRA identifies safety functions and systems required to manage and mitigate severe accident in NPPs. Especially, the safety functions and systems are modeled with the Multi-level Flow Modeling. The result of this study will be used as inputs to the design of SAMSS.

Keywords: Human system interface · Human factors engineering program · Severe accident management support system · Severe accident

1 Introduction

The Severe Accident Management Guideline (SAMG) is a guideline for management and mitigation of Severe Accident (SA) at Nuclear Power Plants (NPPs) [1]. In the case of Korea, operators in Main Control Room (MCR) close the ongoing procedures such as Emergency Operating Procedures (EOPs) and enter the SAMGs when Core Exit Temperature (CET) rises over than 650° of Celsius [2]. Decision making for the SAMG is generally performed by Technical Support Center (TSC) and related actions are mostly carried out by the MCR operators [1].

However, it is known that following the SAMG is one of difficult tasks in NPPs. The current SAMGs generally provides only strategies to mitigate SAs based on symptoms. They do not include specific tasks and criteria for the decision making like EOPs [1]. In addition, according to the lessons learned from previous SAs [3], it is difficult for MCR operators and TSC crew to determine exact state of plant in the SA due to the limitation of available information. Insufficient training is also regarded as a human factors issue relating to the SA management [3–6].

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 131–141, 2020.
https://doi.org/10.1007/978-3-030-20037-4_12

In this light, operator support systems for severe accident management, called severe accident management support system (SAMSS) have been developed. Examples are Accident Management Support Tool (AMST) [7], Computerized Accident Management Support (CAMS) [8], Severe Accident Management EXpert System (SAMEX) [1, 9, 10] and Severe Accident Management and Training simulator (SAMAT) [11].

However, those systems were not successfully applied to actual NPPs. One of reasons for the failure of actual implementation is lack of consideration into human-system interaction and human-factors engineering (HFE). Those were concentrated on the functional capabilities (e.g., accuracy and coverage) but pay little attention to the human-system interaction, like earlier operator support systems in NPPs [12].

This study presents results from two HFE analyses for the design of SAMSS, i.e., operating experience review (OER) and functional requirement analysis (FRA). This study has been performed as a part of the project "Development of Reaction Technologies for Severe Accident Mitigation" led by Korea Atomic Energy Research Institute (KAERI). The OER reviews human and organizational issues in Fukushima NPP accident and functions of SAMSS developed so far. It identifies design requirements for the SAMSS. Then, this study carries out the FRA through a hierarchical analysis method and Multi-level Flow Modeling (MFM). As an example, the FRA on a mitigating function in the SAMG of reference plant, i.e., Steam Generator (SG) Coolant Injection, is provided.

2 Human Factors Engineering Program in NUREG-0711

The OER and FRA for the SAMSS have been performed by following HFE elements suggested by NUREG-0711 human factors engineering program. The objective of HFE element is to verify that the license applicant for NPPs has an HFE design team with the responsibility, authority, placement within the organization, and composition to reasonably assure that the plant design meets the commitment to HFE. Further, a plan should guide the team to ensure that the HFE program is properly developed, executed, overseen, and documented [13]. Figure 1 shows a simplified HFE process suggested by NUREG-0711.

Through OER, a review of the Fukushima accident and the SAMSS developed was conducted, and the result of that, the need for SAMSS to be developed and the design requirements were derived. The FRA was performed to identify the safety functions of SAMSS to be developed.

The objective of the OER is to identify safety issues that are HFE-related. The OER provide information on the performance of the previous design. The issues and lessons learned from operating experience provide a basis to improve the design of NPPs; i.e., at the beginning of the design process in NPPs. The objective of OER is to verify that the applicant identified and analyzed HFE-related issues in previous designs similar to the current one under review. In this way, the negative system of previous designs may be avoided in the current one, while retaining the positive system. The OER must consider the previous systems upon which the design is based, the technological approaches selected, and the HFE issues of NPPs.

The FRA identifies those functions of NPPs that should be performed to satisfy the overall of NPPs operating and safety objectives and goals: To ensure the health and safety of the public by prevention or mitigation the consequences of postulated accidents. FRA to define the high-level functions that must be accomplished to meet the goals of plats and desired performance delineate the relationships between high-level functions and the NPPs systems (e.g., plant configurations or success paths) responsible for performing the functions provide a framework for determining the roles and responsibilities of personnel and automation [13].

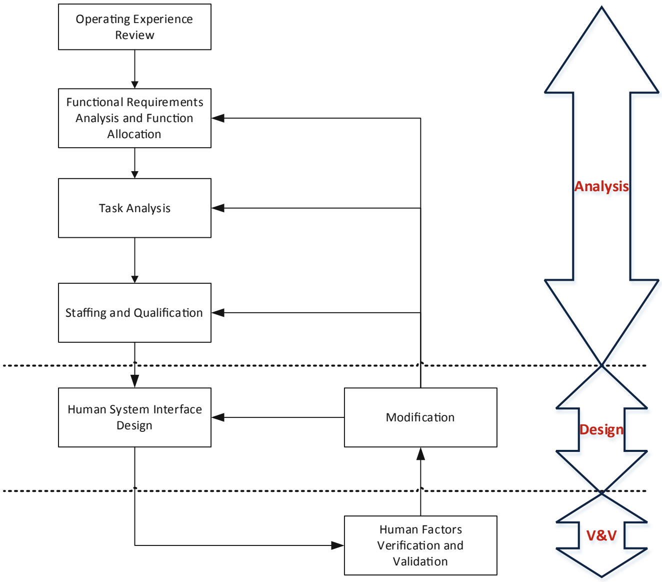


Fig. 1. Process of human factors engineering program in NUREG-0711 for severe accident management support system [13]

3 Operating Experience Review for Design Requirements of Severe Accident Management Support System

For the OER activity, this study reviewed human- and organization-related issues in the Fukushima NPP accident and system functions in the SAMSS's that have been already developed. Based on this review, design requirements for the SAMSS are also identified.

3.1 Review of Fukushima NPP Accident with Human and Organizational Factors Issues

The Fukushima Daiichi Accident is one example of failures in the application of SAMGs. It was the worst accident involving an extreme natural disaster and hydrogen explosion causing radiation to be released outside the plant. In this accident, it is reported that operators did not fully understand that reactor behavior would lead to core damage, although the SAMG presumes that the operator can perform accurately and fast at the time of the accident.

Several human and organizational issues in responding to SA in Fukushima NPP were reported. Some representative issues are as follows.

Training for responding to severe accidents should be conducted thoroughly: Tokyo Electric Power Company (TEPCO) did not have a training simulator for Boiling Water Reactor-3 (BWR-3). Therefore, it was reported that the operators had never operated actual Isolation Condenser (IC) valves [4, 14].

Reliability of the power system should be strengthened: The Fukushima NPP SA was recorded as the worst SA ever caused by natural disasters. During the loss of AC power, the operators relied on portable lights such as flash and mobile phone and then could access very limited information and plant variables. If there was an emergency power source available, or if power could be recovered quickly, the consequences of the accident would be far less serious [14, 15].

NPP monitoring including instruments critical to accident response should be reinforced: In the SA of Fukushima NPP, the lack of direct information on the plant condition, especially, the state of the reactor, caused great difficulties in accident handling and mitigation. Loss of power is one of the major causes, but the instrument itself has lost its function due to the failure. The failure of reactor water level sensor misled the operators to think that the core is not melt down. Therefore, the performance of important sensors and instruments need to be guaranteed even in the extreme condition [3, 4, 14, 15].

3.2 Status of Technology Development in Severe Accident Management Support System

The review of the status of technology has been performed to identify the design requirements for the SAMSS. Four SAMSS's that were already developed are introduced in this chapter.

Severe Accident Management and Training Simulator (SAMAT) developed by KAERI. SAMAT is a system to systematically provide all available additional information to eliminate uncertainties in SA as much as possible, provide the information about plant conditions such as key variables for severe accidents, provide SAMG-related information, verify which strategies can be used to proactively predict plant behavior, and select the best strategy for mitigating severe accidents. SAMAT is based on the logic of SAMG for Korea Standard Nuclear Plant (KSNP), now called Optimized Power Reactor 1000 (OPR1000). It consists of four parts: a training simulator, a variable SPDS, a handbook, and a knowledge base. The training simulator module can virtually perform a strategy; the severe accident SPDS module identifies the status of the plant; and the knowledge-based module contains critical accident scenarios to enable operators to utilize a variety of information when carrying out SAMGs [11].

Severe Accident Management EXpert System (SAMEX) is also developed by KAERI. SAMEX is used when the design basis accident (DBA) of NPP develops into a SA, but even before that, it can be used as a means to predict and respond to the progress of the accident in advance. It can be also used for the purpose of training related to SA in the TSC. Because the existing SAMG mitigation strategies only provide guidance regarding the supply of coolant for the required systems, temperature, pressure, hydrogen concentration, and control of fission products, SAMEX can be used as a means of supplementing them [1, 9, 10].

Accident Management Support Tool (AMST) is a support system developed for the WWER-1000 plant at the Sharif University of Technology in IRAN. AMST is a support system consisting of a tracker to diagnose an accident, a Predicator to predict the progress of an accident, and a decision support function [1, 7].

Computerized Accident Management Support (CAMS) is the support system proposed by the OECD Halden Reactor Project (HRP). The CAMS consists of Signal Validation, Tracking Simulator, Predictive Simulator, Strategy Generator, Critical Function Monitoring, and Man-Machine Interface (MMI) [1, 8].

3.3 Design Requirements from Operating Experience Review

Total 16 design requirements for SAMSS were identified through the OER. Table 1 shows the design requirements with their sources.

Table 1. Design requirements for the developed severe accident management support system

No.	Design requirements	Ref. no.
1	The alarm is required to notify the operator of the occurrence of a severe accident	[3, 16]
2	Support functions (systems) are required to enable response organizations to perform procedures or guidelines quickly and accurately	[4, 15]
3	Even in vulnerable environments such as loss of power, the system should provide information for operators	[17, 18]
4	Functions are required to enable operators to respond creatively to accidents in situations that differ from procedures or guidelines	[4, 14]
5	The ability to support collaboration is required when multiple organizations participate in accident response	[4, 16]
6	In the event of a severe accident, functions are required to predict the behavior of the NPP, progress of accident, and release of radiation sources	[1, 7, 10]
7	Functions are required to support operator decision making in the event of a severe accident	[7, 15]
8	Functions are required to diagnose the cause of a severe accident	[7, 9]
9	Functions are required that support the accident management strategies of response organizations	[3, 17, 19]
10	Functions are required to monitor the major safety functions in the NPPs	[14, 15]
11	The capability to collect the information and to assess the state of plant is required	[15, 17]
12	Functions are required to inform the operator of the possibility of core damage in advance	[3, 4, 17, 19]
13	Functions are required to provide the operator with the information about the inventory of the reactor core	[3, 4, 17]
14	If there is a need to switch from EOP to SAMG, a function to inform it is required	[19]
15	Functions are required to monitor the condition of the containment and the core	[19]
16	Functions are required to inform that the plant has reached a controlled, stable state	[19]

4 Functional Requirement Analysis for Severe Accident Management Support System

For the FRA, this study identifies safety functions for the management and mitigation of SA through a hierarchical analysis. Then, those functions have been modeled by using the MFM.

4.1 Identification of Safety Functions in Severe Accident Management Guideline

Total seven safety functions have been identified on the basis of the OPR1000 SAMG developed by KAERI [2, 20, 21]. The hierarchical structure of SAMG functions is presented in Fig. 2. The ultimate goal of SAMG is the prevention of radiation release. Then, the sub-goals of safety functions are divided into two: (1) cooling down and depressurization of reactor and (2) maintaining the integrity of containment.

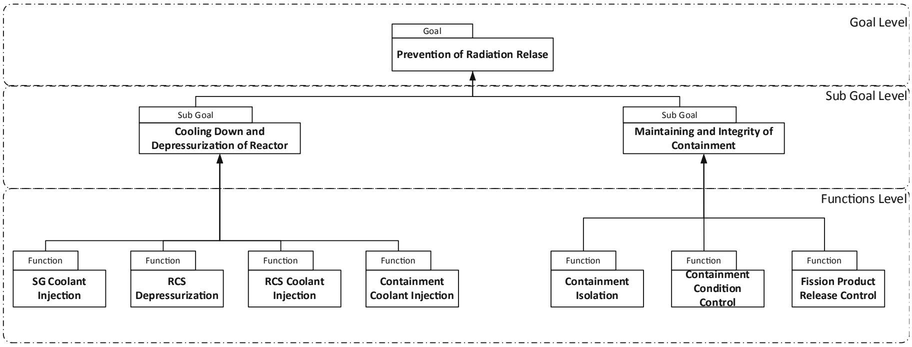


Fig. 2. Safety functions for prevention of radiation release

Four safety functions are identified with relation to the cooling down and depressurization of reactor [2, 20, 21].

- SG Coolant Injection: coolant injection into the SG for Reactor Coolant System (RCS) heat removal and SG tube breakage prevention.
- RCS Depressurization: through depressurization of RCS, enabling the replenishment in RCS using Low Pressure Safety Injection (LPSI), and protection of the Shutdown Cooling System (SCS).
- RCS Coolant Injection: through coolant injection into the RCS, cooling the core and ensuring reactor vessel protection.
- Containment Coolant Injection: through injection into the containment, prevention and delay of reactor vessel damage.

Three safety functions are identified for maintaining the integrity of containment as follows [2, 20, 21].

- Fission Product Release Control: reducing the risk of exposure to people near the NPP, during an SA from the in-containment.
- Containment Condition Control: controlling temperature, pressure, hydrogen and fission product concentration in the containment.
- Containment Hydrogen Control: preventing and controlling the hydrogen explosion in the containment.

This study also identifies the systems that can be applied to achieve the safety functions, also called success paths. The success paths have been identified by the review of piping and instrumentation diagram (P&ID) of reference plants. Table 2 shows the safety functions and their success paths [2, 20, 21].

Table 2. Identified safety functions and their success paths

Ultimate goal	Sub goals	Safety functions	Success paths
Prevention of Radiation Release	Cooling and Depressurization of Reactor	SG Coolant Injection	Auxiliary Feed-water System Main Feed-water System External Injection System
		RCS Depressurization	Reactor Coolant Gas Vent System Safety Depressurization system Pressurizer Auxiliary Spray SG Steaming
		RCS Coolant Injection	Safety Injection System Chemical Volume & Control System External Injection System Containment Spray Pump
		Containment Coolant Injection	Containment Spray System RWT Gravity Drain System
	Maintaining the Integrity of Containment	Containment Isolation	Containment Isolation System
		Containment Condition Control	Containment Cooling System Containment Spray System Combustible Gas Control System Passive Autocatalytic Recombiner (Non-Power)
		Fission Product Release Control	Containment Fan Cooler Containment Isolation System Containment Spray System

4.2 Multilevel Flow Modeling

MFM is a modeling technique proposed by Morten Lind that can easily model complex industrial processes, e.g., NPPs. MFM is a useful technique to deduce systems into multiple stages by applying the concepts of means-end and whole-part. The MFM model divides goals and functions of the system into Mass, Energy and Information, and represents the relationship between the functions associated with its flow. Through the Cause-Consequence or Goal-Means modeling of the system, it is possible to identify the cause of the system failure and the consequences of the system failure [22–24].

This study presents the MFM modeling for the SG Coolant injection as an example for the safety functions.

4.3 Steam Generator Coolant Injection Using Process Modeling

Before building an MFM model, a process model for SG Coolant Injection has been developed from the hierarchical analysis in the previous section. The process model of SG Coolant Injection includes Auxiliary Feed-water System (AFWS) (1, 2), Main Feed-water System (MFWS) and External Injection System (EIS) (1, 2). Figure 3 shows a simplified success paths diagram of SG Coolant Injection function using a process modeling tool of a MFM program [25].

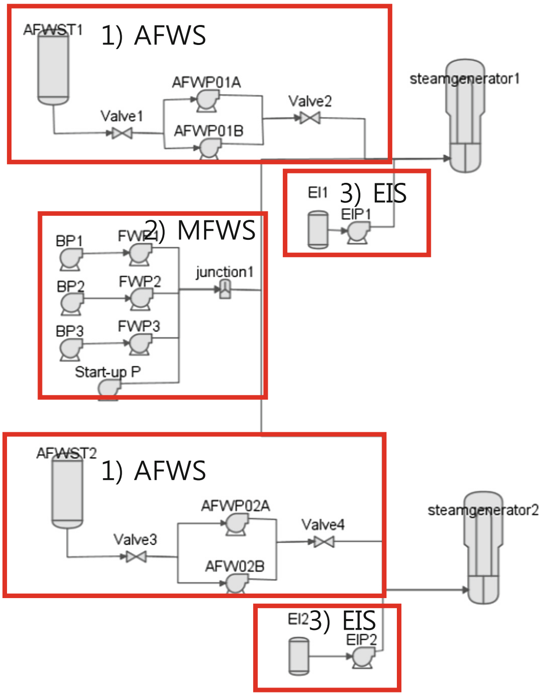


Fig. 3. SG Coolant Injection modeling using process model of MFM

4.4 Steam Generator Coolant Injection Using Multilevel Flow Modeling

Based on the process model in Fig. 3, an MFM model has been developed as shown in Fig. 4. The MFM model consists of three levels. The first level represents the goal structure (red rectangle) in the Fig. 4. It uses the cold-leg and hot-leg temperatures to ensure that the RCS heat is being removed. The second level represents the energy flow structure (green rectangle) that shows the heat exchange between Nuclear Steam Supply System (NSSS), i.e., hot side and success paths, i.e., cold side. The third level represents the mass flow structure (blue rectangle). It includes components and mass flow path in the success paths (AFWS, MFWS, EIS) of SG Coolant Injection function as well as in the NSSS.

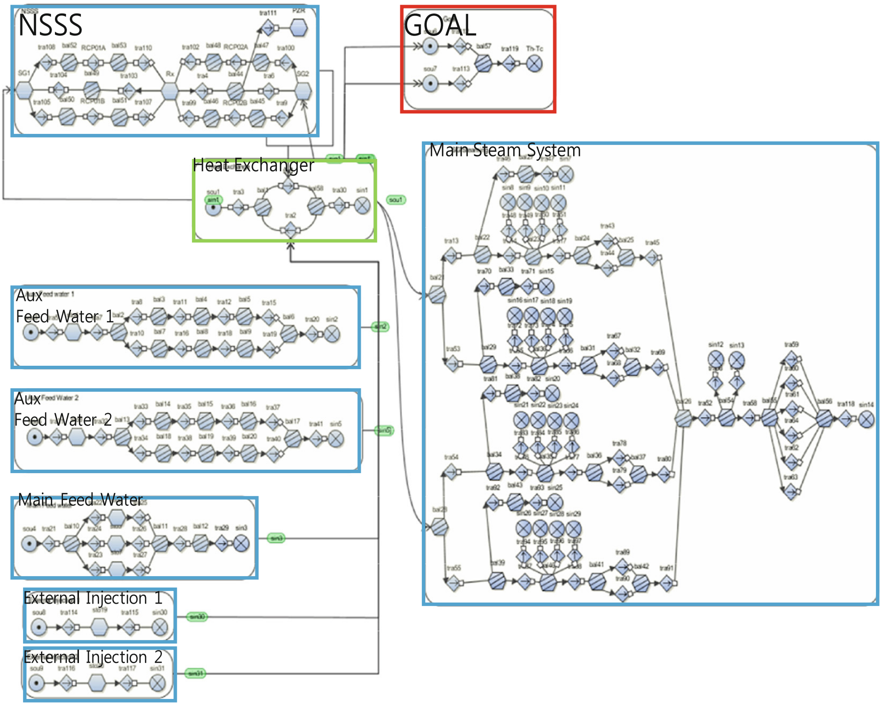


Fig. 4. SG Coolant Injection modeling using MFM model

5 Application

This study identified design requirements for the SAMSS through the OER and analyzes safety functions for the management and mitigation of SA. The design requirements will be directly used for the input to design the HSI of SAMSS.

The safety functions and success paths identified in Sect. 4 will be used for the function allocation and task analysis which follow the FRA in Fig. 1. The function allocation determines whether the safety functions are performed by automatic systems, human operators, or the combination of both. Then, task analysis defines tasks for achieving the function, and identifies information, control, and task support that are also used as inputs for the HSI design of SAMSS.

The structure of MFM model will be used in the HSI as a support to help operators understand plant situations in the SA. It can represent the causal relation between functions, systems, and component. It is also expected to provide diagnostic information when any goal or function is not achieved.

6 Conclusion

This study carried out the OER and FRA for designing the HSI of SAMSS in NPPs. In OER, a literature review on Fukushima NPP accident was conducted to identify human and organizational issues regarding severe accident management and mitigation. Furthermore, the survey on the operator support systems that have been developed to help severe accident management was conducted. As a result of OER, 16 design requirements were identified for SAMSS. The FRA identifies seven safety functions for the SAMG of OPR1000, and success paths for satisfying safety functions. The safety functions and success paths were modeled using the MFM. The result of this analysis will be used as an input to the design stage of SAMSS.

Acknowledgments. This research was supported by Nuclear R&D Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (Grant No. 2017M2A8A4017947).

References

1. Korea Atomic Energy Research Institute (KAERI): Development of an Optimum Severe Accident Management Decision-Support System (SAMEX). KAERI/TR-4686/2012 (2012)
2. Ha, J.-J., et al.: Korea Atomic Energy Research Institute (KAERI). Development of Accident Management Guidance for Korea Standard Nuclear Power Plant. KAERI/RR-1939/98 (1999)
3. Johson, G.: EPRI: severe nuclear accident; lessons learned for instrumentation, control and human factors (2015)
4. Back, W., et al.: Commission of Fukushima in Korea Nuclear Society. Fukushima Nuclear Power Plant Accident Analysis (Final Report) (2013)
5. Rempe, J.L., Knudson, D.L.: Department of energy national laboratory (Idaho). TMI-2 a case study for PWR Instrumentation Performance During a Severe Accident. INL/EXT-13-28043 (2013)
6. International Atomic Energy Agency (IAEA): The Chernobyl Accident: Updating of INSAG-1. Safety Series No. 75-INSAG-7 (1992)
7. Saghafi, M., Ghofrani, M.B.: Department of energy engineering in sharif university of technology. Introduction of a research project on development of accident management support tool for BNPP (WWER-1000) based on the lessons learned from Fukushima accident (2015)
8. Fantoni, P., et al.: Institute for Energy Technology (IFE). The CAMS Prototype (1995)

9. Park, S.-Y., Ahn, K.-I.: SAMEX: a severe accident management support expert. Ann. Nuclear Energy **37**, 1067–1075 (2010)
10. Park, S.-Y., Ahn, K.-I.: Korea Atomic Energy Research Institute (KAERI). Development of an Optimum Severe Accident Management Decision-Support System (SAMEX). KAERI/TR-4686/2012 (2012)
11. Jeong, K.-S., Kim, K.-R., Jung, W.-D., Ha, J.-J.: Development of severe accident management and training simulator (SAMAT). Ann. Nuclear Energy **29**, 2055–2069 (2002)
12. Kim, J., Seong, P.H.: The effect of information types of diagnostic strategies in the information aid. Reliab. Eng. Syst. Saf. **92**, 171–186 (2007)
13. U.S. NRC: Human Factors Engineering Program Review Model Rev-3. NUREG-0711, rev-3, pp. 17–35 (2012)
14. Director General International Atomic Energy Agency (IAEA): The Fukushima Daiichi Accident (2015)
15. Institute of Nuclear Power Operations (INPO): Lessons Learned from the Nuclear Accident at the Fukushima Daiichi Nuclear Power Station-Revision 0. INPO 11-005 (2012)
16. Kurokawa, K.: Nippon Genshiryoku Gakkai-Shi: Fukushima nuclear accident independent investigation commission by the National Diet of Japan (2013)
17. International Atomic Energy Agency: Nuclear Human and Organizational Factors in Nuclear Safety in the Light of the Accident at the Fukushima Daiichi Nuclear Power Plant (2013)
18. Nuclear Regulation Agency in Japan: Analysis of the TEPCO Fukushima Daiichi NPS Accident (2014)
19. International Atomic Energy Agency: Accident Monitoring system for Nuclear Power Plants (2015)
20. Korea Atomic Energy Research Institute (KAERI): Development of Accident Management Guidance for Korean Standard Nuclear Power Plant, Technical Background for Severe Accident Management Guideline. KAERI/RR-1939/98 (1999)
21. Korea Atomic Energy Research Institute (KAERI): Development of Accident Management Guidance for Korea Standard Nuclear Power Plant, Severe Accident Management Guideline. KAERI/RR-1939/98 (1999)
22. LIND Morten: Technical University of Denmark. An introduction to multilevel flow modeling: nuclear safety and simulation, vol. 2, no. 1 (2011)
23. LIND Morten: Technical University of Denmark.: Control functions in MFM: basic principles: nuclear safety and simulation, vol. 2, no. 2 (2011)
24. LIND Morten: Technical University of Denmark. An overview of Multilevel flow modeling: nuclear safety and simulation, vol. 4, no. 3 (2013)
25. MFM-Multilevel Flow Modeling. https://mfm.safepark.dk

Optimization of Information Interaction Interface Based on Error-Cognition Mechanism

Xiaoli Wu[1(✉)], Chengqi Xue[2], Feng Zhou[3], Linlin Wang[1], Weiwei Zhang[1], Tingting Zhang[1], and Ling Xia[1]

[1] College of Mechanical and Electrical Engineering, Hohai University, Changzhou, China
wuxlhhu@163.com, 353248644@qq.com, 962698857@qq.com, {zhangtingting, xialing}@hhu.edu.cn

[2] School of Mechanical Engineering, Southeast University, Nanjing, China
ipd_cqx@seu.edu.cn

[3] College of Mechanical and Electrical Engineering, Donghua University, Shanghai, China
1848001981@qq.com

Abstract. This paper proposed a framework called the "Error-Cognition-Design", which studies the association between error factors, visual cognition and design information. We can combine extended operator's cognitive behavior model with corresponding analytical model of error recognition, and further recognize the cognitive error of the interfacial task. According to the process of operator's cognitive behavior, operator's task, task function as well as task steps and structure are unfolded, corresponding with the analytical opinion of human reliability, which includes task analysis, objection analysis, operation analysis, planning analysis, error analysis, psychological error mechanism analysis, performance shaping factor analysis as well as human error identification in systems tool analysis. Based on the results, we proposed an optimization design scheme for the monitoring task interface. According to the reaction chain of error factors and information features, the avionics interface display can be optimized via the information symbols design and the information block layout.

Keywords: Complex information systems · Error · Design factors · Error-cognition · Mechanism

1 Introduction

With the rapid development of digital and intelligent information systems, display of visual information in interfaces has become an important challenge in the field of human-computer interaction. Transportation hub monitoring system, nuclear power control system, environmental monitoring system, and other large systems have evolved from the traditional control mode to digital control mode. Because digital information interaction interfaces are characterized by the large quantity of information and

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 142–154, 2020.
https://doi.org/10.1007/978-3-030-20037-4_13

complicated information relationships, operators may enter the complex cognition and lead to task failures and even serious system failures or major accidents due to operators' slipping, misreading, misjudgment, late feedback, and other cognitive difficulties.

The human-computer interaction industry has long been exploring reasonable and feasible design methods to improve the information presentation problem faced by current digital, intelligent visual information interfaces. Especially in the field of complex information systems, researchers both at home and abroad have devoted time and effort to seek reasonable methods for information display through research on information coding and layouts. For instance, in the Human Measures and Performance Project (HMPP), NASA specifically studied the problems of color security and availability of the design of various complex graph display interfaces in the aviation field. Yeh and Wickens [1] carried out experiments to investigate how to best exhibit relevant electronic information on battlefield maps. Montgomery and Sorkin [2] performed an experimental study on the effect of luminance on an observer's degree of reliability in identifying information in a visual display. Tullis [3] and Schum [4] investigated the efficiency of identifying digital and graphical information coding. Monnier [5] applied the experimental paradigm of visual delay in search tasks to study the relationship between colors and locations. Parsons et al. [6] summarized ten attributes of interactive visual displays (including importance, relations and adjustability of interactions). Li [7] in 1984, at an early stage of the aviation industry in our country, analyzed and investigated the usage of circular scale instruments in different types of domestic airplane pilot seats to propose suggestions to improve their instrumental scales, pointers, digital display, benchmark of flight attitude, instrumental size and layout. Liu and Ding [8] studied the recognition effects of the relative locations of display interface in instrument panels of fighter planes and helicopters. Neyedli et al. [9] performed research on the information representation of auto-battle recognition system. Zhou, Li, Niu, Wu, Jin, and Xue et al. [10–18] (2013–2015) performed physiological research on the information representation and interface design of human-machine interfaces of complex systems. All these works of research indicate the importance of information presentation in the process of monitoring task execution. They conclude that icons, symbols and colors are critical styles of expression for information presentation.

This paper proposes a framework called the "Error-Cognition-Design", based on our study of the association between error factors, visual cognition and design information. This framework offers a technique that can map the mechanism of error-cognition into the design factors of information interfaces, to perform optimization for the design of complex information interfaces.

2 Analytical Model of Cognitive Error Recognition

We can combine extended operator's cognitive behavior model with corresponding analytical model of error recognition, and further recognize the cognitive error of the interfacial task. Embrey, Altman and Swain et al. [19, 20] tried to use the basic behaviour component of the operator to describe the behavior of the operator with "error" event characteristics from the view of traditional human factors; PHEA and HRMS et al. established the analysis model of human factor from the perspective of cognitive psychology. The technology of HERA proposed by Kirwan [21] which integrates several methods and enables the analysis results to be tested each other, is reliable relatively. Thus, this paper will apply error recognition framework to analyze the cognitive error recognition of information search, recognition, judgment and selection, as well as decision-making. According to the process of operator's cognitive behavior, operator's task (as shown in Fig. 1), task function as well as task steps and structure are unfolded, corresponding with the analytical opinion of human reliability, which includes task analysis, objection analysis, operation analysis, planning analysis, error analysis, psychological error mechanism analysis, performance shaping factor analysis as well as human error identification in systems tool analysis.

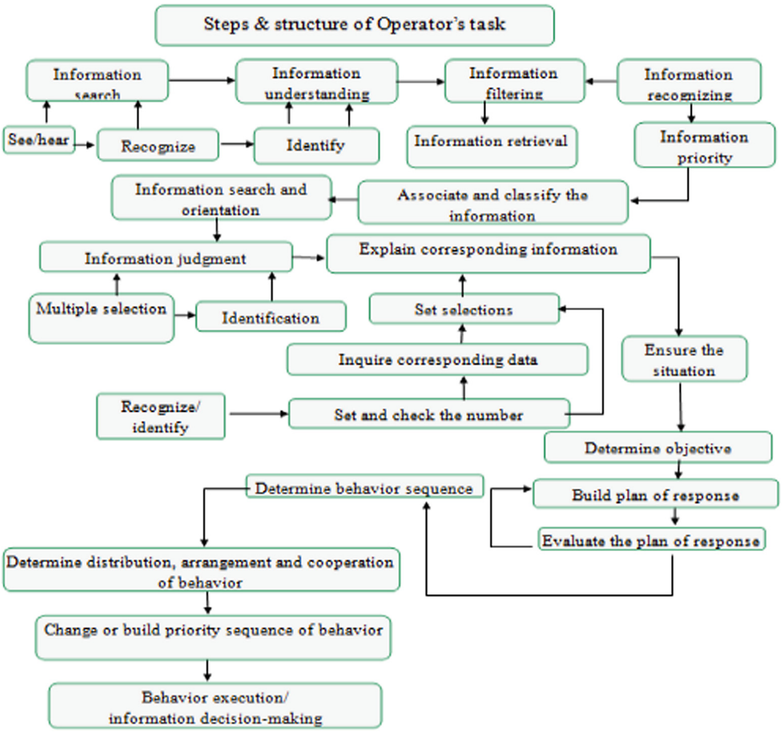


Fig. 1. Error recognition analytical model of operator's cognitive behavior

Digital information task interface is characterized by transforming systematic abstract information into user interfacial elements which are easy to identify and understand. Graphical user interface conveys several elements, including character, text, image, icon, colour, dimension, and so on. When the information displayed is complex, only the reasonable navigation design and structure design of information hierarchy can reach the rationality of information interaction. Thus, the design problem of information interaction interface has evolved into a hot spot and focus problem which was concerned mutually by researchers in human-factors engineering, automatic control, cognitive psychology, systematic science, design science and other disciplines. Then, whether the design factors of information interaction interface could begin with the source of task failures – error factors? The key point lies in how to understand correctly the interaction mechanism between 'error and cognitive', then, can we propose a reasonable design strategy for the optimization of visual information interface. The features of design information, such as locations of information characteristics, visual range, separation distance and visual attention intensity of symbols, are determined in accordance with the reaction chain of error factor-information characteristics. We find a solution to this problem based on error-cognition mechanism.

3 Analytical Model of Cognitive Error Recognition

3.1 Error Factors of Information Interaction Interface

In information interaction interfaces, for example an aviation information display, there are several possible tasks to be executed, such as monitoring status data, querying task information, monitoring threat and security state information, and so on. Display interfaces of complex information system display navigation, situation pictures, status data and other information. The monitoring task is likely to be performed: plan creating, state monitoring, burst scheduling, and so on. We can classify the monitoring interfacial task either by abrupt events and common tasks, or by the order in which tasks are performed. Thus, as shown in Table 1, we listed the monitoring interface tasks and corresponding error factors to extract the error characterization of a monitoring interface of a complex system.

3.2 Information Features of Information Interaction Interface

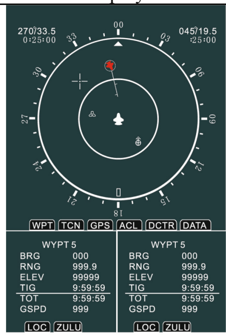
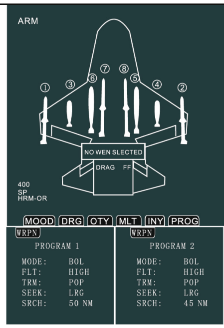
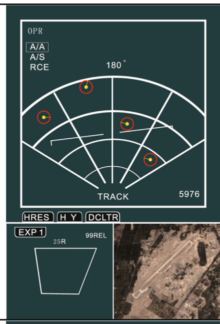
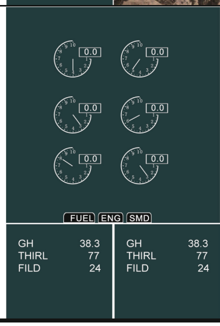
As the analysis object of the design information characteristic, the monitoring task interface of some kind of navigation war is divided into four parts based on different tasks: (i) radar situation interface, (ii) weapon mounting interface, (iii) multi-sensor interface and (iv) flight data display interface, including the navigation, situation charts, state data, alarm reminder and other information display. There are four processes in monitoring tasks which may be performed such as monitoring/detection, state query, response planning and response execution. Table 2 shows the information content displayed in the different monitoring task interfaces, which will be regarded as the content of main information in searching, reading, recognition, judging selection and decision making. The radar situation interface mainly displays the information of the

Table 1. Characterization of error factor of monitoring interface task [22]

Tasks of monitoring interface	Display format of information	Cognitive behavior	Error factor	Representation of error
Monitor/discover Inquire state Plan response Execute response	Dynamic display Static display Navigation Status data Information icon Alarm reminder	Search Recognize Identify Judge & select Decision-making	Ignorance Omission Miss Misreading Misjudgment Misunderstanding Haven't seen Confusion Cannot remember Input error Misregistration Cannot see clearly Hard to distinguish Match incorrectly Cannot find Delay Inadequate Irrelevant React too early No Reaction Select incorrectly Slip	Ambiguity states Visual limitation Visual bluntness Visual illusion Attentional load Visual disturbance Overattention Attention shift and distraction E9 too nervous to do anything E10 cognitive bias Unreasonable match Weak visibility Thinking load Forget Inaccurate recall Lack of memory aids Intentionality decrease False memory Unconsciousness Omission caused by inattention Time pressure

aircraft radar area including the appearance of the target, the database calling, and the different aircraft symbols in the range, the attacked target display, the driving route and other information. The weapon-mounting interface mainly displays the selection of weapons, display, launch selection, the current state of selection and the weapon programming, etc.; the multi-sensor display interface mainly displays radar setting and selection, satellite map information, radar proportional rendering and other information. The flight data display interface mainly displays the machmeter, forecast speed, attack angle, height, horizontal meter calibration, fuel and other indicators.

Table 2 Monitoring task interface of typical navigation war display system

Monitoring interface	Interface information		Monitoring interface display
Radar situation interface	Range selection based on proportional rendering	Data contact target display	
	Airs peed command window	Aircraft symbol	
	Height command t window	Internal range circle	
	Radar target display	Command title symbol	
	Database links (D and L) selection	L & S target display	
	FLIR (FLR) selection	Attacked area reminder	
	Radar (RDR) selection	Current driving route information	
	Aircraft range		
Weapon mounting interface	Weapon program (PROG) selection	A / G gun selection	
	launch Interval (INV)	Weapon Selection Station	
	Multiple Transmitter (MULT)	Major Armed Instructions	
	Number of launches (QTY)	Current armed display	
	Mounting place	Weapon station STEP	
	Bombing mode	Weapons programming storage	
	Weapon selection	Ignore HARM protection mode	
Multi-sensor interface	Azimuth Scanning selection (Radar Trigger)	Radar frozen (FRZ)	
	Tactic range and radar range	Static(SIL)radar	
	Mode selection	Radar Declutter (DCLTR) Vector radar scanning selection	
	Expand 1/2/3 times mode	GPS map navigation	
	Range instruction based on proportional rendering		
Flight data display interface	Machmeter and air speed indicator	Horizontal indicator	
	Target indicator	Bank angle indicator	
	Forecasting speed indicator	Altimeter calibration indicator	
	Attack angle loading indicator	Fuel consumption indicator	
	Altimeter indicator		

4 Case of Task Interface Design of Typical Avionics Display Systems

4.1 Error Factor and Extraction of Information Features of a Typical Avionics Display Interface

We will take the surveillance task interface of a complex information system as an example. This paper takes information features of the surveillance task interface of an avionics display system for analysis (as depicted in Fig. 2), in which the error factors are extracted based on the monitoring interface tasks and corresponding error factors (Table 1) as follows:

1. Visual restriction – omission;
2. Visual mistake – misreading/misjudgment;
3. Visual interference – ignorance;
4. Attention shifting and distraction – miss;
5. Cognitive deviation – misunderstanding;
6. Unreasonable matching – confusion.

With respect to six items of error factors, the designers target the monitoring tasks executed by the operators, and combine them with features from design information to determine seven items from the design information features:

1. Location of an information feature;
2. Visible range of an information feature;
3. Spacing of information features;
4. Intensity of visual attention of an information symbol;
5. Recognition of an information icon;
6. Degree of conciseness of an information icon;
7. Differences between information icons.

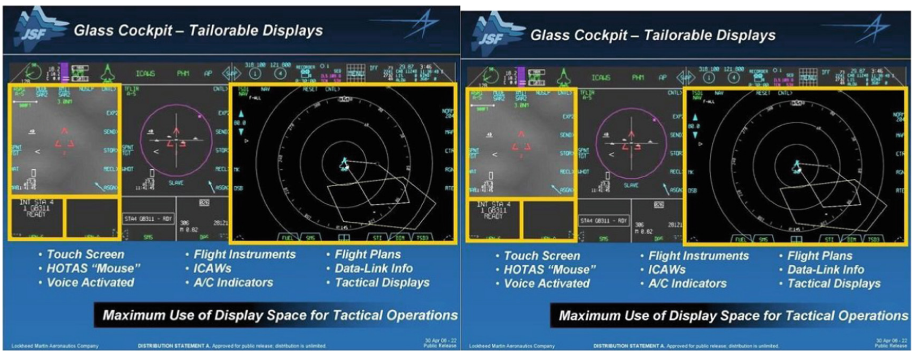


Fig. 2. Monitoring tasks: Mode 1 (left) and Mode 2 (right)

With respect to the design method of complex information interfaces, we explore a use-case for the design of a typical naval warfare display system. Error factors such as omission, misreading/misjudgment, ignorance etc., are extracted along with the description of the relevant information features of tasks for the operators. We identify four significant areas of information display design that can be improved in the monitoring task interface.

Below, we will focus on six items of error factors from these typical error-cognition sets and seven items of design information features extracted by surveillance tasks executed by operators. The analysis will be conducted one by one with respect to the reaction chain. The optimizations of the intensity of visual attention are performed mainly with respect to the line symbol, the character symbol, and a combination of the line and character symbols. We also consider the significance, such as the allocation of colors and line frame symbols which affect the intensity of the visual attention. The relevant symbols of design information are illustrated in Fig. 3.

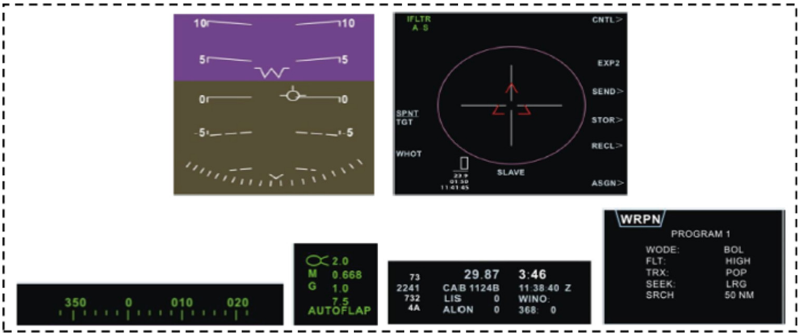


Fig. 3. Information blocks with relatively weaker visual attention

The optimization of the recognition and the distinction of information graphic symbols are mainly from the points of understanding, degree of cognition, similarity of graphic symbols. The relevant symbols of design information are shown in Figs. 4-a, b, c.

With respect to the characterization methods of design factors for interface layout [23], the information structure of the interface can be extracted via abstract layouts. As a result, a layout analysis of the original interface results in the output of the layout abstracts of each sub-interface as depicted in Fig. 5 (A, B, C and D).

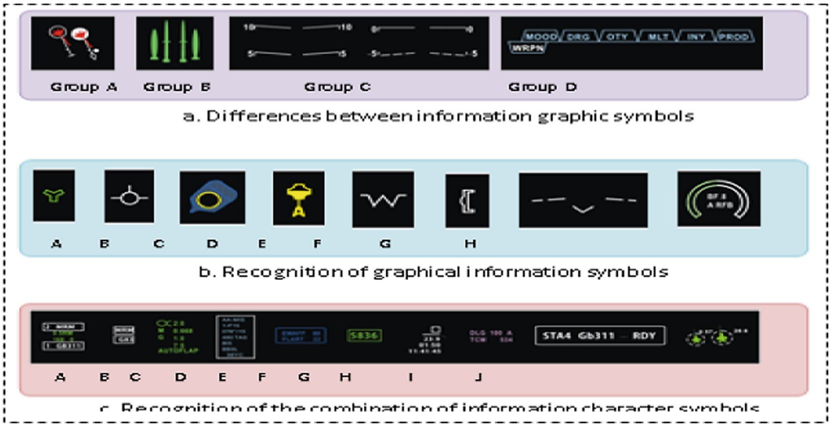
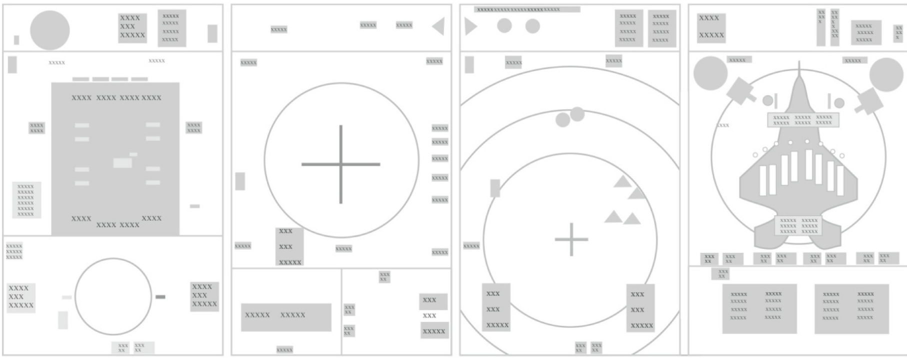


Fig. 4. Information symbols



A. Display of flight data B. Display of horizontal location C. Radar status D. Weapons mounting

Fig. 5. The information layout of a monitoring task interface

It can be seen from the information structure of the interface that this layout can be partitioned into four visual districts, and these districts are divided equally in such a way that they lack concern for visual searching behaviors. What needs to be considered for the factors of information layouts are the locations of information features, the visible ranges of information features and the intervals of the information features.

4.2 Information Display and Interface Design According to Results of Error Factors and Design Features

Based on the results, we also propose an optimization design scheme for the monitoring task interface. According to visual behavioral characteristics and visual searching model, vision is directed from left to right, from up to down, and from the upper left, the upper right, the lower left to the lower right. The distribution of interface layouts is shown, including the main task execution area (optimal visual zone), task execution reserve (secondary visual zone) and task execution reserve (third visual zone). The layout should maximize the task execution area, and hide non-execution areas. It could adjust different information block displays in respect of position and visual range. According to extraction of the error factors, analysis of visual behavior, the reaction chain of error factors and information features, the avionics interface display can be optimized via the information symbols design and the information block layout.

According to visual behavioral characteristics [24] and visual searching model [25], vision is directed from left to right, from up to down, and from the upper left, the upper right, the lower left to the lower right. The distribution of interface layouts is shown in Fig. 6, including the main task execution area (optimal visual zone), task execution reserve (secondary visual zone) and task execution reserve (third visual zone). The layout should maximize the task execution area, and hide non-execution areas. It could adjust different information block displays in respect of position and visual range.

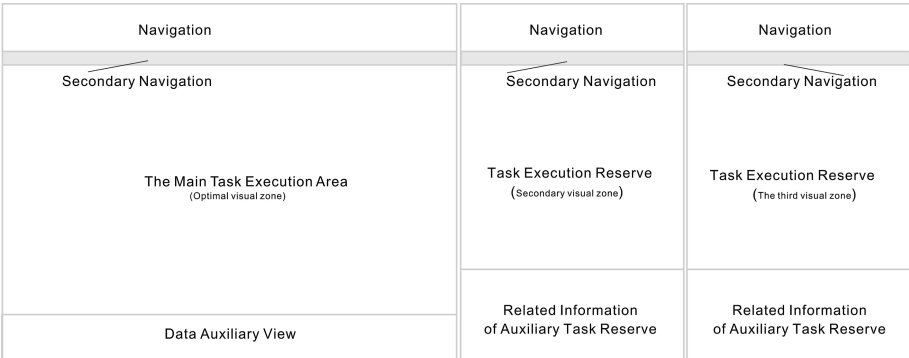
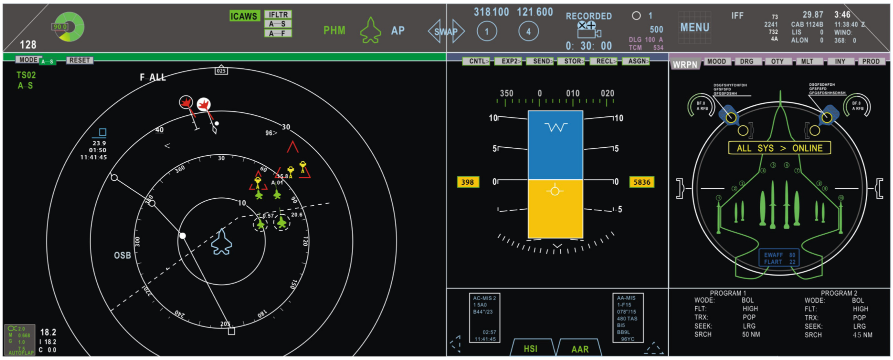


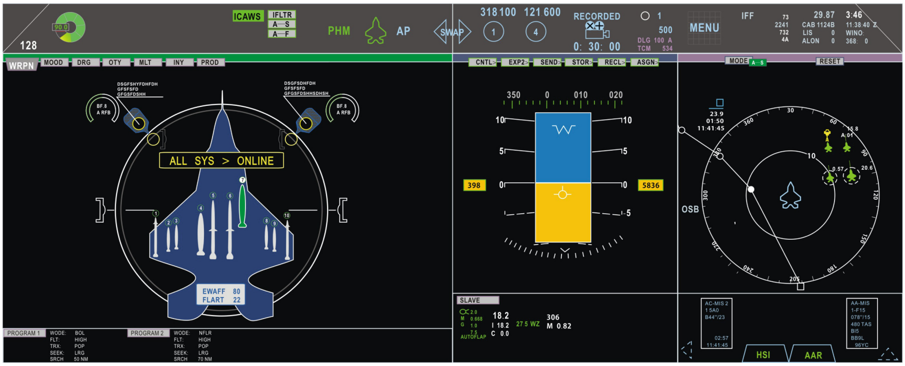
Fig. 6. The information optimized layout of the monitoring task interface

According to extraction of the error factors, analysis of visual behavior, the reaction chain of error factors and information features, and the results of multiple objectives planning, the avionics interface display can be optimized via the information symbols design and the information block layout. The 3 optimized modes interface display are shown in Fig. 7 (Mode 1, Mode 2, and Mode 3).



An optimized interfaces display（Mode 1）

An optimized interface display（Mode 2）



An optimized interface display（Mode 3）

Fig. 7. 3 optimized modes interface display

5 Conclusions

This paper takes information features of the surveillance task interface of an avionics display system for analysis, in which the error factors are extracted based on the monitoring interface tasks and corresponding error factors. With respect to the design method of complex information interfaces, we explore a use-case for the design of a typical naval warfare display system. Error factors such as omission, misreading/misjudgment, ignorance etc., are extracted along with the description of the relevant information features of tasks for the operators. We find a solution to this problem based on error-cognition mechanism. Based on the results, we also propose an optimization design scheme for the monitoring task interface. It could adjust different information block displays in respect of position and visual range.

According to extraction of the error factors, analysis of visual behavior, the reaction chain of error factors and information features, the avionics interface display can be optimized via the information symbols design and the information block layout.

Acknowledgements. This work was supported by Jiangsu Province Key Project of philosophy and the social sciences (2017ZDIXM023), Jiangsu Province Nature Science Foundation of China (BK20181159, BK20170304), the National Nature Science Foundation of China (Grant No. 71601068, 61603123, 61703140), Science and technology projects of Changzhou (CE20175032), Outstanding Young Scholars Joint Funds of Chinese Ergonomics Society- King Far (No. 2017-05), Overseas research project of Jiangsu Province (2017), and Fundamental Research Funds for the Central Universities (Grant No. 2015B22714).

References

1. Yeh, M., Wickens, C.D.: Attentional filtering in the design of electronic map displays: a comparison of color coding, intensity coding, and decluttering techniques. Hum. Factors **43**(4), 543–562 (2001)
2. Montgomery, D.A., Sorkin, K.D.: Observer sensitivity to element reliability in a multi-element visual display. Hum. Factors **38**(3), 484–494 (1996)
3. Tullis, T.S.: An evaluation of alphanumeric, graphic, and color in-formation displays. Hum. Factors **23**(5), 541–550 (1981)
4. Schum, D.A.: The weighting of testimony in judicial proceeding from sources having reduced credibility. Hum. Factors **33**(2), 172–182 (1991)
5. Monnier, P.: Redundant coding assessed in a visual search task. Displays **24**(1), 49–55 (2003)
6. Parsons, P., Sedig, K.: Adjustable properties of visual representations: improving the quality of human-information interaction. J. Assoc. Inf. Sci. Technol. **65**(3), 455–482 (2014)
7. Li, L.: Investigation on engineering psychology of aircraft cockpit instrument on the flight eight. Aviat. Mil. Surg. **1**, 60–61 (1984)
8. Ding, Y., Liu, B., Cao, B.: Simulating the visual effects of various visual districts of main instrumental board of military helicopters. J. Aviat. Med. China **5**(2), 108–111 (1994)
9. Neyedli, H.F., Hollands, J.G., Jamieson, G.A.: Beyond identity: incorporating system reliability information into an automated combat indentification system. Hum. Factors: J. Hum. Factors Ergonoimics Soc. **53**(4), 338–355 (2011)

10. Zhou, L., Xue, C., Tang, W., Li, J., Niu, Y.: Aesthetic evaluation method of interface elements layout design. J. Comput.-Aided Des. Comput. Graph. **25**(5), 758-766 (2013)
11. Wu, X., Xue, C., Tang, W.: Experimental study on visual limitation experiment of goal-seeking in radar situation-interface. J. SE Univ. (Nat. Sci. Ed.) **44**(6), 1166–1170 (2014)
12. Wu, X., Xue, C., Wang, H.: E-C mapping model of human-computer interaction interfaces of complicated systems. J. Mech. Eng. **50**(12), 206–212 (2014)
13. Zhou, L., Xue, C.: Research of interface composition design optimization based on visual balance. In: The 2013 International Conference on Intelligent Systems and Knowledge Engineering, Shenzhen, 20–23 November 2013, vol. 279, pp. 483–493 (2013)
14. Li, J., Xue, C., Wang, H., Zhou, L., Niu, Y.: Encoding information of human-computer interface for equilibrium of time pressure. J. Comput.-Aided Des. Comput. Graph. **25**(7), 1022–1028 (2013)
15. Wu, X., Xue, C., Tang, W.: Study on eye movement of information omission misjudgment in radar situation-interface. In: Engineering Psychology and Cognitive Ergonomics - 11th International Conference, EPCE 2014, Held as Part of HCI International 2014, Proceedings, Crete. LNAI, vol. 8532, pp. 407–418 (2014)
16. Wu, X., Li, J., Zhou, F.: An experimental study of features search under visual interference in radar situation-interface. Chin. J. Mech. Eng. **31**(1), 45–58 (2018)
17. Wu, X.: Error-Cognition Mechanism in Complex Information Task Interface. Science Press (2017)
18. Wu, X., Gedeon, T., Wang, L.: The analysis method of visual information searching in the human-computer interactive process of intelligent control system. In: Proceedings of the 20th Congress of the International Ergonomics Association, IEA 2018, Volume X: Auditory and Vocal Ergonomics, Visual Ergonomics, Psychophysiology in Ergonomics, Ergonomics in Advanced Imaging Florence, Italy, 26–30 August 2018, pp. 73–84 (2018)
19. Embrey, D.E.: SLIM-MAUD: a computer-based technique for human reliability assessment. Int. J. Qual. Reliab. Manag. **3**(1), 5–12 (1986)
20. Swain, A.D., Guttmann, H.E.: Handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG/CR-1278, US Nuclear Regulatory Commission, Washigton (1983)
21. Kriwan, B.: The development of a nuclear chemical plant human reliability management approach: HRMS and JHED. Reliab. Eng. Syst. Saf. **56**(2), 107–133 (1997)
22. Wu, X., Chen, Y., Zhou, F.: An interface analysis method of complex information system by introducing error factors. In: 13th International Conference on Engineering Psychology and Cognitive Ergonomics, EPCE 2016 and Held as Part of 18th International Conference on Human-Computer Interaction, HCI International, Toronto, Canada, July 17–22. LNAI, vol. 9736, pp. 116–124 (2016)
23. Wu, X.: Research on the Mechanism of Error-Cognition for Interaction Interface of Complex Information. Doctoral Dissertation of Southeast University, Nanjing (2015)
24. Ding, Y.: Ergonomics. Institute of Technology Press, Beijing (2000)
25. Wickens, C.D., Holiands, J.G.: Engineering Psychology and Man Peformance. East China Normal University Press (2003)

A Pilot Error Analysis Method for the Human-Machine Interface Design Process of Civil Aircraft Cockpit

Shasha Lu(✉), Yang Zhou, Hongtao Liu, and Zhefeng Jin

COMAC ShangHai Aircraft Design and Research Institute,
No. 5188 Jinke Road, Pudong New Area, Shanghai, China
lushasha88126@126.com

Abstract. CS25.1302 requires that the design of aircraft cockpit human-machine interface (HMI) shall have error management capability, thus human error analysis is a crucial part in the cockpit human factor design process. Focusing on potential pilot errors and observed pilot errors during human factor evaluation tests during the civil aircraft cockpit HMI design process, this paper proposed a practical pilot error analysis method which considers CREAM method as well as the characteristics of flight. This method analyzed the type of error, effects and criticality of the error on operation and safety, root cause of the error and existing safety strategy and means of risk mitigation.

Keywords: Pilot error · Safety · HMI design

1 Introduction

Human error was defined as the intended action that dose not achieve the desired effect. Errors committed by the pilots were one of the major causes for the aircraft accidents.

AMC 1302 [1] requires that the design of aircraft cockpit human-machine interface (HMI) shall have error management capability, thus human error analysis is a crucial part of cockpit human factor certification. Pilot errors derived from two sources should be analyzed according to experience and the requirement of the authority. The first source was potential errors which refers to those errors that can be reasonably expected in service from qualified and trained flight crews. The term ‘reasonably expected in service’ means those errors that have been seen in service with similar or comparable equipment or which can be projected to occur based upon general experience and knowledge of human performance capabilities and limitations related to the use of controls, information, or system logic of the type being assessed. Once a potential error is identified, it has to be demonstrated that the cockpit offers sufficient means for its management (protection, identification by the crew, recovery). The other source refers to observed human error, which means flight crew errors that are actually occurred and observed during evaluations and tests based on simulated scenarios. Similarly, regarding these observed error, it has to be demonstrated that the cockpit offers sufficient means for its management.

This document specifies a methodology of cockpit human error analysis for aircraft cockpit HMI design process.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 155–161, 2020.
https://doi.org/10.1007/978-3-030-20037-4_14

2 Pilot Error Analysis Methodology

Hollnagel [2] suggested several human error analysis methods. When it was applied in the industry, especially in aircraft cockpit HMI design and certification process, the methods have been hybrid and adopted in a practical way. After modifying for several

Table 1. Pilot error analysis template

General Information	Error ID:	
	Error name:	Number of occurrences:
	Error Type:	Pilot position:
Scenario Description	Flight Phase:	
	Event:	
	Description of the error:	
	Consequences:	
	Detection of the error:	
	Recovery of the error:	
Contributing factors	Probable Cause:	
	Current safety strategy and means of risk mitigation:	
Criticality	Criticality level:	
	Criticality rationale:	
Proposal	Proposed safety strategy and means of risk mitigation:	
	Design change recommendation:	

times, a general template has been developed to analyze the pilot errors. As it can be seen from Table 1, five aspects should be considered when analyzing the pilot errors. The sections below explain the details one by one.

2.1 General Information

This part consists of the following information:

1. Error ID. Error ID is used to number the errors. Numbering is a good way to manage and track the status of the errors.
2. Error name. Each error should have a short name which could simplify the discussing process.
3. Error type. Classifying pilot errors could enable the description the error information completed and help analyze the root causes of the errors. Pilot errors could be classified into the following types according to the analysis of several accidents as well as referring to literature [2, 3].
   (a) Omission. This type of error has two more detailed sub-types. One is that the pilot has no reaction to stimulus (alerting or abnormal behavior of the aircraft). The other is forget to do one or several steps of the procedure.
   (b) Deviation from procedure. This error refers that the pilot intentionally breaks out the standard operational procedure for some reasons, such as skip some steps or perform the actions in the wrong sequence.
   (c) Action in the wrong manner. This error refers that the pilot performs the action too fast or move the controls in the wrong direction.
   (d) Action at the wrong object. This error refers that the pilot selects the other object instead of the right one.
   (e) Incorrect data entry. This error is typical when keyboard or control display unit is used.
   (f) Action too late. This error refers that the pilot detects the abnormal and perform the action too late.
4. Pilot position. It is essential to describe the which seat (left or right) the pilot sit when making the error since the crew have different tasks during flight.

2.2 Scenario Description

This part focuses on the scenario of the error.

1. Flight phase. It should be interpreted in which flight phase the error occurred (Push back/Taxi, Take-off, Climb, Cruise, Descent, Approach, Landing, etc.).
2. Event. It refers to event/activity/task that is in progress when the error occurred, e.g., cabin depressurized and perform emergency descent procedure.
3. Consequence. Describe the possible consequences that may result from the error.

4. Detection and recovery of the error. It is important to interpret whether the error was detected and recovered after committing when analyzing observed human error, since AMC 1302 [1] requires that the HMI design should have the ability to help pilot detect and correct the errors.

2.3 Contributing Factors

Analyzing contributing factors is critical for analyzing the pilot errors. The mitigation strategies could be proposed only when the causes were found out.

Both internal and external factors could cause pilot errors based on literature review and interviews with airline pilots. Internal factors could be summarized as following aspects [4, 5]:

1. Lack of adequate knowledge. This factor is the main cause for errors of action in the wrong manner.
2. Too much workload. Several omission errors made because the abnormal situation causing too much workload which resulting in omitting the alerting.
3. Distraction. Distraction is another reason for omission. The common situation is that once distracted to another task, the remained actions for current task were omitted.
4. Fatigue. Fatigue could result in acting or reacting slowly.

As for external factors, natural environment, design of equipment and team work are highly relevant. Natural environment refers to the visibility of the outside environment, turbulence, storm, etc. Team work relates to team communication and coordination between flight crew members. Design of equipment is the most important point that should be deeply analyzed since it is directly related to AMC 1302.

The following factors related to the design of equipment were summarized through Delphi method.

1. The HMI of the equipment is too complex.
2. The information provided is not adequate for the task.
3. The feedback is not sufficient after operating the controls.
4. The intent function does not match the task.
5. Too many steps to accomplish one task.
6. The HMI could not help situation awareness.
7. Ambiguous labels.
8. The cues for next action are not sufficient.

2.4 Criticality

Identification of the criticality of pilot error could help to decide the strategies that are used for mitigating the errors.

AMC 25.1309 [6] defines and classifies the criticality of system failure condition according to the severity of its effects, this approach is also applicable to the criticality analysis of human error. The criticality of effect of human error is defined as five levels according to the severity of its effects, which are No safety effect, Minor, Major, Hazardous and Catastrophic. The rational of the criticality could compare the consequence of the errors to the definition of the criticality level [6].

2.5 Proposal

This part describes newly proposed safety strategy and means of risk mitigation regarding this particular error. The strategy could be design change, additional training, procedure modification, etc.

If design change was proposed, the recommendation related to design change should be detailed and clarify how the new design could mitigate the error.

3 Examples that Applied the Methodology

Table 2 describes and analyzes the potential human error “Not turning X FEED VLV switch to OFF after performing FUEL IMBALANCE procedure”, and is for example only.

Table 2. Potential human error analysis - Example

General Information	Error ID: A929-P-28-001	
	Error name: Not turning X FEED VLV switch to OFF after performing FUEL IMBALANCE procedure (no fuel leak)	Number of occurrences: N/A
	Error Type: Omission	Pilot position: Right seat
Description	Flight Phase: Cruise	
	Event: Fuel quantity difference in left and right wing tanks triggers FUEL IMBALANCE (Caution) alert, crew performs corresponding procedure to correct the imbalance.	
	Description of the error: Flight crew performs FUEL IMBALANCE procedure after checking that there is no fuel leak; after the imbalance has been corrected and the alert disappeared, the crew did not turn X FEED VLV switch to OFF as stated in the procedure.	
	Consequences: If there is no fuel leak, keeping X FEED VLV open will not reduce aircraft safety; only when there is fuel leak while X FEED VLV is open may lead the aircraft to a more serious criticality level.	
	Detection of the error: N/A	
	Recovery of the error: N/A	
Contributing factors	Probable Cause: Interruption by other events (e.g., ATC communication, cabin crew enters cockpit, etc.) while performing a procedure may lead to omission or skipping of one or more steps of the procedure by the crew.	
	Current safety strategy and means of risk mitigation: a) Turning X FEED VLV switch to OFF is clearly stated in the FUEL IMBALANCE procedure; b) When X FEED VLV is open, X FEED VLV switch on the overhead panel will illuminate a cyan "OPEN"; c) Fuel quantity difference in left and right wing tanks is less than 50kg, but X FEED VLV is still open, EICAS will indicate FUEL EQUAL XFEED OPEN (Advisory).	
Criticality	Criticality level: Minor	
	Criticality rationale: Current safety strategy and means of risk mitigation shall be able to make the crew realize that X FEED VLV switch is still open in minimal time and take action to handle the problem; if there is no fuel leak, keeping X FEED VLV open will not reduce aircraft safety; only when there is fuel leak while X FEED VLV is open may lead the aircraft to a more serious criticality level.	
Proposal	Proposed safety strategy and means of risk mitigation: No.	
	Design change recommendation: Add an aural alert along with EICAS indication of FUEL EQUAL XFEED OPEN (Advisory) when fuel quantity difference in left and right wing tanks is less than 50kg, but X FEED VLV is still open. (For example only)	

4 Conclusion

This paper proposed a practical methodology which could be applied to analyze the pilot errors during aircraft cockpit HMI design process. The methodology proposed were summarized by the experience of industrial engineering experience as well as literature review. It has been validated by industrial practice.

References

1. EASA: Certification Specifications For Large Aeroplanes CS25.1302. EASA, 2009/017/R, Cologne, Germany (2009)
2. Hollnagel, E.: Cognitive Reliability and Error Analysis Method. Elsevier Ltd., Amsterdam (1998)
3. Guo, B., Wang, M.: Civil Aircraft and Incident Statistics Analysis Manual. China Aviation Publishing & Media Co., Ltd., Beijing (2015)
4. Rosecoe, S.N.: Aviation Psychology. The Iowa State University Press, Iowa City (1980)
5. Chen, D., Jiao, Z.: Human Factors and Flying Safety. National Defence Industry Press, Beijing (2016)
6. EASA: Certification Specifications for Large Aeroplanes CS25 AMC. AMC25.1309, EASA, 2009/017/R, Cologne, Germany (2009)

Personal Resilience

Training for Readiness and Resilience: Supplemental Findings

Lauren Hallal-Kirk[1](✉), Debra J. Patton[2], Joan Johnston[2], Laura Milham[3], Lisa Townsend[3], and Dawn Riddle[3]

[1] Cognitive Performance Group, 3662 Avalon Park East Blvd Suite 205, Orlando, FL, USA
Lauren@cognitiveperformancegroup.com
[2] United States Army Research Laboratory, Orlando, FL, USA
{Debra.J.Patton.civ,Joan.H.Johnston.civ}@mail.mil
[3] Naval Air Warfare Center Training Systems Division, 12211 Science Drive, Orlando, FL, USA
{Laura.Milham,Lisa.Townsend,Dawn.Riddle}@navy.mil

Abstract. Conventional military training might not fully prepare Soldiers for some of the physical and psychological skills vital to reducing stress as well as avoidable causalities in combat situations. The present study operationalized a complex Integrative Training Approach (ITA) [1, 2] to address cognitive-based gaps regarding First Responder actions and decisions. This approach included three live, scenario-based exercises calibrated with scenarios meant to elicit increasing psychophysiological pressure via combat stressors embedded within key events. The associated mental and physical arousal was recorded using subjective and objective measures during each scenario. Participants in the experimental condition were provided psychological training and practical application before proceeding to the live environment. This paper serves as the continuation of the 2016 Training and Readiness Resilience study, published in 2018 [2]. With the incorporation of additional data, we found continued reliability with previously reported results; indicating that training was a mitigator of measured stress, as compared to Control squads, who received no training.

Keywords: Training · Stress · Arousal · Performance

1 Introduction

In 2018, Patton et al. [2] reported on a study conducted in 2016 in which they found that stress exposure training reduced subjective and objective stress responses in U.S. Army squad members following live training exercises, compared to squads that did

Disclaimer. This work was funded by the US Army Medical Research and Materiel Command Joint Program Command-1 (MRMC JPC1). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 165–175, 2020.
https://doi.org/10.1007/978-3-030-20037-4_15

not receive training. Stress exposure training is taught in three steps: (a) the initial stage, in which information is provided regarding stress and stress effects; (b) skills acquisition, in which specific cognitive and behavioral skills are acquired; and (c) application and practice of these skills under conditions that increasingly approximate the task environment [1, 3]. These findings and methodology were used to inform a new team training research experiment conducted in 2018 wherein Soldiers were taught specific resilience skills to help them recognize and reduce their responses to acute stressors. Regardless of the modifications to training content, the 2018 study had the same research objectives as the 2016 study, with the focus on cognitive and emotional resilience at the forefront of the study objectives. In this paper we report these findings as a follow-on to the Patton et al. study [2].

In combat, the effects of immediate and profound stress can impair a multitude of vital cognitive processes such as decision making, information processing, attention, and situation awareness [4], with catastrophic, sometimes life-threatening consequences for military members and their teams. Additionally, those who have experienced mission failure or civilian and squad member casualties have shown challenges with mental health resilience later in life [5, 6]. The 2016 study collected psychophysiological data, with the focus on assessing combat team performance related to Tactical Combat Casualty Care (TC3). The 2018 study employed a revised resilience training curriculum to focus on presenting squad members with actionable objectives to identify and mitigate a stress-related psychological event on behalf of a civilian or other team member [7–9]. A major goal of the 2018 study was to test the effectiveness of this additional training component as it related to psychological wellness on the battlefield. As in the 2016 study, the 2018 study was a quasi-experimental mixed design and employed an integrated training approach (ITA) first using classroom-based instruction and simulation-based training (SBT), and then live team training. In addition, the 2018 study sought to test the efficacy of simulated versus classroom-based practical application using additional resilience training components. The experimental group received the instructional content in the same manner; however, half the participants were directed to practice the content in a classroom environment among their peers, and the other half participated in an individually-focused computer-based curriculum. The Control group received the ITA based on the 2016 study. Both groups participated in three live scenarios. To emphasize psychological resilience, scenario events in both simulated and live training were designed to elicit increasing psychophysiological pressure via combat stressors embedded within key events.

Researchers were able to replicate the 2016 data collection method by using identical materials, methods, and artifacts throughout the 2018 experiment. Subjective measures (anxiety, depression, hostility, positive affect, sensation seeking and dysphoria) and objective measures (heart rate and respiration rate) of arousal were collected. In this paper we present findings for the stress assessments during the live scenarios to address the following hypotheses: the Experimental group would show lower levels of stress, as measured by subjective and objective arousal measures, compared to the Control group; and all participants, regardless of group, reporting higher levels of motivation would show lower levels of subjective and objective stress.

2 Methods

2.1 Participants

Volunteer affidavits were obtained from all participants in accordance with Institutional Review Board requirements, 32 CFR 219 and DoDI 3216.02. A total of 67 participants were recruited for this study made up of ten squads drawn from the 2nd Infantry Division and 6th Marine Regiment, each augmented with a 68W (Army) Medic or 8404 (Navy, attached to Marine squad) Corpsman. The first squad (Army) participated as a rehearsal squad to allow researchers the opportunity to pilot the comprehensive suite of assessments, as well as make scheduling or other adjustments based on qualitative verbal feedback received from the soldiers. No data was assessed pertaining to their tactical performance. Of the remaining nine squads, a total of six (4 Army, 2 Marine) participated in either the Experimental or Control condition. These nine squads had data collected on their psychophysiological state during the training events. Study condition assignment was randomized and resulted in four squads in the Experimental condition (n = 44) and two squads in the Control condition (n = 22).

Due to scheduling conflicts of one 68W (Army) medic, the first squad to participate under the Experimental condition was augmented with a soldier who was not a trained medic but the Soldier had experience in a First Responder capacity as a civilian and was designated by the Platoon Leader to perform the medic's role. Two squads (Army) participating within the Experimental condition were ad-hoc due to the nature of individualized recruitment. Additionally, two squads (Marine) participating within the Control and Experimental conditions had less than the maximum number of Marines (i.e., 13) that are normally assigned to an infantry squad. At the recommendation of the Marine Platoon Leader, and prior to the administration of the Informed Consent form, several members from a third squad (present, but not participating in any study-related activities) were assigned to the first and second (i.e., Experimental or Control) squads to participate in the study.

2.2 Task and Stimuli

All simulation and live training scenarios required pre-existing tactical and TC3 skills. The intensity of the combat situation increased with each scenario from a suicide bombing taking place away from the squad, to a sniper attack on the squad, to a combination of a civilian bombing and sniper attack taking place simultaneously and in the immediate vicinity of the squad.

2.3 Measures

Subjective. Paper and pencil self-report measures were collected to assess motivation and traits, and state arousal measures (see references for detailed descriptions of the individual measures). The Revised Ways of Coping Checklist (RWCCL) [10] was employed to assess the use of both adaptive (problem focusing, seeking social support) and maladaptive (avoidance, wishful thinking, blaming others) coping behaviors. Motivation was measured by asking participants to rate how important this training was to them and how willing they were to participate on a 1 to 100-point scale [11]. The Multiple Affect Adjective Checklist-Revised (MAACL-R) [12] was used to assess six validated state and trait subscales for anxiety, depression, hostility, sensation seeking, positive affect, and dysphoria. Participants were instructed to check all of the words (132 in total) that described how they "feel generally" (trait measure), how they "feel right now" (pre-measure), or "felt during a specific time" (post-measure).

Objective. Psychophysiological recordings were made with the Equivital™ EQ02 LifeMonitor which is a lightweight sensor belt worn around the participant's chest and abdomen. Electrocardiogram sensors embedded into the belt allow measures of heart rate (HR) and heart rate variability measures (inter-beat-interval; IBI), and a pressure transducer in the belt allowed a measure of respiration rate (RR).

2.4 Procedures

Data collection was conducted over a five-week period, with one Control and one Experimental condition squad participating during a single week. Both Control and Experimental squads received several sessions of training over four days (Table 1). Days 1 and 2 were conducted with classroom instruction in the morning and SBT in the afternoons. The Motivation, Trait MAACL-R and RWCCL, followed by the Baseline MAACL-R and accompanying Baseline Physiological Measures, were completed simultaneously by both groups on Days 1 and 2. During instructional periods related to psychological intervention techniques, the Control Group was asked to take an extended break while the Experimental group participated in classroom instruction, as well as a practical application of the content. No psychological or physiological data was collected during these training events.

Table 1. Experimental and control schedule of events

Time of day	Day 1	Day 2	Day 3	Day 4
Morning	Session 1 Classroom Instruction	Session 3 Classroom Instruction	Exp. Session 5 & 6 Live Exercise Scenario M1 & M2	Exp. Session 7 Live Exercise Scenario M3
Afternoon	Session 2 SBT	Session 4 SBT	Ctrl. Session 5 Live Exercise Scenario M1	Ctrl. Session 6 & 7 Live Exercise M2 & M3

On Days 3 and 4, both groups participated in three live scenarios in an outdoor urban complex. In the morning of Day 3, the Experimental squad participated in live scenarios M1 and M2. In the afternoon, the Control squad participated in M1. On the morning of Day 4, the Experimental squad participated in live scenario M3, and the Control squad participated in M2 and M3 in the afternoon. The MAACL-R was administered directly before and immediately following each live exercise scenario, and physiological measures were collected throughout all live training sessions.

2.5 Data Analysis

The three conditions were computer-based enhanced resilience training (Experimental condition, n = 21), traditional lecture-based enhanced resilience training (Experimental condition, n = 23), and standard resilience training (Control condition, n = 22). To control for the unequal sample size in each condition we opted to conduct a Monte Carlo, in which the statistical program (i.e., SPSS) randomly selects participants. An analysis of the two Experimental conditions did not yield significant differences, therefore we combined the two Experimental conditions to compare against the Control group. This left the two conditions; Experimental (n = 44) and Control (n = 22). Following the selection, we operationalized a Kruskal-Wallis analysis to gauge the normality of distribution of the Experimental and Control group to ensure a valid sample with which to conduct analyses.

The data were evaluated for missing or incomplete measures. We also evaluated for outliers which were removed from all analyses. When a data point was missing it was assigned with a missing value or the participant's data was not included in that analysis. In circumstances when the sample size of a particular group was reduced to less than 85% of the comparable group's sample size, non-parametric analysis results were considered. GLM Repeated Measures ANOVA was conducted to compare the Control and Experimental groups on the dependent variables (MAACL-R). If the GLM was not significant, data were grouped and paired t-tests were conducted to determine differences among the individual variables. For comparison of the other variables, paired t-Tests were computed. One-way ANOVAs were conducted to test differences between groups for motivation and confidence.

3 Results

3.1 Motivation

As with the 2016 study [2] motivation to participate was moderate to high in both groups. Further, both groups found the task of equal importance (Control: $M = 85.45$, $SEM = 3.97$, Experimental: $M = 82.50$, $SEM = 2.99$), [$F(1,40) = .361$, $p = .551$] and were equally willing to participate (Control: $M = 95.91$, $SEM = 6.67$, Experimental: $M = 93.33$, $SEM = 10.64$), [$F(1, 41) = .914$, $p = .345$].

3.2 Psychological Interactions

A GLM MANOVA of the Trait MAACL-R and Trait RWCCL results for both groups produced no significant differences (Wilks'$\lambda = .922$, $F = (6,36)$, $p = .798$, Wilks'$\lambda = .878$, $F = (5, 37)$, $p = .414$).

Trait Coping and Trait Affect Relationships. Examination between Trait Coping (i.e., Trait RWCCL) and Trait Affect (i.e., Trait MAACL-R) indicated individuals within the Experimental group with increased depression ($r = -.489$, $p < .05$), hostility ($r = -.678$, $p < .01$) and dysphoria ($r = -.630$, $p < .01$) were less likely to use the maladaptive behaviors such as blaming others. Further, individuals with increased dysphoria (r = .480, p < .05) were more apt to participate in wishful thinking, and those with increased avoidance were less likely to participate in maladaptive risk-taking behaviors (r = −.453, p < .05). Data for the Control group indicated a significant, positive relationship between the coping mechanisms of problem focusing (r = .443, p < .05), seeking social support (r = .456, p < .05), and wishful thinking (r = .454, p < .05) with those having positive affect.

Motivation, Coping, and Arousal. We investigated the relationship between coping, motivation, and subjective stress. Results showed participants in the Experimental group who were more willing to train prior to training events were less inclined to the maladaptive behavior of blaming others (r = −.449, p = .041). Additionally, these participants reported less hostility before M2 (r = −.453, p = .039), more hostility following M3 (r = .447, p = .042), and were more likely to report risk-taking behaviors during M1 (r = .551, p = .010). Those individuals who were more willing to train following the training events were less anxious (r = −.449, p = .041) and reported less sense of failure (r = −.473, p = .030) prior to M1, less dysphoric (r_s = .508, p = .019) prior to M2, as well as more hostile (r = .480, p = .028) and more likely to report risk-taking behaviors (r = .442, p = .045) following M3 (non-parametric results). Individuals who reported a greater importance of the training, following all training, also indicated a lower positive affect before M1 (r = −.673, p = .002), and lower anxiety following M3 (r = −.458, p = .049).

For participants in the Control group, analyses indicated individuals who were more willing to train prior to training events were less likely to report risk-taking behaviors following M1 (r = −.491, p = .020). Additionally, they reported less anxiety (r = −.489, p = .021), and a higher positive affect (r = −.436, p = .042) prior to M2. Additionally, participants who reported a greater importance of training following training events had a lower positive affect following M1 (r = −.521, p = .013).

3.3 Psychological Arousal

A GLM Repeated Measures MANOVA was conducted on the State MAACL-R variables between groups across all three live scenarios. Results were not significant between or within subjects. (Wilks'λ = .933; F(6, 36) = .430 p = .854; Wilks'λ = .336; F(30, 12) = .792 p = .710, respectively). Therefore, as with the 2016 study, the groups were combined and a within-subjects ANOVA was run, however no significant results were found.

3.4 Physiological Arousal

A GLM ANOVA was conducted on the baseline physiological measures. Results indicated no significant differences between groups; thus, unlike the 2016 study, we proceeded with analyses without the use of a covariate measure. A GLM Repeated Measures MANOVA was conducted on the physiological responses during each live scenario. Results showed no interaction for group (Wilks'λ = .898; F(6, 116) = 1.068 p = .386), but did reveal a significant interaction of scenario, session, and group (Wilks'λ = .798; F(6, 116) = 2.313 p = .038). Assessing the mean results indicated a similar finding to the 2016 study, with the finding that the Experimental condition participants experienced a significant decrease in heart rate (HR) and increase in inter-beat-interval (IBI) during M2 and M3 compared to the Control group. There was also a significantly lower respiratory rate (RR) across all scenarios for individuals within the Experimental group as compared to the Control group. (See Fig. 1).

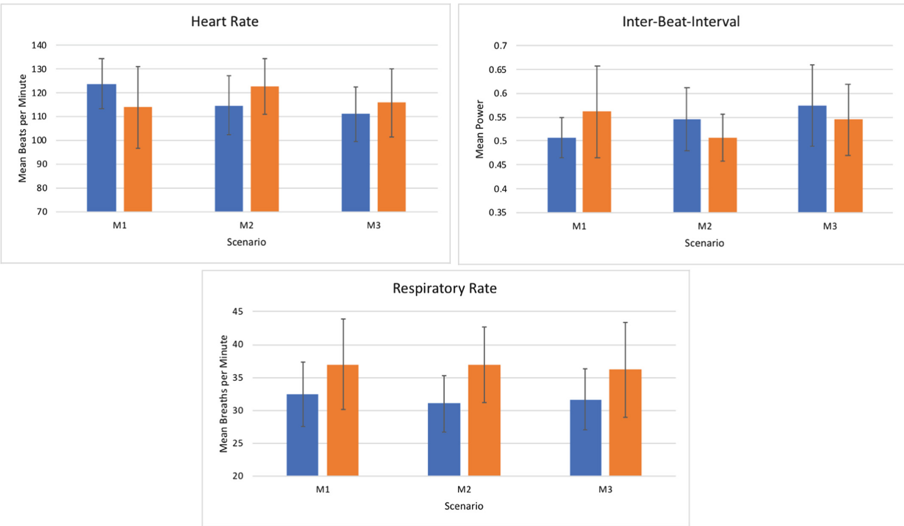


Fig. 1. Mean $\pm$ SEM for heart rate, inter-beat-interval and respiration rate by scenario (M1, M2 and M3) and condition (*Experimental: Dark, Left; Control: Light, Right*).

3.5 Psychophysiological Interaction

Pearson Correlations were conducted to investigate whether any psychological values (Trait MAACL-R) or coping mechanisms (RWCCL) that participants held coming into the training, along with psychological response values reported immediately following each scenario (State MAACL-R), had a significant relationship with an individual's physiological responses during a given live training event.

Trait MAACL-R. Results for M1 found two significant findings across both groups. A significant, positive relationship was found between IBI and Trait Positive Affect for individuals within the Experimental group (r = .561, p = .029), and a significant, negative relationship was found between RR and Trait Positive Affect for participants in the Control group (r = −.556, p = .009).

State MAACL-R. M1 results indicated participants within the Experimental group who reported higher risk-taking behaviors following the scenario experienced an increase of HR ($r = .715$, $p < .01$) and decrease in IBI ($r = -.777$, $p < .01$). These results were also found in M3 ([HR: $r = .499$, $p < .05$] [IBI: $r = -.568$, $p < .05$]); M3 data further demonstrated this negative relationship with RR ($r_s = -.461$, $p < .05$). M2 results indicated a decrease in participants' anxiety, as RR increased ($r = -.675$, $p < .05$); M3 results showed a decrease in hostility as RR increased ($r = .591$, $p < .05$).

Participants in the Control condition demonstrated an increase in dysphoria as RR decreased ($r = -.545$, $p < .05$) during M1. Additionally, participants with a lower RR reported an increase in hostility (M1; $r = -.494$, $p < .05$). During M2, data showed an increase in HR ($r = -.499$, $p < .01$) and decrease in IBI ($r = .502$, $p < .05$) with regard to lower reported anxiety.

Trait RWCCL. Results for M2 indicate a significant, negative relationship between RR and Wishful Thinking ($r = -.480$, $p < .05$) for participants in the Experimental Group. Additionally, individuals in the Control group who were more inclined to seek social support also experienced a decrease in RR ($r = -.455$, $p < .05$). The same finding was present for M3 ($r = -.439$, $p < .05$).

4 Discussion, Limitations, and Future Directions

We set out to assess psychophysiological stress experienced during live training exercises that implemented the SET paradigm. Our intent was to gauge the efficacy of psychological training modules [7] as part of the ITA [14]. Additionally, we sought to investigate the reliability of results reported in the 2016 study conducted by Patton et al. [13].

Psychological Traits and Motivation. Traits and motivation are factors that persons have and are considered stable over time [15, 16]. We investigated these variables to determine if they would mitigate or aggravate the effect of stress. Our findings indicated no significant differences between groups on these variables, as well as minimal significant relationships (one for each group) with respect to psychophysiological interactions during the first exercise, only. Thus, our focus was directed toward differences identified between groups during the live training.

Psychological. Replicating findings from the 2016 study we found that individuals in the Experimental group reported greater frustration and were more likely to engage in risk-taking behaviors during the training. Additionally, we found these individuals were less anxious during the exercises. On the contrary, the Control group participants were less likely to engage in risk-taking behaviors and reported greater negative affect throughout the training. Thus, it appears the psychological intervention may serve as the explanation for the difference in beneficial and maladaptive behaviors.

Physiological. Part of the training provided was to remain alert to breathing patterns such as when to use breathing prior to a potential known stressor or when the body is responding to a stressor. Breathing is a known factor in reducing stress due to its effect on the vagus nerve [17]. Breathing during stress affords the body the proper oxygen it needs to provide better physical and cognitive control for enhancing performance. In this study, across all three scenarios, RR was significantly lower for the Experimental group than the Control group, replicating findings in the 2016 study. Additionally, the IBI is considered a measure of the cognitive workload reflected in the ECG. In this study we found that the Experimental group demonstrated a decrease in HR, and subsequent increase in IBI from M1 to M3, indicating they were more relaxed during the increasingly stressful scenarios [18], also replicating findings in the 2016 study. These results support the training they received mitigated the effects of stress.

Psychophysiological. With regard to psychophysiological findings, participants in the Experimental group who engaged in risk-taking behaviors showed an increase in HR, subsequent decrease in IBI, and a decrease in RR in the first and last scenarios (M1, M3). Since the psychological responses reported here were collected after the event, and the physiological data was collected throughout each live event, it is possible this finding reflects that participants in the Experimental group capitalized on the breathing exercises presented during training. This technique slowed the RR and allowed the individual to engage the executive brain centers [7, 18], and thus were able to consider the consequences and/or meaning of their potential behaviors. This risk/benefit analysis then manifested in the physiological arousal reflected in our findings, during a given moment, while still supporting the overall lower heart-rate reflected in average findings for each event, respectively.

Though we found a significant, negative relationship between RR and anxiety, and hostility for the Experimental group during M2 and M3, we must consider that these two scenarios were inherently more active than M1, and likely required more physical expenditure.

The negative relationship between dysphoria and RR found in the Control group during M1 may be indicative of several individuals experiencing a negative psychological reaction (e.g., “going into the black” [7]). During such an event, individuals completely lose touch with what is occurring in the moment and this can be displayed in a variety of ways, one of which is standing still and staring into space. This could account for the decrease in RR as individuals are much less active and adopt an almost calm demeanor during such an event.

Limitations. As our experiment was part of a larger study which sought to test the efficacy of either a classroom-based or computer-based practical application of content, there were twice the number of Experimental participants than in the Control condition. Though we employed statistical methods to reduce the Experimental sample size to equal the Control condition, we recognize the omitted data may have influenced our findings had it been included. While we found possible evidence of individuals in the Control group going “into the black” during the first exercise, we did not collect data regarding an individual’s level of engagement until roughly an hour after the event due to participants moving directly into an after-action-review. This “cooling off’” period might have mitigated participant’s recollection of the severity of their psychological experiences during a given event, resulting in our inability to support the physiological findings with additional, relevant data. We did not collect any psychophysiological data from participants following a “cooling off” period to gauge how quickly in which they returned to their baseline thresholds. One aspect of study design not considered was the potential availability of psychological training and practices (i.e., mindfulness) to Soldiers and Marines outside the study; the popularity of mindfulness practices may have mitigated the outcome of results, specifically as it related to Control findings.

Conclusion. Our findings suggest the training the Experimental group received mitigated stress based on the psychophysiological changes noted throughout the live training. Additionally, our objective findings replicated those found during the 2016 study, further supporting the efficacy of the SET method and the ITA. Though we found significant differences, and those differences favored the Experimental participants, future research should include each participant's depth of knowledge of mindfulness training and practices, using such data as a covariate during the analytic process.

References

1. Johnston, J.H., Napier, S., Ross, W.A.: Adapting immersive training environments to develop squad resilience skills. In: International Conference on Augmented Cognition, pp. 616–627. Springer, Cham (2015)
2. Patton, D., Johnston, J., Gamble, K., Milham, L., Townsend, L., Riddle, D., Phillips, H.: Training for readiness and resilience. In: International Conference on Applied Human Factors and Ergonomics, pp. 292–302. Springer, Cham (2018)
3. Driskell, J.E., Salas, E., Johnston, J.H., Wollert, T.N.: Stress exposure training: an event-based approach. In: Hancock, P.A., Szalma, J.L. (eds.) Performance Under Stress, pp. 271–286. Ashgate, London (2008)
4. Stokes, A., Kite, K.: Flight Stress. Ashgate, Brookfield (1994)
5. Grieger, T.A., Cozza, S.J., Ursano, R.J., Hoge, C., Martinez, P.E., Engel, C.C., Wain, H.J.: Posttraumatic stress disorder and depression in battle-injured soldiers. Am. J. Psychiatry **163** (10), 1777–1783 (2006)
6. Hoge, C.W., Castro, C.A., Messer, S.C., McGurk, D., Cotting, D.I., Koffman, R.L.: Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. N. Engl. J. Med. **351**(1), 13–22 (2004)
7. Allard, Y.S., Start, A.R., Knust, S.K., Thompson, J., Svetlitzky, V., Adler, A.B.: iCOVER for acute stress reactions in combat: transitioning an Israeli intervention for U.S. service members. In: Military Health System Research Symposium, Kissimmee, Fl, August 2018
8. Svetlicky, V., Farchi, M.: The YAHALOM protocol: immediate, peer-based CSR intervention at the point-of-injury. Poster presented at the Meeting of the International Society of Traumatic Stress Studies, Washington, DC, November 2018
9. Farchi, M.: A training curriculum for Homefront command units for stress trauma situations. IDF Homefrong Command, Mental Health Division (Heb) (2010)
10. Vitaliano, P.P., Russo, J., Carr, J.E., Maiuro, R.D., Becker, J.: The ways of coping checklist: revision and psychometric properties. Multivar. Behav. Res. **20**(1), 3–26 (1985)
11. Fatkin, L.T., Hudgens, G.A.: Stress perceptions of soldiers participating in training at the chemical defense training facility: the mediating effects of motivation, experience, and confidence level (ARL-TR-365). U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD (1994)
12. Lubin, B., Zuckerman, M.: Manual for the MAACL-R: Multiple Affect Adjective Checklist-Revised. Educational and Industrial Testing Service, San Diego (1999)
13. Patton, D., Townsend, L., Milham, L., Johnston, J., Riddle, D., Start, A.R.: Optimizing team performance when resilience falters: an integrated training approach. In: Schmorow, D.D., Fidopiastis, C.M. (eds.) AC 2018. LNAI, vol. 10916, pp. 339–349 (2018). https://doi.org/10.1007/978-3-319-91467-1_26
14. Cobb-Clark, D.A., Schurer, S.: The stability of big-five personality traits. Econ. Lett. **115**(1), 11–15 (2012)

15. Patton, D.: Perspective taking, cultural stress, and the individual: from the inside out (No. ARL-TR-6923). U.S. Army Research Laboratory, Human Research and Engineering Directorate, Aberdeen Proving Ground, MD (2014)
16. Gerritsen, R., Band, G.: Breath of life: the respiratory vagal stimulation model of contemplative activity. Front. Hum. Neurosci. **12**, 397 (2018). https://doi.org/10.3389/fnhum.2018.00397
17. Morgan III, C.A., Aikins, D.E., Steffian, G., Coric, V., Southwick, S.: Relation between cardiac vagal tone and performance in male military personnel exposed to high stress: three prospective studies. Psychophysiology **44**(1), 120–127 (2007). https://doi.org/10.1111/j.1469-8986.2006.00475.x
18. Lombroso, P.J., Arnsten, A.F.: Development of the cerebral cortex: XIV. Stress impairs prefrontal cortical function. J. Am. Acad. Child Adolesc. Psychiatry **38**(2), 220–222 (1999)

Physical Readiness is More Than Physical Fitness: Relationships Between Army Physical Fitness Test Scores and Self-reports of Physical and Psychological Fitness

Gary L. Boykin Sr.(✉) and Valerie J. Rice

U.S Army Research Laboratory, RSQ Branch,
Army Medical Department Field Office, 3161 McIndoe Road,
Fort Sam Houston (JBSA), San Antonio, TX 78234, USA
{gary.l.boykin.civ,valerie.j.rice.civ}@mail.mil

Abstract. Physical Readiness (PR) is essential to the U.S. Army's ground fighting lethality and essential for combat medics trained to evacuate the sick and wounded from the battlefield. However, physical readiness, as measured by the Army Physical Fitness Test (APFT), may involve more than physical capabilities. The purpose of this study was to examine the relationships between APFT scores and self-reports of physical fitness, self-esteem, self-concept, and positive and negative self-talk among soldiers attending combat medic training (n = 473). Pearson Correlations with a p-value of .05, showed low to moderate associations between APFT scores and self-reports in each category of consideration ($p < .05$). These results suggest that physical readiness, as determined by APFT scores, is about more than physical performance, and previous research demonstrating the highest physical readiness is achieved among those with high physical, cognitive, and emotional performance. Findings also suggest the need for mental fitness training offered alongside physical training to build strong, resilient soldiers.

Keywords: Training · Mental · Performance · Conditioning · U.S. Military Body Systems · Performance · Conditioning · Exercise

1 Introduction

Physical Readiness (PR) for military troops is described as, "…the ability to meet the physical demands of any combat or duty position, accomplish the mission, and continue to fight and win", and "Soldier physical readiness is acquired through the challenge of a precise, progressive, and integrated physical training program" [1]. As such, physical readiness is the responsibility of military leadership and the individual Soldier [2–4]. Maintaining optimal physical and mental wellbeing is vital in warfighting or peacekeeping deployments [5], and especially during continuous operations missions [4, 6]. Currently, to measure Soldiers' physical strength and endurance, the Army conducts a semiannual standardized Army Physical Fitness Test (APFT) consisting of three timed physically demanding performance measures – push-ups, sit-ups and a

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 176–183, 2020.
https://doi.org/10.1007/978-3-030-20037-4_16

two-mile run [2]. Successful completion of each performance task is necessary to receive an overall passing score, and passing score benchmarks are established based on gender and age [2]. Soldiers who fail the APFT are offered remedial physical training as an opportunity to build their physical skills and retake (and potentially pass) the test. This remedial support increases training costs for the expended man-hours of both the trainer and the individual soldier. Furthermore, after repeated APFT failures, Soldiers are subject for release from the Army [2], which may diminish the Army's core fighting and conservation strength.

There is no doubt Soldiers' must develop necessary muscular strength and physical endurance to successfully complete the APFT [7]. However, other, non-physical factors may contribute to a soldiers APFT scores. Self-esteem, self-efficacy, and personal self-talk may also influence physical performance outcome [8–10]. Research suggest that positive self-esteem serve as a positive motivator during sports activity – thus leading to better sports performance [11]. Furthermore, the relationship between higher self-esteem and perfectionism is prevalent among athletes [12]. When examining the role of self-efficacy in the realm of sports and physical training, many researchers refer to the work of Bandura [13]. Bandura suggests individuals learn from past events–whether positive or negative [14]. These events help to develop greater confidence relating to a future event [14]. Therefore, soldiers that have successfully completed the APFT can predict future APFT success. However, those that have failed the APFT can learn from their shortcomings and build solutions to improve their physical and emotional states. It is believed that what a person thinks can influence their actions, and is one of the foundations for cognitive behavioral therapies [15]. Self-talk training strategies have been developed to facilitate physical performance among athletes. Studies reveal that self-talk is more effective with novel (rather than well-learned) tasks and fine (rather than gross) motor performance, and athletic training that includes self-talk is more effective than training that does not for facilitating learning and improving sports performance [16].

According to crisis theory, individual performance (physical and non-physical) relies on one's ability to assess and respond accordingly to stressful conditions [17]. Although the APFT is a physical performance test, the principles and theories associated with test anxiety are also relevant to Soldiers pending an APFT, as the consequences of failing the fitness test can be high. Consequences can include remedial physical training, not attaining a promotion, inability to attend training programs necessary for building their professional skills sets (and attaining promotion points, and ultimately the inability to remain on active duty. These potential consequences may heighten levels of stress and anxiety [8, 17–19].

Current AFPT standards [20] have been modified into a new test of six demanding performance test measures [21]. The Army's new fitness test is more physically challenging. Therefore, it is prudent to gain an understanding of personal and physical factors associated with physical readiness [22].

The purpose of this study was to determine whether self-reports of fitness, self-concept, positive and negative self-talk, and tiredness were associated with APFT scores, among soldiers attending combat medic training. We hypothesized that both physical and psychosocial factors would be significantly correlated with overall APFT scores. The new Army Combat Fitness Test will be introduced in the fiscal year 2020

[21], with the purpose of improving soldier Readiness for combat scenarios. The results of this study may contribute information on training that will help Soldiers pass this new assessment.

2 Methods

Procedures: Soldiers attending advanced individual training in the Military Occupational Specialty (MOS) 68W (Army Combat Medic) (n = 473) participated in the study. Participants completed an informed consent document before taking part in the study. The medic trainees provide self-report responses about their background, past life events, physical activity, personality traits, coping strategies, affect, and beliefs before joining the military and some questions asked about activities after joining the military. In addition, 68W course administrators' supplied current APFT training scores of all trainees involved in the study.

2.1 Self-report Survey and Questions (See Table 1)

Table 1. Self-report survey and questions.

Survey	Survey summary
Demographics	Demographics included general information on age, race, gender, marital status, and education level
"Fitness" Self-Ratings	Participants rated their physical activity levels prior to active duty and currently; endurance, speed, strength, and flexibility ratings; overall health rating; and feelings of being tired, each with a 5-pt. Likert scale rating. They also listed their typical hours of sleep/night during weekdays and weekends
Rosenberg Self Esteem Scale (SES)	The SES is a 10-item self-report measure of global self-esteem with statements rated on a 4-pt. Likert scale from strongly agree to strongly disagree - and situational self-efficacy, and self-ratings of satisfaction with their own performance and their leadership ability [23]
Self-Concept	Four self-concept questions were included. The first three addressed their 68W MOS training, the first was a situational self-efficacy question asking them to rate their confidence in their ability to complete their training on a 10-pt. scale anchored with 1 = not at all confident and 10 = extremely confident. The second item asked them to rate whether they were satisfied with their performance in their training thus far, and the third asked if their work performance was consistently the best they could do, both rated on an anchored 5-pt Likert scale from 1 = not at all true and 5 = very true. The final question asked them to rate their personal leadership ability on a 5-pt. Likert Scale
Self-Talk Survey (STS)	The STS has 17 items rated on a 5-pt. Likert scale. Adapted for this study, the Cronbach alphas were >.75, an example statement is "I think often about personal issues which interfere now with my performance." The survey provides scores for both positive and negative self-talk [24]

Army Physical Fitness Test (APFT): The APFT consists of scores for the number of sit-ups, push-ups completed independently within two minutes for each event, and the time it takes to complete a 2-mile run. The maximum score for a single event is 100 points. These scores can be consolidated into one overall score for a total up to 300 points, however, a failure (inability to meet the Army's physical performance standard based on age) in any one of the three events will consist a complete APFT failure. Subsequently, the APFT must be retaken at a later period.

Statistics: Stata statistical software [25] was used for all analyses. Pearson's correlations were used to identify covariances of variables with alpha level of 0.05.

3 Results

Table 2 shows participants' demographics. Table 3 shows the correlations between demographics and APFT scores. Higher APFT scores were seen among men, men tended to be older than women in this population, and those who were married had higher levels of education. None of the other demographic measures were significantly associated with APFT scores.

Table 2. Participant demographics

Total n = 473*	N (%)
Gender	
Male	286 (60.4)
Female	186 (39.3)
Age	
17–18	191 (40.3)
19–20	97 (20.5)
21–24	68 (14.3)
25–29	41 (.08)
30 or over	21 (.04)
Ethnicity	
African American	33 (.06)
Caucasian	330 (69.7)
Hispanic	62 (13.1)
Asian	16 (.03)
Other	29 (.06)
Marital status	
Married	59 (12.4)
Divorced	14 (.02)
Partnered	30 (.06)
Single	367 (77.5)

Table 3. Demographic correlations r (significance).

	Age	Race	Gender	Marital status	Education	APFT
Age					.	
Race	−.02 (.65)					
Gender	−.14 (.00)	.03 (.55)				
Marital status	−.46 (.00)	.00 (.92)	−.02 (.62)			
Education	.46 (00)	.08 (.04)	−.05 (.21)	−.19 (.00)		
APFT	.04 (.38)	−.02 (.65)	−.13 (.01)*	−.04 (.41)	−.06 (.17)	

*Some participants did not respond, therefore, numbers presented throughout this table may not add up precisely to the totals provided.

Physically, those who reported being more physically active prior to entering active duty service had higher APFT Scores ($r(471) = .37$, $p < .00$), as did those who were exercising in addition to their company-led physical training ($r(471) = .21$, $p < .00$). Those who rated their overall health higher also had higher APFT scores ($r(467) = .31$, $p < .00$). Soldiers whose self-ratings of their own endurance ($r(471) = .51$, $p < .00$), sprint speed ($r(471) = .40$, $p < .00$), strength ($r(471) = .33$, $p < .00$), and flexibility ($r(471) = .10$, $p < .02$), scored higher on their APFT. Feeling tired was not significantly correlated with APFT scores ($r(471) = .00$, $p = .97$).

Self-concept scores showed positive correlations between APFT scores and scores on the Rosenberg self-esteem scale ($r(473) = .10$, $p = .04$) and the situational self-efficacy scale ($r(473) = .09$, $p = .04$). Soldiers whose satisfaction with their own performance during their current Combat Medic training and rated their of their current work performance as high also had higher APFT scores ($r(471) = .10$, $p = .03$) and ($r(473) = .12$, $p = .01$), respectively. Finally, those who rated their personal leadership ability as high had higher APFT scores ($r(471) = .18$, $p < .00$).

Positive self-talk with positively correlated with APFT scores ($r(471) = .09$, $p = .04$), while negative self-talk was negatively correlated, but not significantly related with APFT scores ($r(471) = .09$, $p = .05$).

4 Discussion

The purpose of this study was to examine the relationship between self-reports of fitness, self-concept, positive and negative self-talk, and tiredness, were associated with APFT scores, among Soldiers attending combat medic training. Our hypothesis was that both physical and psychosocial factors would be significantly correlated with overall APFT scores.

Our hypothesis was supported in the findings that self-ratings of fitness were positively correlated with physical readiness, as measured by the APFT test, with associations being strongest for self-ratings of endurance, sprint speed, and activity level prior to active duty accounting for 26% 16%, and 14% of the variance respectively. The hypothesis was further supported in that ratings of self-esteem, self-concept, and positive and negative self-talk were significantly associated with APFT scores.

However, although these findings were statistically significant, the strengths of the associations were extremely small. The self-rating of personal leadership ability was the highest, accounting for only 3% of the variance. Physical exercise, past and present, and self-ratings of ones' own fitness and health are more closely associated with physical readiness; than self-esteem, self-concept, and self-talk, however, all are involved to some degree.

Soldier's responsibility and effort will be more impactful to passing the APFT [7]. A soldier that understand their specific physical weaknesses from a brutally honest perspective, for example, the inability to complete a timed two-mile run event – can be targeted with additional physical training (PT) external to military's PT program. Extra physical training that provides rigorous and task-oriented objectives could support APFT success [26]. Additionally, research suggests that anxiety hinders peak athletic performance in competitive sports activity [27]. Although the APFT is an individual performance test, scoring high could promote competitiveness, and competitiveness could lead to higher levels of anxiety among soldiers during physical testing. Soldiers that are required to receive a passing APFT score in order to achieve a next level accomplishment, e.g. a promotion is rank – could face higher levels of stress and anxiety during the APFT. However, soldiers that can gauge their emotional factors concerning their physical abilities with the APFT can begin to develop a self-accepting attitude and mental toughness [17, 28] thus improving upon their self-esteem and self-efficacy [13, 14]. When combined, soldiers can aggressively demote perceived negative responses to their immediate physical limitation and emotional state and begin building confidence.

This research also generally supports findings by Hatzigeorgiadis and colleagues [16], showing that positive self-talk was associated with better physical performance. These improvements, although small in this case, may be influenced by improvements in attentional focus, increases in confidence, enhancements in attention and regulating efforts, increases in control over cognitive and emotional reactions, and the initiation of programmed action, as may occur from positive self-talk [29]. Soldiers attending training could undoubtedly benefit from each of these enhancements, in academic, as well as physical performance. In addition, improvements in confidence and self-control might lessen sports-related test anxiety and fear of failure. When physical and emotional factors are considered at the individual level, there could be a huge benefit as APFT failures and additional retraining cost could substantially decrease. However, sequential APFT failures will do more harm than good, as Soldier could be expelled from the military and returned to a civilian lifestyle [2].

Military leaders and physical trainers may not contemplate the links between Army physical readiness as being associated with other personal factors. Understandably, as a responsibility and military standard, there is a need for Soldiers to be physically fit at all times. Physical readiness is a global requisite for military engagements whether for war or peacekeeping [30]. However, accomplishing physical readiness is about more than progressive exercise, as beliefs in oneself in both physical and non-physical realms were also related to physical fitness. These findings support theories and prior research showing the highest physical readiness is achieved among those with high functioning physically, cognitively, and emotionally [3, 7, 30–32], thus advocating the need for mental fitness training offered alongside physical training [33]. Having one, without the other, is unlikely to achieve the same level of physical readiness. Commanders must not only remain cognizant of soldiers' fitness before, during and after deployments [5], they must also maintain awareness of their soldiers' mental and emotional status, for both will impact physical readiness. These findings suggest the need for a holistic, comprehensive military fitness program. Similar questions to those used in this study might be useful in the Global Assessment Tool survey. Adopting a combat mind and body fitness training and testing routine is likely to give U.S. Soldiers the most significant advantage when faced with the challenges of war.

5 Conclusion

These findings support prior research showing that the highest physical readiness is achieved among those with high body and mind attributes, advocating the need for mental training offered alongside physical training. Having one, without the other, is unlikely to achieve the same level of physical readiness. We suggest further research is warranted given the finding within this report.

References

1. Army Public Health Center: Army Physical Fitness (2017). https://phc.amedd.army.mil/topics/healthyliving/al/Pages/ArmyPhysicalFitness.aspx
2. Department of the Army: Physical Fitness Training: FM 21-20 (1999)
3. East, W.: A historical review and analysis of army physical readiness training and assessment. Combat Studies Institute, U.S. Army Combined Arms Center Fort Leavenworth, Kansas (2013)
4. Destadio, F.J.: Peacetime physical fitness and its effects on combat readiness: an air force perspective. Individual Study Project. U.S. Army War College, Carlisle Barracks, PA (1991)
5. Lester, M.E., Knapik, J.J., Catrambone, D., Antczac, A., Sharp, M.A., Burrnel, L., Darakjy, S.: Effects of a 13-month deployment to Iraq on physical fitness and body composition. Mil. Med. **175**, 6 (2010)
6. Kruegerm, G.P., Balkin, T.J., Belenky, G.L., Headley, D.B., Sollick, R.E.: Effects of continuous operations (CONOPS) on soldier and unit performance. U.S. Army Research Institute for the Behavioral and Social Science, Walter Reed Army Institute of Research, Washington, D.C. (1986)
7. Friedl, K.E., Knapik, J.J., Hakkinen, K., Baumgartner, N., Groeller, H., Taylor, N., Duarte, A., Kyrolainen, H., Jones, B., Kraemer, W., Nindl, B.: Perspectives on aerobic and strength influences on military physical readiness: report of an international military physiology roundtable (2017)
8. Rathschlag, M., Memmert, D.: The influence of self-generated emotions on physical performance: an investigation of happiness, anger, anxiety, and sadness. J. Sport Exerc. Psychol. **35**, 197–210 (2013)
9. Beal, S.A.: The roles of perseverance, cognitive ability, and physical fitness in U.S. Army special forces assessment and selection (2010)
10. Tenenbaum, G., Edmonds, W.A., Eccles, D.W.: Emotions, coping strategies, and performance: a conceptual framework for defining affect-related performance zones. Mil. Psychol. S11-S37 (2008)

11. Arun, M.N.: Impact of self-esteem on sport performance - a comparative study. Int. J. Interdisp. Stud. **1**(1) (2016)
12. Koivula, N., Hassmen, P., Fallby, J.: Self-esteem and perfectionism in elite athletes: effect on competiveness anxiety and self-confidence. Pers. Ind. Diff. **32**(5), 865–875 (2002)
13. Mortiz, E., Feltz, D., Fahrbach, R., Mack, D.: The relation of self-efficacy measures to sports performance: a meta-analytic review. Res. Q. Exerc. Sport **71**(3), 280–294 (2000)
14. Bandura, A.: Self-efficacy: toward a unifying theory of behavioral change. Psychol. Rev. **84** (20), 191–215 (1977)
15. Meichenbaum, D.H.: Cognitive Behavior Modification: An Integrative Approach. Plenum, New York (1977)
16. Hatzigeorgiadis, A., Zourbanos, N., Galanis, E., Theodorakis, Y.: Self-talk and sports performance: a meta-analysis. Perspect. Psychol. Sci. **6**(4), 348–356 (2011)
17. Bar-Eli, M., Tenenbaum, G.: A theory of individual psychological crisis in competitive sport. Appl. Psychol.: Int. Rev. **38**, 107–120 (1989)
18. Spielberger, C.D., Vagg, P.R.: Test anxiety Theory Assessment and Treatment. Taylor and Francis, Washington, D.C. (1995)
19. Deffenbacher, J.L.: Worry, emotionality, and task-generated interference in test anxiety: an empirical test of attentional theory. J. Educ. Psychol. **70**(2), 248–254 (1978)
20. Department of the Army: Army Physical Readiness Training: TC3–22-20 (2010)
21. Department of the Army: The Army Combat Fitness Test. Center for Lessons Learned. Fort Leavenworth, KS (2018)
22. Seligman, M.E.P., Fowler, R.D.: Comprehensive soldier fitness and the future of psychology. Am. Psychol. **66**(1), 82–86 (2011)
23. Rosenberg, M.: Society and the Adolescent Self-image, Revised edn. Wesleyan University Press, Middletown (1989)
24. Banderet, L., Rice, V., Allison, S., Creedon, J., Sharp, M.: Medical indicators of successful United States military-specialty training. In: Proceedings, International Military Testing Association, pp. 190–199 (2004)
25. StataCorp.: Stata Statistical Software: Release 9. StataCorp LP, College Station (2005)
26. Haddock, C.K., Poston, W.S.C., Heinrich, K.M., Jahnke, S.A., Jitnarin, N.: The benefits of high-intensity functional training fitness programs for military personnel. Mil. Med. **181**(11), e1508–e1514 (2016)
27. Kumar, V., Singh, A., Sandhu, J., Gupta, N., Pandey, R.M.: Comparative study of sports competition anxiety among contact and non-contact sports persons, **2**(2). 77–79 (2017)
28. Zeiger, J.S., Zeiger, R.S.: Mental toughness latent profiles in endurance athletes. PLoS ONE **13**(2), 1–15 (2018)
29. Theodorakis, Y., Hatzigeorgiadis, A., Chroni, S.: Self-talk: it works, but how? Development and preliminary validation of the functions of self-talk questionnaire. Meas. Phys. Educ. Exerc. Sci. 12, 10–30 (2008)
30. LoRusso, N.: Maintaining the standard: why physical fitness will always matter in war. Modern War Institute at West Point (2017)
31. Conway, T.L., Trent, L.K., Conway, S.W.: Physical readiness and lifestyle habits among U. S. Navy personnel during 1986, 1987, and 1988. Naval Health Research Center, San Diego, CA (1989)
32. Folkins, C.H., Aime, W.E.: Physical fitness training and mental health. Am. Psychol. **36**(4), 373–389 (1981)
33. Ramnath, U., Rauch, L., Lambert, E.V., Kolbe-Alexander, T.L.: The relationship between functional status, physical fitness and cognitive performance in physically active older adults: a pilot study (2018)

Which Mindfulness Meditation Delivery Method Yields the Greatest Benefits?

Valerie J. Rice[1(✉)] and Cory Overby[2]

[1] United States Army Research Laboratory HRED FSTD RSQ Army Medical Department Field Office, Fort Sam Houston, San Antonio, TX 78234, USA
valerie.j.rice.civ@mail.mil
[2] DCS Corporation, 6909 Metro Park Drive, Suite 500, Alexandria, VA 22310, USA
coverby@dcscorp.com

Abstract. The purpose of this paper is to compare three mindfulness training delivery methods: 8 weekly classes delivered in-person (8-IP), 5 days of classes delivered in-person, 8 weekly classes delivered in a virtual world, and an 8 week wait-list control group. U.S. military service member and veteran volunteers completed pre- and post-assessments of self-reported anxiety, stress, positive and negative affect, and the Post Traumatic Stress Disorder (PTSD) Checklist – Military Version. Within group pre- to post-reductions in all measures occurred in the 8-IP group, with reductions in PTSD symptoms, anxiety, and negative affect reaching significance ($p < .05$). One group difference was seen - negative affect was reduced more in the 8-IP than in the other groups ($p < .05$). The results demonstrate that the delivery method has an impact on outcome. Future research should continue to address the effectiveness and feasibility of varied mindfulness programs.

Keywords: MBSR · Mindfulness · Resilience · Military

1 Introduction

The nature of military operations often carry with them a heavy psychological burden manifesting in the form of stress, anxiety, depression and/or other behavioral health consequences [1]. The ability of personnel to weather the challenges that military operations present is of interest to the military [2]. Thus, research has focused on personal resilience [3].

One definition of resilience is, "…the process of effectively negotiating, adapting to, or managing significant sources of stress or trauma" [4]. However, simply managing sources of trauma may not be sufficient, as a resilient individual will also be "… continuing to work, interacting with and maintaining relationships with friends and relatives, and remaining interested and involved in leisure pursuits" [5]. In other words, they must continue to function successfully, even while managing their reactions to stressors and trauma. While there are many assessments to measure resilience itself [4], and characteristics of resilience have been identified in the civilian sector [5] and among and with military and veterans [6], it is often through correlational research. For

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 184–195, 2020.
https://doi.org/10.1007/978-3-030-20037-4_17

example, research has found resilience to be positively associated with coping strategies [6, 7], optimism [8] and positive emotions [9] and negatively associated with anxiety and depression [10], perceived stress [11], negative affect [12], and PTSD symptoms [13]. There is also some evidence that the underlying characteristics and processes that contribute to resilience may change across the lifespan [14] indicating that it is not a static trait [5]. To that end, research has focused on methods to bolster resilience, and the characteristics associated with resilience, in areas such as public health [15], nursing [16], and the military application [17–19]. One potential method for building resilience is mindfulness-based stress reduction [20, 21].

1.1 Mindfulness-Based Stress Reduction (MBSR)

MBSR is an adaptive, eight-week mind/body intervention developed by Dr. Kabat-Zinn at the University of Massachusetts Medical [22]. Research into the utility of MBSR has found it useful for reducing anxiety [23], and to have a greater impact on reducing social anxiety than aerobic exercise [24]. Attending MBSR training also lessens perceived stress [25, 26], aids in emotion regulation [27], and decreases symptoms associated with PTSD [28].

MBSR training classes consist of either (a) an 8-week program with weekly classes of 2.5 h duration, and an all-day silent retreat between weeks six and seven or (b) the same classes (same content) offered over 4 ½ days [29]. MBSR training includes a prescribed set of topics and practices, and instructors complete training to qualify them to teach MBSR. As the popularity of mindfulness meditation has grown, training variations have emerged, with differing curriculums and timeframes have evolved. The differences in the interventions make research comparisons difficult, if not impossible [30]. Kabat-Zinn [31] has shared that part of his original intent for the MBSR program was for it serve as a 'vehicle' for individual self-change. No published research has compared the pre/post outcomes ('self-change') of the two MBSR delivery methods or compared MBSR with other mindfulness training programs.

This lack of information on whether the different configurations have an impact upon the program's effectiveness is of considerable interest. Organizations and individuals may prefer engaging in the most effective training, given their available time and circumstances. While two reviews have looked into the length of various mindfulness training programs [32, 33], none have compared the two methods of teaching MBSR with each other for their impact on anxiety, stress, affect, or PTSD symptoms, all of which have been shown to be positively impacted by MBSR training [34].

Our laboratory conducted a series of studies to examine the efficacy of different mindfulness training delivery systems including: traditional 8-week in-person MBSR training; 5-day in-person MBSR training, and an 8-week virtual world training program based on MBSR, with a control group [35].

Virtual worlds are a relatively recent but useful platform for training and education [36] and ehealth [37]. They allow for the engagement of geographically or socially isolated individuals through the use of virtual humans (avatars) in a co-located online world [38]. VWs have a wide range of uses from helping those with developmental disabilities [39] to medical education [40], and to military training [41]. Should such virtual world training over a VW be effective, it could be useful for reaching active duty service members stationed where such in-person training is unavailable and to veterans living in remote areas without access to in-person MBSR training.

The purpose of this paper is to share research results comparing the impact of these three delivery methods with a control group, on outcomes measures of stress, anxiety, affect, and PTSD symptoms, and to determine which method shows the greatest benefits for participants.

2 Methods

2.1 Participants

The participant sample consisted of 60 active duty military, reserve, and veteran volunteers. Participants enrolled as part of two larger studies. The first study examined (a) the traditional eight-week MBSR program delivered in person (8-IP) and (b) a mindfulness program based on MBSR (same content), but with shorter class times (1.5 h with a 4 h silent retreat) conducted over a virtual world (8-VW), and (c) a non-training wait-list control group (Control). The second study investigated d) the shorter, more intensive MBSR program, known as the Five-Day program (4 ½ days), conducted in-person (5-IP).

2.2 Instruments

Demographic data included age, gender, veteran status, marital status, and time on active duty service. Four outcome measures were used.

Zung Self-Rating Anxiety Scale (SAS). The SAS is 20-question survey used to measure the level of anxiety an individual is currently experiencing. It uses a Likert Scale rating from 1 to 4, with 1 = none or a little of the time and 4 = most or all of the time [42].

Perceived Stress Scale (PSS). The PSS is a 10-item instrument that uses a four-point Likert response scale (0 = never, 1 = Almost Always, 2 = Sometimes, 3 = Fairly Often 4 = Very Often) to quantify stress levels [43].

Positive and Negative Affect Schedule (PANAS). The PANAS consists of two 10-item mood scales to measure two distinct dimensions of mood –positive and negative affect. Positive affect (PA) reflects the extent to which a person feels enthusiastic, active, and alert, whereas negative affect (NA) is a general dimension of subjective distress and unpleasurable engagement [44].

PTSD Checklist Military Version (PCL-M). The PCL-M is a 17 item, interval-level rating scale used to screen for PTSD in military groups. The scale was derived from the 17 DSM-IV symptoms of PTSD [45].

2.3 Procedure

Participants in the two studies signed an informed consent document, completed a pre-intervention assessment, were given their group assignment, participated in the group to which they were assigned, and completed a post-intervention assessment. The pre and post assessments included the same four outcome measures previously listed. The group assignments were as previously described (8-IP, 8-VW, 5-IP, and Control). The curriculum for each of the mindfulness classes were the same, however the duration of the classes differed (described above). Homework assignments were included in the 8-wk classes.

2.4 Statistics

Only 15 research volunteers completed the 5-IP class. Therefore, 15 participants from each of the other groups were selected for comparison, matched as closely as possible on age, gender, military status (active duty vs. veteran), and time spent on active duty.

The normality of the distribution of scores were examined as an assumption check for Dependent Sample T-Tests using the Shapiro-Wilk test. If the assumption of normality was violated, a non-parametric alternative was utilized in the form of the Wilcoxon Signed-Rank Test. Change scores were calculated for the dependent variables and examined using an Analysis of Variance to assess differences been groups (therefore group is the area of interest. Where baseline normalized differences were large (>1.0), a Kruskal-Wallis test was employed for change score comparisons.

3 Results

3.1 Demographics

Participants had a mean age of 41 (±15) years and were comprised of mostly females (n = 41, 68%) and were approximately equivalent in terms of military status (veterans = 32, 53%; active duty = 26, 43%; reserve = 2, 3%). Most participants were married (n = 25, 42%) or single (n = 16, 11%), with an average of 13 years of military service.

3.2 Baseline Scores

As detailed in Table 1, participants showed average to elevated anxiety at baseline, per established cut off scores [42]. Participants' perceived stress was slightly higher than adult normative data for 2009 in the U.S. [46]. Affect scores were within normal range for positive affect, but nearly double the average for negative affect [47]. Finally, the mean results of the participant group on the PCL-M were below the cut off score for indication of trauma related distress [48].

Instrument scores by group provide more granularity (Table 2). For anxiety scores, the 8-IP and 8-VW groups showed minimal to moderate anxiety as opposed to the Control and 5-IP groups whose average scores were within normal range. Perceived stress for the 5-IP, 8-IP week and 8-VW groups were above the mean normative scores. Affect scores for all the groups were close to normative means for positive affect. However, negative affect was well above normative means for all groups. For PTSD, only the 8-IP group had a mean score indicative of some trauma related distress. Baseline scores were examined for balancing using the normalized differences method [49]. Problematic differences (Δ_{ct} > 1.0) were found for perceived stress necessitating a non-parametric test.

Table 1. Descriptive statistics for outcome measures at baseline.

	Descriptive statistics			
	Minimum	Maximum	Mean	Std. deviation
Anxiety	29	65	46.17	9.052
Perceived stress	2	34	19.59	7.582
Positive affect	17	49	31.83	3.927
Negative affect	10	42	23.43	3.943
PTSD symptomology	17	83	32.44	17.195

Table 2. Descriptive statistics by MSBR group

Group		Descriptive Statistics			
		Minimum	Maximum	Mean	Std. deviation
Control	Anxiety	29	60	44.93	9.81
	Perceived stress	6	34	16.00	7.39
	Positive affect	17	49	33.67	10.14
	Negative affect	10	42	20.67	10.27
	PTSD symptomology	17	54	28.57	12.14
FD	Anxiety	30	56	43.73	6.02
	Perceived stress	16	27	22.00	3.72
	Positive affect	19	43	31.33	7.57
	Negative affect	12	42	23.27	9.35
	PTSD symptomology	17	83	32.83	20.02
In person	Anxiety	34	66	49.07	10.75
	Perceived stress	8	34	21.53	8.35
	Positive affect	18	45	29.33	8.90
	Negative affect	13	41	26.27	8.45
	PTSD symptomology	17	80	39.21	19.30
Virtual world	Anxiety	35	66	46.93	8.93
	Perceived stress	2	34	19.00	8.77
	Positive affect	18	49	33.20	9.17
	Negative affect	11	37	23.73	7.47
	PTSD symptomology	17	71	28.67	16.19

3.3 Zung Anxiety Index

Pre to post change scores, examined by group (within subjects) were significant for the 8-IP $t(14) = 2.22$ $p = .03$ and 8-VW groups $t(14) = 2.36$, $p = .04$. An ANOVA found the difference scores across groups approached significance, $F(3, 56) = 2.74$, $p = .05$, $\eta p = .121$. Compared to the other groups, the 8-IP group showed the largest reduction in anxiety, pre to post training.

3.4 Perceived Stress Scale

A Wilcoxon test found the pre to post change in perceived stress scores was not statistically significant for any group. A Kruskal-Wallis test did not find a statistically significant difference between groups $\chi^2 = 2.15$, $p = .14$. However, the 8-IP group showed the largest decrease in perceived stress compared to other groups and the Control showed a slight increase.

3.5 PANAS: Positive

A paired Student's t-test found a significant pre to post decrease in positive affect for the Control group $t(14) = 2.18$, $p = .04$, but no significant changes in the other groups. An ANOVA did not find a significant difference between groups, $F(3, 56) = 2.31$, $p = .086$, $\eta p = .11$. Despite the lack of statistical significance, all the groups except the 8-VW group increased pre to post positive affect scores.

3.6 PANAS: Negative

A paired Student's t-test found a significant pre to post decrease in negative affect for the 8-IP group $t(14) = 2.89$, $p = .01$. An ANOVA found a statistically significant difference between groups, $F(3, 56) = .48$, $p = .03$, $\eta p = .03$. All the groups showed a decrease in negative affect scores, with the 8-IP group showing the largest decrease.

3.7 PTSD Checklist – Military Version

A Wilcoxon test found a significant pre to post change for the 8-IP group only ($Z = -2.97$, $p = .00$). An ANOVA did not find a statistically significant between subjects difference for group, $F(3, 55) = 1.38$, $p = .26$, $\eta p = .07$. While the group differences were not statistically significant, the 8-IP group had the largest decrease in PTSD symptomology, followed by the 5D group, with the 8-VW group showing a very slight increase in symptoms, and no change in the Control group.

3.8 Summary of Results

The within subjects results were compiled for a side-by-side comparison (Fig. 1). Between group comparisons revealed significantly greater decreases in negative affect among the 8-IP group compared with the other three groups. No other significant between group differences were noted.

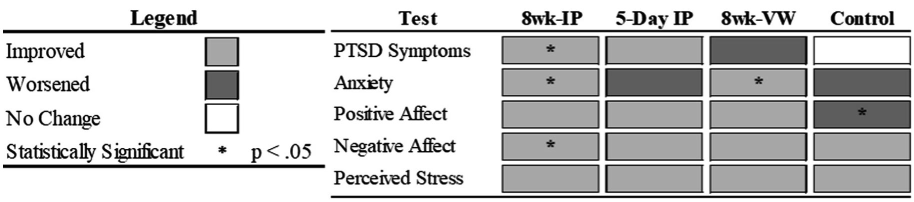


Fig. 1. Positive and negative changes from pre to post, along with significance levels are shown for each outcome measure and each mindfulness delivery method.

4 Discussion

Much of the research in the area of mindfulness training has focused on efficacy [21]. MBSR has been compared with a clinical derivative of MBSR, Mindfulness-Based Cognitive Therapy [50]. The purpose of this study was to compare the two delivery methods of MBSR training, an 8-wk mindfulness program delivered over a virtual world, and a wait-list control group on outcome measures of anxiety, stress, affect, and PTSD symptoms.

The primary finding of this study is that the delivery method has a measurable effect on outcomes. While the sole group difference showed greater reductions in negative affect among the 8-IP group, no other statistically significant differences were seen between delivery systems. When combining this with the within group comparisons, he 8-IP format produced the most beneficial outcomes with participants improving in all areas, and significant pre to post improvements in anxiety, negative affect, and reduction of PTSD symptoms.

The lack of differences between groups may be due to the small sample size coupled with relatively large standard deviations of the outcome measures. This study was conducted with individuals who did not have a specific diagnosis, and baselines showed near normal scores for each measure. These findings support earlier studies showing more moderate improvements are found among non-patient (healthy) populations [26], than patient populations [51]. The reduction in negative affect for the 8-IP group supports research showing a 20% reduction in negative affect following MBSR training [52].

The superiority of the 8-IP group in this study may be due to two things: class duration and in-person interaction. A class that includes both didactic learning and behavioral change may require additional time for information integration and both learning and assimilation of new techniques into ones' lifestyle. For example, to achieve behavioral change among those with addictions, experts recommend an intervention duration of at least three months [53]. Changes in health-related behaviors tend to occur in stages of awareness, contemplation of whether a change is desired and preparation to make a change, an action stage, and a maintenance and relapse prevention stage – and these stages take time [54]. While both IP and VW training allow for establishing personal relationships with the instructors and group members, only in-person training allows for moment-by-moment accountability, eye contact, seeing one-another's facial expressions and body language, and personal physical contact (such as a handshake, hand-on-shoulder, or a hug). In a 'learning by doing' context, such as mindfulness meditation, this personal connection may be particularly helpful. However, the slight increase in PTSD symptoms among the 8-VW group is important to notice, as we previously found those with higher scores on the PCL-M were less likely to complete mindfulness training over a VW [55]. Additional research on the use of telehealth by those with PTSD is necessary to fully assess its' utility. The 8-VW group did also appeared to have some benefits for attendees, showing significant improvements in anxiety and non-significant improvements in stress and affect. These improvements may also be due to the greater allotted time to learn, practice, ask questions, and integrate the meditation into one's daily life. Although there is considerable personal variation, research shows habit formation takes approximately 66 days [56], ten days longer than the traditional 8-week MBSR program, and considerably longer than the 5-day training.

While the intensive five-day format showed improvements in all except anxiety, none were significant pre to post training. Previous research found traditional 8-week in-person MBSR to improve emotion regulation, thereby improving symptoms among patients with generalized anxiety disorder [27]. The compressed nature of this method may not alleviate anxiety, as participants learn didactic information and practice meditation activities/behaviors all day for only 4½ days. It is unlikely they have time to integrate either the knowledge or the practice. Our findings are contrary to previous research investigating abbreviated forms of mindfulness training that found decreases in anxiety immediately post class [57] and two months later [58], but support other studies finding non-significant reductions in anxiety [59].

The Control group showed non-significant improvements in perceived stress and negative affect, no change in PTSD symptoms, and a significant decrease in positive affect. The Control group experienced the least benefit of the four groups, although this was not statistically significant.

5 Limitations

The primary limitation of this study was the small number of participants per group Also, these results are based upon a sample of U.S. military active duty, reserves, and veterans. Care should be taken in generalizing these results to other populations.

6 Conclusions

Mindfulness meditation training is offered in a multitude of ways, differing in content, duration, and delivery, with positive results shown in the majority of published studies [32]. However, this is the first study to demonstrate that different delivery systems yield different results for MBSR training and for mindfulness training based on MBSR. One statistically significant group difference was found, and other noticeable differences in outcomes were seen, suggesting that longer duration, in-person MBSR training may be superior for improving overall well-being. Further research is encouraged to explore benefits gained from various mindfulness training programs in terms of costs (time) and benefits (self-change improvements). Our findings also warn against the inclusion of all forms of mindfulness meditation training in meta-analyses, as the various trainings do not produce the same results. Finally, these results suggest identifying goals of mindfulness training, that is, if an introduction to the concepts and practices of mindfulness is the goal, then a five-day MBSR class should suffice. However, if the goal is to improve overall well-being, this study suggests the 8-wk in-person class is preferable.

Acknowledgements. Special thanks to the U.S. military service members and veterans who participated in this study. This research was supported in part by the U.S. Army Study Program

Management Office (ASPMO). Our appreciation is offered to our colleagues Mr. Gary Boykin, Ms. Angela Jeter, Mr. Jim Hewson, Dr. Baoxia Liu, Ms. Jessica Villarreal, and Mr. Jim Hewson.

Disclaimer. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Army, Department of Defense, or the U.S. Government.

References

1. Vanderploeg, R.D., Belanger, H.G., Horner, R.D., Spehar, A.M., Powell-Cope, G., Luther, S.L., Scott, S.G.: Health outcomes associated with military deployment: mild traumatic brain injury, blast, trauma, and combat associations in the Florida National Guard. Arch. Phys. M. **93**, 1887–1895 (2012)
2. Britt, T.W., Sinclair, R.R., McFadden, A.C.: Introduction: the meaning and importance of military resilience. In: Sinclair, R.R., Britt, T.W. (eds.) Building Psychological Resilience in Military Personnel: Theory and Practice. American Psychological Association, Washington, D.C. (2013)
3. Meredith, L.S., Sherbourne, C.D., Gaillot, S.J., Hansell, L., Ritschard, H.V., Parker, A.M., Wrenn, G.: Promoting Psychological Resilience in the U.S. Military. RAND Corporation, Santa Monica (2011)
4. Windle, G., Bennett, K.M., Noyes, J.: A methodological review of resilience measurement scales. Health Qual. Life Outcomes **9**, 8 (2011)
5. Rice, V.J., Liu, B.: Personal resilience and coping with implications for work part I: a review. Work **54**, 325–333 (2016)
6. Rice, V.J., Liu, B.: Personal resilience and coping Part II: identifying resilience and coping among U.S. military service members and veterans with implications for work. Work **54**, 335–350 (2016)
7. Beasley, M., Thompson, T., Davidson, J.: Resilience in responses to life stress: the effects of coping style and cognitive hardiness. Pers. Individ. Diff. **34**, 77–95 (2003)
8. Souri, H., Hasanirad, T.: Relationship between resilience, optimism and psychological well-being in students of medicine. Procedia Soc. Behav. Sci. **30**, 1541–1544 (2011)
9. Ong, A.D., Bergeman, C.S., Bisconti, T.L., Wallace, K.A.: Psychological resilience, positive emotions, and successful adaptation to stress in later life. J. Pers. Soc. Psychol. **91**, 730–749 (2006)
10. Beutel, M.E., Glaesmer, H., Wiltink, J., Marian, H., Brahler, E.: Life satisfaction, anxiety, depression and resilience across the life span of men. Aging Male **13**, 32–39 (2010)
11. Friborg, O., Hjemdal, O., Rosenvinge, J.H., Martinussen, M., Aslaksen, P.M., Flaten, M.A.: J. Psychosom. Res. **61**, 213–219 (2006)
12. Tugade, M.M., Fredrickson, B.L.: Resilient individuals use positive emotions to bounce back from negative emotional experiences. J. Pers. Soc. Psychol. **86**, 320–333 (2004)
13. Pietrzak, R.H., Johnson, D.C., Goldstein, M.B., Malley, J.C., Rivers, A.J., Morgan, C.A., Southwick, S.M.: Psychosocial buffers of traumatic stress, depressive symptoms, and psychosocial difficulties in veterans of Operation Enduring Freedom and Iraqi Freedom: the role of resilience, unit support, and post-deployment social support. J. Affect. Disord. **120**, 188–192 (2010)
14. Gooding, P.A., Hurst, A., Johnson, J., Tarrier, N.: Psychological resilience in young and older adults. Int. J. Geriatr. Psychiatry **27**, 262–270 (2012)

15. Pfefferbaum, B.J., Reissman, D.B., Pfefferbaum, R.L., Klomp, R.W., Gurwitch, R.H.: Building resilience to mass trauma events. In: Doll, L.S., Bonzo, J.A., Sleet, D.A., Mercy, J. A. (eds.) Handbook of Injury and Violence Prevention. Springer, New York (2007)
16. Jackson, D., Firtko, A., Edenborough, M.: Personal resilience as a strategy for surviving and thriving in the face of workplace adversity: a literature review. J. Adv. Nurs. **60**, 1–9 (2007)
17. Cornum, R., Matthews, M.D., Seligman, M.E.P.: Comprehensive soldier fitness: building resilience in a challenging institutional context. Am. Psychol. **66**, 4–9 (2011)
18. Jha, A., Morrison, A.B., Dainer-Best, J., Parker, S., Rostrup, N., Stanley, E.A.: Minds "at attention": mindfulness training curbs attentional lapses in military cohorts. PLoS ONE **10**, e0116889 (2015)
19. Rice, V.J., Schroeder, P.J.: Self-reported stress, fatigue, pain and sleep among U.S. Military Service members. In: Proceedings of the 8th International Conference on Applied Human Factors and Ergonomics (AHFE) 2017 and Affiliated Conferences, Los Angeles, CA (2017)
20. Rice, V.J., Schroeder, P.: Are those who are more mindful also more resilient? Dispositional resilience and mindfulness among active duty and veteran U.S. Military. In: International Conference on Mindfulness, 10–13 July, Amsterdam (2018)
21. Sharma, M., Rush, S.E.: Mindfulness-based stress reduction as a stress management intervention for healthy individuals: a systematic review. Evid. Based Complement. Alternat. Med. **19**, 271–286 (2014)
22. Kabat-Zinn, J.: Some reflections on the origins of MBSR, skillful means, and the trouble with maps. Contemp. Buddhism **12**, 281–306 (2011)
23. Vollestad, J., Sivertsen, B., Nielsen, G.H.: Mindfulness-based stress reduction for patients with anxiety disorders: Evaluation in a randomized controlled trial. Behav. Res. Ther. **49**, 281–288 (2011)
24. Goldin, P.R., Ziv, M., Jazaieri, H., Hahn, K., Gross, J.J.: MBSR vs aerobic exercise in social anxiety: fMRI of emotion regulation of negative self-beliefs. Soc. Cogn. Affect. Neurosci. **8**, 65–72 (2013)
25. Baer, R.A., Carmody, J., Hunsinger, M.: Weekly change in mindfulness and perceived stress in a mindfulness-based stress reduction program. Clin. Psychol. (New York) **68**, 755–765 (2012)
26. Khoury, B., Sharma, M., Rush, S.E., Fournier, C.: Mindfulness-based stress reduction for healthy individuals: a meta-analysis. J. Psychosom. Res. **78**, 519–528 (2015)
27. Goldin, P.R., Gross, J.J.: Effects of mindfulness-based stress reduction (MBSR) on emotion regulation in social anxiety disorder. Emotion **10**, 83–91 (2010)
28. Rice, V.J., Liu, B., Schroeder, P.J.: Impact of in-person and virtual world mindfulness training on symptoms of post-traumatic stress disorder and attention deficit and hyperactivity disorder. Mil. Med. **183**, 413–420 (2018)
29. Center for Mindfulness. https://www.umassmed.edu/cfm/mindfulness-based-programs/mbsr-courses/mindfulness-tools
30. Kabat-Zinn, J.: Personal communication, 29 March 2012
31. Kabat-Zinn, J.: Mindfulness-based interventions in context: past, present, and future. Clin. Psychol. Sci. Pract. **10**, 144–156 (2003)
32. Carmody, J., Baer, R.A.: How long does a mindfulness-based stress reduction program need to be? A review of class contact hours and effect sizes for psychological distress. Clin. Psychol. (New York) **65**, 627–638 (2009)
33. de Vibe, M., Bjorndal, A., Fattah, S., Dyrdal, G.M., Halland, E., Tanner-Smith, E.E.: Mindfulness-based stress reduction (MBSR) for improving health, quality of life and social functioning in adults: a systematic review and meta-analysis. Campbell Systematic Reviews, vol. 11 (2017)

34. Shian-Ling, K., Smoski, M.J., Robins, C.J.: Effects of mindfulness on psychological health: a review of empirical studies. Clin. Psychol. Rev. **31**, 1041–1056 (2011)
35. Rice, V., Alfred, P., Villarreal, J., Jeter, A., Boykin, G.: Human factors issues associated with teaching over a virtual world. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 56, pp. 1758–1762 (2012)
36. Jarmon, L., Traphagan, T., Mayrath, M., Trivedi, A.: Virtual world teaching, experiential learning, and assessment: an interdisciplinary communication course in second life. Comput. Educ. **53**, 169–182 (2009)
37. Gorini, A., Gaggioli, A., Vigna, C., Riva, G.: A second life for eHealth: Prospects for the use of 3-D virtual worlds in clinical psychology. J. Med. Internet Res. **10**, e21 (2008)
38. Morie, J., Chance, E.: Extending the reach of health care for obesity and diabetes using virtual worlds. J. Diabetes Sci. Technol. **5**, 272–276 (2011)
39. Stendal, K., Balandin, S., Molka-Danielsen, J.: Virtual worlds: a new opportunity for people with lifelong disability? Am. J. Intellect. Dev. Disabil. **36**, 80–83 (2011)
40. Brown, R.A., Rasmussen, R.K., Baldwin, I., Weth, P.: Design and implementation of a virtual world training simulation of ICU first hour handover processes. Aust. Crit. Care **25**, 178–187 (2012)
41. Zook, A., Lee-Urban, S., Riedl, M.O., Holden, H.K., Sottilare, R.A., Brawner, K.W.: Automated scenario generation: toward tailored and optimized military training in virtual environments. In: Proceedings of the International Conference on the Foundations of Digital Games, 29 May–1 June, Raleigh, NC (2012)
42. Zung, W.W.K.: A rating instrument for anxiety disorders. Psychosomatics **8**, 371–379 (1971)
43. Cohen, S., Kamarck, T., Mermelstein, R.: A global measure of perceived stress. J. Health Soc. Behav. **24**, 385–396 (1983)
44. Watson, D., Clark, L.A., Tellegen, A.: Development and validation of brief measures of positive and negative affect: the PANAS scales. J. Pers. Soc. Psychol. **54**, 1063–1070 (1988)
45. Weathers, F., Litz, B., Herman, D., Huska, J., Keane, T.: The PTSD checklist (PCL): reliability, validity, and diagnostic utility. Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX (1993)
46. Cohen, S., Janicki-Deverts, D.: Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J. Community Appl. Soc. Psychol. **42**, 1320–1334 (2012)
47. Crawford, J.R., Henry, J.D.: The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br. J. Clin. Psychol. **43**, 245–265 (2004)
48. Kimerling, R.: Examining the diagnostic and clinical utility of the PTSD checklist. Health Services Research & Development Study, SHP 08–169 (2009) https://www.hsrd.research.va.gov/research/abstracts.cfm?Project_ID=2141698713
49. Imbens, G.W., Rubin, D.B.: Causal Inference for Statistics, Social, and biomedical Sciences: An Introduction. Cambridge U.P, New York (2015)
50. Fjorback, L.O., Arendt, M., Ornbol, E., Fink, P., Walach, H.: Mindfulness-based stress reduction and mindfulness-based cognitive therapy – a systematic review of randomized controlled trials. Acta Psychiatr. Scand. **124**, 102–119 (2011)
51. Khoury, B., Lecomte, T., Fortin, G., Masse, J., Therien, P., Vouchard, V., Chapleau, M., Paquin, K., Hofmann, S.G.: Mindfulness-based therapy: a comprehensive meta-analysis. Clin. Psychol. Rev. **33**, 763–771 (2013)
52. Martin-Asuero, A., Garcia-Banda, G.: The mindfulness-based stress reduction program (MBSR) reduces stress-related psychological distress in healthcare professionals. Span. J. Psychol. **13**(2), 897–905 (2010)

53. National Institute on Drug Abuse: Principles of Drug Addiction Treatment: A Research-Based Guide, 3rd edn. https://www.drugabuse.gov/publications/principles-drug-addiction-treatment-research-based-guide-third-edition/principles-effective-treatment
54. Zimmerman, G.G., Olsen, C.G., Bosworth, M.F.: A 'stages of change' approach to helping patients change behavior. Am. Fam. Physician **61**, 1409–1416 (2000)
55. Liu, B., Rice, V.J.: Characteristics of United States military personnel and veterans who complete mindfulness training. In: Duffy, V., Lightner, N. (eds.) Advances in Human Factors and Ergonomics in Healthcare and Medical Devices. Advances in Intelligent Systems and Computing, vol. 590. Springer, Cham (2017)
56. Lally, P., van Jaarsveld, C.H.M., Potts, H.W.W., Wardle, J.: How are habits formed: modeling habit formation in the real world. Eur. J. Soc. Psychol. **40**, 999–1009 (2010)
57. Smith, B., Metzker, K., Waite, R., Gerrity, P.: Short-form mindfulness-based stress reduction reduces anxiety and improves health-related quality of life in an inner-city population. Holist. Nurs. Pract. **29**, 70–77 (2015)
58. Rice, V.J., Boykin, G.L., Overby, C.R., Jeter, A., Villarreal, J.: A pilot study evaluating the utility, acceptability, and feasibility of an abbreviated mindfulness meditation program before and during Army warfighting training. In: Duffy, V., Lightner, N. (eds.) Advances in Human Factors and Ergonomics in Healthcare and Medical Devices. Advances in Intelligent Systems and Computing, vol. 590. Springer, Cham (2017)
59. Duarte, J., Pinto-Gouveia, J.: Effectiveness of a mindfulness-based intervention on oncology nurses' burnout and compassion fatigue symptoms: a non-randomized study. Int. J. Nurs. Stud. **64**, 98–107 (2016)

Human Reliability Analysis for Safety Critical Industries

Experimental Research on Measurement of Team Situation Awareness in Nuclear Power Plants

Guimei Zuo, Jianhua Chen, and Licao Dai(✉)

Human Factor Institute, University of South China, Hengyang, Hunan, China
632616255@qq.com

Abstract. Endsley (1995) systematically put forward the theory of Situation Awareness (SA). The theory emphasizes people's perception and understanding of current operating environment information, and prediction of future state under certain time and space conditions. Endsley and Jones (2012) further proposed a SA-Oriented Design system. With the deepening of SA research, it has become the frontier of engineering psychology research. At present, Nuclear Power Plants (NPP) are highly computerized, the complexity of its main control room has increased, man-machine interaction is frequent, tasks are complex and changeable, time pressure is high, and cognitive load is heavy, so the workload of operators is greatly increased. In this complex technical system, the accuracy of Team Situation Awareness (TSA) of NPP is an important factor for high quality decision-making and efficient operation. If the problem of real-time and accurate SA measurement can be solved, it will be helpful to the design of Real-Time Adaptive man-machine interaction. Firstly, this paper analyses the elements and types of Team Situation Awareness of nuclear power plant, and develops a team SA measurement scale. Then, in different accident scenarios, the level of team SA of different types was measured by the simulation experiment, and the influencing factors were analyzed. These provide theoretical and experimental support for improving TSA of NPP.

Keywords: Nuclear Power Plant (NPP) · Team · Situation Awareness (SA)

1 Introduction

1.1 Situational Awareness

The concept of situational awareness (SA) first appeared in the field of aviation to describe the pilots' understanding of the combat flight process. With the continuous progress of science and technology, more and more complex systems need people to operate to accomplish specific tasks, so SA has also been studied by more and more researchers as an important concept of anthropology. In the early stage of SA research, different scholars have different definitions. Endsley (1995) defines SA as "the perception of various elements in the environment in a specific time and space, the understanding of its meaning and the prediction of its future state", and divides SA into three levels. The first level is perception, which is the most basic link in SA. The

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 199–213, 2020.
https://doi.org/10.1007/978-3-030-20037-4_18

second level is understanding, that is, integrating different information and making decisions about goals. The third level is prediction on future situation events, which is the highest level of SA.

Compared with the definition, the SA model can describe the cognitive structure in more detail and intuitively. Researchers can use SA model to develop corresponding measurement methods and applications. Therefore, in the study of SA, besides definitions, researchers also put forward different views on the conceptual model of SA.

In addition to the three stages of "perception-understanding-prediction", Endsley also studied the interaction between SA and other cognitive processes such as long-term memory, schema, working memory, etc. In this model, mental models based on past experience and knowledge is the key elements for developing and maintaining SA. The mental model guides the operator to obtain the key information needed by SA from the environment, and to understand the information based on past training and experience, so as to simulate and forecast the possible future situation. Because of its operability, this model can guide researchers to measure and train SA at different stages, so it is also the most widely accepted model.

1.2 Team Situational Awareness

Although the concept of team situational awareness (TSA) was put forward very early, due to the complexity of the team itself, there is still a relatively preliminary stage in the study of TSA. Many scholars have different opinions on its concept and measurement method, so there has been no unified concept and measurement method. From the existing studies, TSA is considered to be an important factor affecting team output, evaluating team performance and improving team performance, which deserves further study by researchers. Salas et al. (1995) believed that TSA should be "the shared cognition of current situation among team members at a specific time", and he emphasized the important influence of team process and communication on TSA.

1.3 TSA Model

Based on the definition of TSA, different scholars put forward different models of TSA. Generally speaking, there are two main conceptual models.

(1) Model Based on Individual SA and Shared SA. For a team, the individual plays an important role as the basic unit, and the team itself depends on communication and exchange with the individual, so the team can be regarded as the whole of communication and exchange between the individual and the individual. From this point of view, Endsley (2001) proposed that TSA includes individual SA and the Sharing SA brought about by communication between individuals, as shown in Fig. 1.

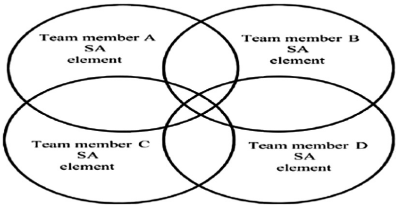


Fig. 1. Endsley TSA model

For individuals, they only need to understand the goals of their actions and key operational factors. But for teams, members need to understand the goals of the team and other members' SA. TSA requires members to have a common awareness of the current environment and team situation at specific time points in the operation process.

Shu and Furuta (2005) proposed a model based on individual SA and mutual SA. Interactive SA refers to team members' understanding of other team members' level of SA and other team members' confidence in their level of SA.

(2) Distributed TSA Model. Garbis et al. (2008) proposed a distributed TSA model. This model does not start from the shared SA among individual members, but from the point of view of the whole team, it considers that TSA should be the result of communication and convergence of SA among different members. Stanton et al. (2009) summarized the distributed TSA model in detail. They believed that the focus of TSA was team communication and exchange. Distributed TSA Model considers that communication exists not only between individuals, but also between individuals and non-individuals, and even between non-individuals and non-individuals. TSA is influenced by these three kinds of communication.

(3) Model comparison. Endsley model is suitable for small and medium-sized teams. In these two teams, there is a large overlap and sharing of SA between members. At this time, Endsley model can describe TSA more accurately and concisely. In large teams, it is entirely possible that there is no and no need for overlapping parts of SA between two members. Therefore, for large teams, the distributed TSA model can do better analysis.

1.4 Necessity of SA Study for NPP Teams

With the expanding scope of SA research, it has become the frontier of engineering psychology research. There are several common points in the research field of SA at the present stage: firstly, the environment is dynamic, the information is complex and the amount of information is large; secondly, accurate state and goal setting is needed; thirdly, the psychological load of operators may be heavy; fourthly, a rich knowledge system and a lot of operation training are needed. In these complex technological systems, the accuracy of operator's SA is an important factor in deciding high-quality decision-making and efficient operation. If the SA is very poor, it will not only fail to complete complex cognitive tasks, but also may lead to disastrous consequences.

With the highly computerized development of NPP, the main control room of NPP has become a digital control room. Compared with the traditional main control room, although the digital main control room has strong data acquisition ability, automatic monitoring ability, more abundant parameter information display and automatic control, it also brings unprecedented challenges to the main control room operators: (1) too much information, (2) errors of operation and control, (3) difficulties on information exchange. Digital control system provides a lot of information; it can also lead to poor intuition of monitoring and control, which can easily lead to human error. This also weakens the operator's SA.

In high-risk complex industrial systems such as NPPs, the complexity of the system is improved, human-computer interaction is frequent, tasks are complex and changeable, time pressure is great, and cognitive load is heavy. Much work needs to be done by teams together. Team performance is particularly important. TSA is the key factor to determine the level of team decision-making and behavioral performance, which has been paid more and more attention by researchers. Therefore, with the increasing collaboration tasks of NPP team, how to comprehensively evaluate the SA of NPP team from the perspective of human factor engineering, detect possible human error factors in NPP team tasks and eliminate them through training is an important direction to be studied. This evaluation method is of great significance to the NPP teams performing complex tasks.

2 SA and Its Types of NPP Teams

2.1 Elements of SA of NPP Teams

The level of individual SA is influenced by individual knowledge structure, cognitive ability, physiological and psychological status and social aspects, and there are great differences. The level of TSA is related to the collocation of NPP teams and the leadership art of team leaders. The overall work of NPP teams is accomplished by the division, coordination and cooperation of members. Operators constructs scenario model by using of their understanding of power plant knowledge and operation experience. When the situational model can accurately reflect the actual state of the power plant, it shows that the NPP team has good SA.

Shu and Furuta (2005) studied the elements of TSA, including sharing information display, communication, common knowledge base, role assignment and so on. Salas et al. (2005) conducted an extensive literature review and identified five TW factors and three coordination mechanisms. On the basis of Salas's research, this paper considers that the elements of SA of NPP teams are shown in Table 1. At the same time, in order to quantitatively analyze the SA level of NPP teams, according to different grade criteria, the variables are divided into good, medium and poor grades, which are expressed by 5, 3 and 1 respectively.

Table 1. Elements of SA of NPP teams

Variable	Status grade	Criteria/description	Explanation
Leadership (Mearns et al. 2001)	Good (5) Medium (3) Poor (1)	As a team leader, he has the determination to make accurate judgments and decisions; the ability to select the necessary procedures and put them into practice; and the ability to judge the changes of the situation flexibly	The shift manager and shift supervisor typically function as the team leaders in NPP MCRs Task-oriented skills: e.g. Setting goals and delegating tasks; social skills: e.g., conflict resolution
Mutual performance monitoring (Weil et al. 2004)	Good (5) Medium (3) Poor (1)	Monitoring other team members and providing feedback and support	In NPPs, operators often check each other's instrument readings to ensure accuracy; known as peer checking
Backup behavior (Marks et al. 2001)	Good (5) Medium (3) Poor (1)	Advising, assisting, or performing a task for an overloaded team member	If a control board operator has to interrupt his or her monitoring of the board, another operator will step in temporarily
Adaptability (Stachowski et al. 2009)	Good (5) Medium (3) Poor (1)	In an emergency, the team must quickly adapt to rapidly changing conditions	Adapting the individual actions of team members to produce coordinated team action
Team orientation (Driskell and Salas 1992)	Good (5) Medium (3) Poor (1)	Being motivated towards and capable of working constructively with others	Coordination, evaluation and utilization of task inputs from other members is critical for the NPP operator
Coordinating mechanisms	Good (5) Medium (3) Poor (1)	The tacit understanding of team members' coordination and cooperation; the effectiveness of communication; the quality of communication	The ability to communicate and inquire on questions that are questioned and to obtain useful results; coordination, assessment and utilization of task inputs from other members are essential for NPP operators.
Shared mental models (Lim and Klein 2006)	Good (5) Medium (3) Poor (1)	Team updates ensure that all team members have the same information about plant functioning	Shared understanding and knowledge helps each team member to take Appropriate, coordinated steps for maintaining task performance

(*continued*)

Table 1. (*continued*)

Variable	Status grade	Criteria/description	Explanation
Closed-loop communication (Lin et al. 2011)	Good (5) Medium (3) Poor (1)	The delivery of information from the giver to the receiver, the acknowledgement that the information was received by repeating what was heard, and the approval from the giver that the information was processed correctly.	Verifiably accurate communication is critical for accomplishing tasks in the mcr e.g., use of three-way communication protocols
Mutual trust (Salas et al. 2005)	Good (5) Medium (3) Poor (1)	Mutual reliance and accepting risk by allowing others to take responsibility	Trust among crew members is imperative to cohesive and Effective team performance

2.2 Influencing Factors of TSA in NPPs

In the working environment, the factors affecting human behavior are called behavioral influencing factors (PIFs) or behavioral forming factors (PSFs). Bad PIFs more or less increase the possibility of failure.

Lee et al. (2012) analyzed the PIFs influencing the shared mental model, and obtained the factors including team knowledge, team skills, team attitude, team motivation, team environment and so on. In the "three-level" model, Endsley et al. pointed out that the influencing factors of SA can be divided into two categories: external factors and internal factors. External factors include interface design, complexity, automation level, workload and pressure of equipment and facilities, and internal factors include attention, memory, and experience and so on.

According to our previous research, SA of operators is affected by individual factors, situational state factors and organizational factors. Based on the existing research results, considering that SA errors are human errors, TSA will also be affected by these PIFs. Therefore, we include external factors, internal factors and organizational factors into the PIFs of TSA, as detailed in Table 2.

When each factor is in a different state, the degree of influence on perception, understanding and prediction is also different. Therefore, in order to quantitatively analyze the main factors affecting TSA, according to different grade criteria, the variables are divided into good, medium and poor grades, which are expressed by 5, 3 and 1 respectively.

Table 2. Factors affecting SA of NPP teams

Influence factor	Subclass	Status grade	Criteria/description	Explanation
External factors	Interface design of equipment and facilities	Good (5) Medium (3) Poor (1)	Interface color, display screen and other settings meet the requirements of human factors engineering design	Interface design can provide the required information and perform tasks in a simple and error-free manner; meet the requirements of human factor engineering design
	Complexity	Good (5) Medium (3) Poor (1)	The matching degree between task complexity and human processing ability	Complex tasks may occupy a large amount of cognitive resources and affect SA
	Rules	Good (5) Medium (3) Poor (1)	Integrity of rules: detail and adequacy of rules; comprehensibility of rules, etc.	Regulations are available and state-oriented. They can keep power plants in a safe state without accurate diagnosis of events. All they need to do is mitigate events
	Workload	Good (5) Medium (3) Poor (1)	Requirements for cognition, action (speed, strength, accuracy)	Although workload and SA are two independent structures, they are related to each other
	Time pressure	Good (5) Medium (3) Poor (1)	Requirements for communication and cooperation, calculation, completion time and speed, etc.	If the operator is difficult to diagnose and deal with the problem within the available time, the pressure will be high
Internal factors	Attention	Good (5) Medium (3) Poor (1)	The degree to which attention is focused in this situation	Attention to current tasks; attention and alertness to the surroundings
	Memory	Good (5) Medium (3) Poor (1)	Mastery of knowledge and tasks related to NPPs	Strong knowledge of NPPs

(*continued*)

Table 2. (*continued*)

Influence factor	Subclass	Status grade	Criteria/description	Explanation
		Good (5) Medium (3) Poor (1)	Training opportunities; frequency and experience in performing similar tasks	Training system; peacetime training for handling abnormal events; accident handling experience
Organizational factors	Organizational objectives and strategies	Good (5) Medium (3) Poor (1)	Integrity and concreteness of objectives and strategic systems; contradictions between current and long-term objectives	Formulation of high-level organizational plans
	Organization structure	Good (5) Medium (3) Poor (1)	Rationality of organizational structure	Definition of organizational structure, responsibility and power
	Organization and management	Good (5) Medium (3) Poor (1)	Effectiveness of organizational management	Working methods/strategies, management priorities and secondary issues
	Organizational culture	Good (5) Medium (3) Poor (1)	Effectiveness of organizational culture	Affect staff cohesion and collective identity; affect safety attitude, safety measures, safety and benefit balance, etc.

2.3 SA Model of NPP Teams

Salas et al. (1995) proposed a conceptual analysis model of TSA, which includes two main parts: individual SA and team process. This model describes the internal dynamic interaction characteristics of TSA and the interaction among its components, but it does not analyze the influencing factors of TSA and their interaction with TSA in detail. Salmon et al. (2009) established TSA model on the basis of previous research, including environmental data, SA development, individual SA, team process and TSA, and each part affects each other. Information-Decision-Behavior Model (IDAC) and O'Hara et al. (2009) general operator model have been applied to human factor engineering and human reliability analysis (HRA) in digital NPPs.

Salmon et al. (2008) holds that as a higher level organism, team has SA which is independent of individual SA, dependent but higher than individual level. The perception-understanding-prediction model of individual SA can also be used in TSA model.

This paper constructs the formation process of TSA from the perspective of macro-cognitive process, and identifies the possible failure modes and influencing factors of TSA, as shown in Fig. 2.

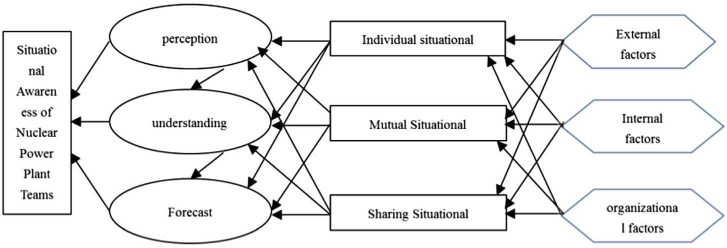


Fig. 2. SA model for NPP teams

3 SA Measurement and Experiments of NPP Teams

3.1 Measurement of TSA

1. SA Measurement Based on Endsley Model
The Endsley model can be used to measure the SA of individual members, and at the same time to measure their awareness of the current situation of the team and their understanding of other members' SA. For example, Shu and Furuta (2005) conducted a questionnaire survey based on individual SA and mutual SA. This study has a strong task dependence, that is, the content of the measurement and the evaluation method are highly related to the task itself, and the level of TSA of the two subjects is represented by their behavior consistency. Although this measurement method is more objective, it cannot reflect that TSA belongs to team process, and the dimension of measurement is relatively single.
2. Expert Evaluation Method
Expert evaluation method refers to inviting experts in the field to evaluate the team as a whole. Based on the fluency of team communication process, problems and final team performance, experts score and evaluate team performance on the basis of natural observation, record and later analysis. This method is mostly used in the evaluation and measurement of operating room teams and military teams. In this method, on the one hand, the evaluator needs real-time observation and detailed record of the teams task completion process, and then interviews different individuals according to their needs by using critical event assessment method, so as to obtain first-hand interview materials for teams and individuals; on the other hand, the evaluator needs to be an expert in the context of the task and be familiar with the team's standard ethics. Work flow and corresponding performance standards, on this basis, we also need to have a more professional understanding of team building,

team communication and other content. This method obviously cannot meet the requirements of rapid and simple measurement of TSA.

3. CAST (coordinated awareness of situation by teams)
   CAST measurement method considers that the core feature of TSA is to adapt to the dynamic changes of the environment in order to achieve team goals. Then, in the process of accomplishing a specific goal, if we introduce unexpected changes and obstacles artificially, and then record a series of behaviors that the team produces in response to the unexpected situation, including information collection, communication and communication, as well as the final treatment methods, and evaluate and analyze these behaviors, we can evaluate the team's SA.

Based on the fluency of team communication process, problems and final team performance, this paper evaluates the team as a whole. Subjective assessment is used to measure SA based on the opinions of operators and observers. The advantages of subjective evaluation method are easy to implement and low cost. It can be used to simulate scenarios, and can also be used in the actual task environment.

3.2 SA Experiment of NPP Teams

The existing research on SA reliability mainly obtains the required data through simulation experiments of complete accidents. In digital NPPs, once abnormal conditions (such as LOCA and SGTR accidents) occur, operators should calmly and diligently think and understand the abnormal information they observe and make decision response. Fifty-six LOCA and SGTR accident reports were collected from 2011 to 2013. On this basis, simulation experiments were carried out on the process of handling LOCA and SGTR accidents, and 56 sets of sample data were obtained for analysis.

1. Experimental design
   In this study, trainees were selected as subjects to conduct simulation experiments on LOCA and SGTR accident handling process. Firstly, the trainees are grouped into groups, and each group is assigned a group leader. The team leader then chooses the members of the team and makes them play different roles according to the different characteristics of each team member. The team leader plans, coordinates and makes decisions according to the simulation tasks, communicates with the members of the team fully, makes feasible plans, and then divides work and cooperates. The experimenter acts as an observer and evaluator to evaluate each link of the simulation experiment.
   ① The experimenters observed the performance and action of the team during the whole process, and filled in the Situation Awareness Scale and the Influencing Factors Scale, as shown in Tables 1 and 2.
   ② After the end of the experiment, the trainees supplement the Situation Awareness Scale and the Influencing Factors Scale according to their current understanding/memory.

2. Data Analysis and Processing
   SPSS17.0 statistical software was used to collate and analyze the data.
   ① Validity tests of collected data are conducted to eliminate invalid data.
   ② Estimate the team's SA by subjective assessment.
   ③ Using correlation analysis to identify the main influencing factors of SA.

4 SA Level of NPP Teams and Analysis of Its Influencing Factors

4.1 SA Level Analysis of NPP Teams

After completing the scale, the experimenter counted the test results of each scale, and used the Fuzzy Comprehensive Evaluation to evaluate the SA level of NPP teams.

1. Scores of Each Dimension
   Each dimension of NPP TSA can be reflected by the score of each item in the scale. The higher the score of each item, the better the SA of the NPP team, as shown in Table 3.

Table 3. SA assessment Form for NPP Teams

Variable	Weight	Variable score
Leadership	0.1	2.389
Mutual performance monitoring	0.1	3.278
Backup behavior	0.1	2.611
Adaptability	0.1	3.611
Team orientation	0.1	2.500
Coordinating mechanisms	0.15	3.667
Shared mental models	0.15	2.333
Closed-loop communication	0.1	3.444
Mutual trust	0.1	3.722
Composite score	1	3.056

   According to the results of the scale, the scores of each dimension of TSA rank from high to low: mutual trust, coordination mechanism, adaptability, closed-loop communication, mutual performance monitoring, standby behavior, team positioning, leadership, sharing psychological model.
2. Weight
   Five senior operators were invited to be evaluation experts according to their working years and experience. After fully explaining the contents of each scale, the weights of each scale factor were scored. The final weight of each dimension of the TSA scale is shown in Table 3.

3. Comprehensive Score
   The TSA was comprehensively assessed by weighted comprehensive evaluation method. The comprehensive score is shown in Table 3.

4.2 Analysis on the Influencing Factors of SA of NPP Teams

The correlation can quantitatively measure the correlation intensity between two variables. Pearson correlation analysis is used to measure the correlation level between influencing factors and group SA. Fifty-six sample data are analyzed by statistical analysis software SPSS17.0, and the correlation coefficient between influencing factors and group SA is obtained, as shown in Table 4.

Table 4. Relevant levels between influencing factors and TSA

Influencing factors of SA	Pearson correlation	Significance (bilateral)
Interface design of equipment and facilities	.197	.250
Complexity	.297	.079
Rules	.099	.564
Workload	−.661**	.000
Time pressure	.025	.886
Attention	.808**	.000
Memory	.433**	.008
Training/experience	.681**	.000
Organizational objectives and Strategies	.281	.097
Organization structure	.128	.455
Organization and management	.759**	.000
Organizational culture	.490**	.002

*. Significant correlation at 0.05 level (bilateral)
**. Significant correlation at 0.01 level (bilateral)

Determine the correlation of influencing factors according to the absolute value of correlation coefficient greater than 0.3. Table 4 shows that there is a strong correlation between the level of TSA and workload, attention, memory, training experience, organizational management, organizational culture. The factors with strong correlation are described as follows.

1. Workload and TSA of NPP
   The influence of operator's situational cognitive measurement depends on the change of workload in specific situations. At present, when measuring Situational Cognition in foreign countries, most of them study the change of workload in corresponding situations. The latest research results show that there is no consistent relationship between the two. However, from the experimental results, different workloads have a certain impact on TSA.
   Through the experiment, we can preliminarily guess that the workload at normal level has little influence on TSA. In the case of high workload, the team's level of SA is relatively low. The reason is that high workload occupies too much cognitive

resources of operators, which leads to the low level of TSA. In the case of low workload, it cannot effectively improve SA. The reason is that under low workload, the cognitive resources of crew members are surplus, and there is no sense of urgency in situational prediction, which will result in a low level of SA.

2. Relationship between Attention and TSA of NPP
   According to the results of the scale evaluation, the NPP teams pay more attention in the process of operation, especially when the environment changes, the corresponding operators need to spend a lot of energy to increase the attention supply. When the attention is not focused, the most common is the visual and auditory ignorance of some information; it is difficult for operators to focus their attention on the handling of accidents. At this time, the SA of the NPP team is poor.
3. Relationship between Training/Experience and TSA of NPP
   There is evidence that the operator's sensory acuity can be trained. Trained/experienced operators can often be demonstrated by proficiency in specific environments and can quickly and accurately assess environmental cues. It is generally believed that trained/experienced operators have a memory reserve of multiple event patterns and can be extracted when needed.
4. Relation between Organizational Factors and TSA of NPP
   Organizational factors mainly include the arrangement of working hours and the design of workload. The unreasonable work design will cause the operator's mental load, which may result in the operator's misjudgment of information. When the attention is not focused, it will directly lead to the correct rate of people's visual, auditory and tactile senses to obtain information, which directly affects the level of group SA.

5 Discussion and Conclusions

5.1 Conclusion

This paper analyses the elements and types of SA of NPP teams, and then develops a SA measurement scale based on this. Then, the level of SA of different types of teams in different accident scenarios was measured by the simulation experiment, and the influencing factors were analyzed.

The experimental results show that the scores of mutual trust, coordination mechanism, adaptability, closed-loop communication and mutual performance monitoring are higher in TSA of NPP. Workload, attention, training/experience, organizational management and organizational culture have great influence on TSA.

From the analysis of SA level of NPP teams, in order to improve SA, we need to strengthen the following aspects: improving leadership ability, strengthening team positioning/clarifying the responsibilities of members, implementing mutual performance monitoring, strengthening closed-loop communication, etc.

From the perspective of the influencing factors of SA of NPP teams, in order to solve the problem of invalidation of SA, attention should be paid to the following aspects: suitable workload, accumulated manipulation knowledge, systematic training/strengthening safety education, etc.

5.2 Problems in Research

1. Complete accident scenario assessment should include three stages: surveillance stage, treatment stage and decision-making implementation stage. Considering the complexity of the experiment, the experiment has not yet distinguished the three.
2. The measurement and evaluation method is based on the manipulator's subjective emotions and perception factors. Due to the limitation of personal ability and experimental conditions, it lacks the corresponding verification of objective indicators.

5.3 Research Trends

1. Evaluation methods. To establish more accurate evaluation index, especially in the experimental process, the scale is relatively general, and the design of scale factors is not meticulous enough, which may lead to inaccurate evaluation results. In the future research work, we should refine the influencing factors as much as possible to make the evaluation results more scientific and reasonable.
2. The accuracy of TSA data and measurement. At present, TSA data is very scarce. The data basically come from expert judgment, or through HRA to assess the individual's diagnostic reliability, and then through correlation analysis to estimate the team's diagnostic reliability. However, due to many uncertainties, this analysis increases the deviation of TSA reliability results.
3. The causality between TSA errors and PSF, and the interaction between PSF. A situational environment-based identification technology should be developed to identify the relationship between TSA and PSF.

References

Adams, M.J., Tenney, Y.J., Pew, R.W.: Situation awareness and the cognitive management of complex systems. Hum. Factors **37**, 85–104 (1995)

Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–64 (1995)

Lee, S.W., Park, J., Kim, A.R., et al.: Measuring situation awareness of operation teams in NPPs using a verbal protocol analysis. Ann. Nucl. Energy **43**, 167–175 (2012)

Lee, S.W., Kim, A.R., Park, J., et al.: Measuring situation awareness of operating team in different main control room environments of nuclear power plants. Nucl. Eng. Technol. **48**(1), 153–163 (2016)

Li, P.C., Zhang, L., Dai, L.C., et al.: Study on operator's SA reliability in digital NPPs. Part 1: the analysis method of operator's errors of situation awareness. Ann. Nucl. Energy **102**, 168–178 (2017)

Lin, C.J., Hsieh, T.L., Yang, C.V., et al.: The impact of computer-based procedures on team performance, communication, and situation awareness. Int. J. Ind. Ergon. (2015). https://doi.org/10.1016/j.ergon.2014.12.001

O'Hara, J., et al.: Identification and evaluation of human factors issues associated with emerging nuclear plant technology. Nucl. Eng. Technol. **41**, 225–236 (2009)

Park, J.K.: Analysis of communication contents and the associated crew performance data collected from abnormal operating conditions for nuclear power plants. Korea Atomic Energy Research Institute, KAERI/TR-4159/2010 (2010)

Salas, E., Prince, C., Baker, D.P., et al.: Situation awareness in team performance: implications for measurement and training. Hum. Factors **37**, 123–136 (1995)

Salmon, P.M., et al.: Measuring situation awareness in complex systems: comparisons of measures study. Int. J. Ind. Ergon. **39**, 490–500 (2009)

Salmon, P.M., Stanton, N.A., Walker, G.H., et al.: What really is going on? Review of situation awareness models for individuals and teams. Theor. Issues Ergon. Sci. **9**(4), 297–323 (2008)

Salmon, P.M., Stanton, N.A.: Situation awareness and safety: contribution or confusion? Situation awareness and safety editorial. Saf. Sci. **56**(7), 1–5 (2013)

Shu, Y., Furuta, K.: An inference method of team situation awareness based on mutual awareness. Cogn. Technol. Work **7**, 272–287 (2005)

Stanton, N.A., Salmon, P.M., Walker, G.H., et al.: Genotype and phenotype schemata as models of situation awareness in dynamic command and control teams. Int. J. Ind. Ergon. **39**, 480–489 (2009)

Yang, C.W., Yang, L.C., Cheng, T.C., et al.: Assessing mental workload and situation awareness in the evaluation of computerized procedures in the main control room. Nucl. Eng. Des. **250** (9), 713–719 (2012)

Yim, H.B., Seong, P.H.: A quantitative team situation awareness measurement method considering technical and nontechnical skills of teams. Nucl. Eng. Technol. **65**, 144–157 (2014)

![Check for updates]

A Simulator Experimental Study on Individual SA and Team SSA in Digital MCRs of NPPs

Pengcheng Li[1,2], Xiao Jin[1,2], Yanxing Wang[1,2], Jianhua Chen[1,2], and Licao Dai[1,2(✉)]

[1] Human Factor Institute, University of South China, Hengyang 421001, Hunan, People's Republic of China
dailpc@sina.com

[2] School of Nuclear Science and Technology, University of South China, Hengyang 421001, Hunan, People's Republic of China

Abstract. Operator's situation awareness (SA) issues are more prominent in digital nuclear power plants (NPPs). In order to identify the SA levels of operators and levels of team shared situation awareness (SSA), a SA measure method was developed based on Situation Awareness Global Assessment Technique (SAGAT). Furthermore, it was used to measure the SA levels of operators by simulation experiments and a method was established to determine the team SSA level. The experimental results showed that the individual SA level was related to the SSA level. The higher the individual SA level is, the higher the SSA level is. Furthermore, for different experimental scenarios, the SA level of operators and team SSA level is different, which means that the more obvious the symptoms of the risky scenarios are, the relatively higher the SA and SSA level are.

Keywords: Situation awareness · Shared situation awareness · Simulator experiment · Nuclear power plant

1 Introduction

Team Situation Awareness (TSA) is very important for team tasks in high-risk systems. It is attributed to the monitoring of operating state of systems and the handling of abnormal states are completed by a team in complex industrial systems such as nuclear power plants (NPPs). A team is an important unit of organizational work because they bring diverse expertise, skills, and resources to complex tasks that may be too large or complex for a single individual to undertake [1]. Researches have shown that system safety is more dependent on team performance rather than individual performance in complex dynamic systems such as NPPs and air traffic control (ATC) [2] as well as team performance is positively correlated with TSA [3, 4]. Furthermore, if an individual makes a SA error, it may be detected, corrected and restored by other team members. Consequently, there is no doubt that the TSA, like individual SA, is also critical to the successful execution of team tasks. From the viewpoint of a broad level, TSA includes Individual Situation Awareness (ISA), Shared Situation Awareness (SSA) and Mutual Awareness (MA). However, from the perspective of a narrow level,

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 214–224, 2020.
https://doi.org/10.1007/978-3-030-20037-4_19

TSA only refers to SSA. For example, Salas et al. [5] thought that TSA is team members' common understanding of certain state of system or component at a certain point of time. In this paper, we adopt the latter view as our research object for measuring. Despite some research done in team performance [5, 6] and team situation awareness [7, 8], valid and reliable measures of team and shared SA are still lacking [9], especially in digitized NPPs. Therefore, based on the developed measure method, this paper wants to measure individual SA and team SSA by simulator experiments in digitized NPPs.

2 Team Shared SA

Endsley and Jones [10] claimed that team SA is mainly a matter of shared SA, i.e. possessing a common picture of a situation. Lloyd and Alston [11] argued that team members acquire individual SA and then communicate this throughout the team, which leads to a common team understanding. Nofi [12] described the content of SSA as shown in Fig. 1, which shows that the three circles represent the individual SAs of operator A, B, and C, respectively. The overlapping parts represent their shared SA. For example, the area AB means common awareness of person A and person B, the area ABC is their common SA among person A, B, C, namely team SSA. The Fig. 1 is a static representation of team SSA at a certain time point.

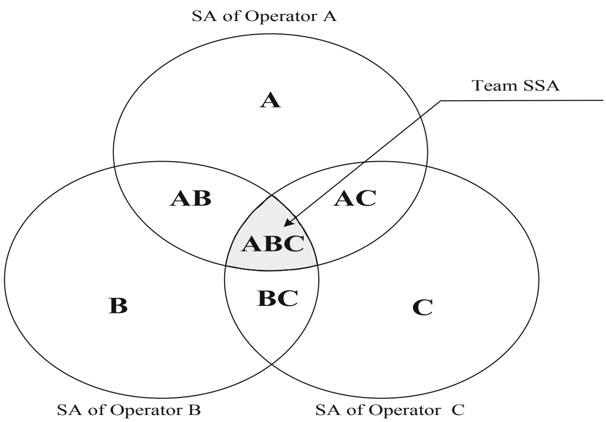


Fig. 1. Team shared SA including three people [12]

One of the main goals of system design is to enhance the operator's SA. In order to ensure that SA is promoted without being reduced, it needs to accurately measure the individual's SA and TSA. So far, a widely tested and validated approach for assessing SA is the Situation Awareness Global Assessment Technique (SAGAT) [13]. It has been widely used in SA and TSA measurement. Therefore, we develop a new SA measure method based on SAGAT to measure individual SA and team shared SA in digital MCRs of NPPs.

3 Method

3.1 The Principle of Obtained Team SSA

Shared SA is dependent on the SA of the individuals involved [5], but it is more complicated, in that often not all team members in a given situation need to be aware of all the same information [14]. In many situations, individuals in a team possess the specialized knowledge to help them perform particular tasks, and they rely on others to do their jobs properly. Although each team member needs to monitor and has good SA of the information that is relevant to his or her job, the overlap of the individual SA between two team members is important only when they have the same task requirements. Saner et al. [8] provided that team SSA can be evaluated by directly comparing any two given team members with regard to the similarity of their understanding of the situation elements that are relevant to both of them (see Fig. 2).

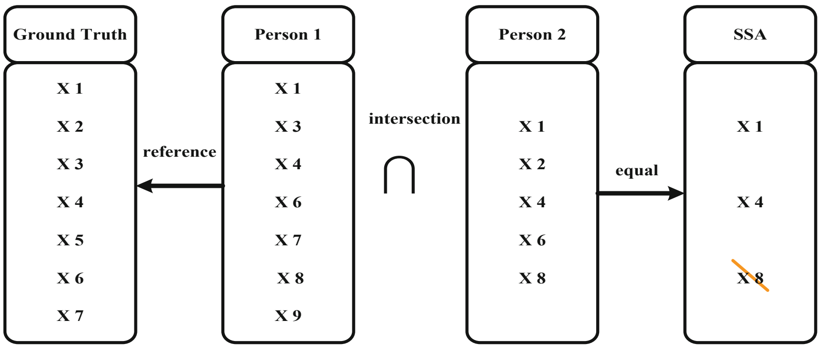


Fig. 2. The principle or theoretical basis of SSA

3.2 Participants

A total of 15 male operators in digital NPPs participated in the simulator experiments. They were divided into 5 groups, each group included 3 operators, who were work as primary loop operator (RO1), secondary loop operator (RO2) and team leader/coordinator (US) (not including shift supervisor/safety engineer (SS/STA) due to some objective reasons such as limited manpower). The main role and responsibility of RO1 is the operating/manipulation of the primary loop system. RO2 is mainly responsible for the monitoring and control of the secondary loop system to coordinate the operation of primary loop. The coordinator is mainly responsible for the understanding of system state, coordinating and monitoring the two operators. Their ages are from 24 to 35 (Mean = 28.6, SD = 2.34). They have the experiences from 6 months and 5 years for operating digital control systems in NPPs. All subjects' naked eyes or corrected visions are normal to meet the operating requirement.

3.3 Experiment Platform

The full-size simulator of a digital NPP was used to carry out experiments, which is a digital control system (DCS). The overall design features of a computer-based advanced MCR in a NPP. There are four large screens in front of the MCR, which are used to display the overall overview of the picture of an NPP and main parameters of primary loop, secondary loop, electrical system, P-T diagram separately. There are 4 computer-based workstations in the center of MCR, which correspond to the operator of primary loop (RO1), operator of secondary loop (RO2), team leader/coordinator (US), shift supervisor/safety engineer (SS/STA). Each computer-based workstation has 5–6 screens to display partly related systems, alarms, procedures, control screens etc.

3.4 Experiment Scenarios and SA Scales

In 2013, our team and the teachers of training center of a digital NPP discussed how to select some representative and suitable accident scenarios for implementing experiments. Finally, the four typical events/accident scenarios were selected, including steam generator tube rupture (SGTR) accident, main steam line break (MSLB), loss of off-site power (LOOP), and small break loss-of-coolant accident (SBLOCA). In order to measure operator's SA level in digital NPPs, a measure method of SA was developed based on SAGA. Compared with SAGAT, the main difference is the measurement scale. As for the above 4 risk events, the measurement tables/questionnaires relevant to SA of operator were developed by human factor experts (4 persons), teachers of simulation training (2 persons) and senior operators (6 persons). The developed scales of SA measurement included three aspects: demographics questionnaire, SA measurement questionnaire, and state level measurement of PSFs questionnaire. An example of rating scale of measurement of SA level in the sub-task —"isolation of the failed SG (Steam generator)" in SGTR simulation experiment was shown in the reference [15, 16]. Demographics questionnaire was used to collect data on their age, gender, and work experience in digital control system etc. SA measurement questionnaire was mainly used to identify the SA level of operators about system/unit state.

3.5 Procedure

In order to measure the SSA of a team, there are three main experiment steps as follows:

(1) Experiment training. After determining the selected staff who take part in these experiments, the two-hour experiment training course/session is carried out. (2) Pre-experiment. After the training, there was a pre-test to verify whether they had an adequate understanding of the procedure of experiment and to examine possible problems in the experimental design process and to test the effectiveness and credibility of the experimental measurement tool. (3) Formal experiment. The experimental organizer will explain the experiment and give illustrations to the operators. Then the teacher of simulator training center announced that the experiment starts. The experiment scenario is determined by the teacher randomly. During each risk scenario, the

simulator system was randomly frozen and the displays were hidden. Then individual SA questionnaires are distributed to them to fill in. The operator's SA level was determined through the questionnaire (namely rating scale) based on the operator's understanding level on the state of units.

4 Results

4.1 Individual SA

In order to measure the level of operator's SA. We assume that the importance of all the question entries of the questionnaire is the same, that is, all SAGAT questions are considered equally important. The individual SA level is measured and calculated at each pause point of simulator. The individual's SA level is that the total correct number of answers to all question items divided by the total number of question items of questionnaire. The formula is as follows:

$$ISA_{ik} = \frac{C_i}{T_k} \tag{1}$$

Where ISA_{ik} indicates the level of SA of the i operator in the k experimental scenario. C_i indicates the total correct number of answers in the questionnaire. T_k indicates the total number of entries in the k experimental scenario. The data obtained is partly shown in the reference [15, 16].

Based on the data obtained from the simulator experiments, we can calculate the level of SA of RO1, RO2 and US, respectively. We assume that the impact of different risk scenarios on individual SA is negligible so that the mean value of individual SA with different risk scenarios can be obtained as follows:

$$ISA_j = \frac{\sum_{i=1}^{n} ISA_{ij}}{n} \tag{2}$$

Where, ISA_j indicates the average level of SA of the j kind of operator, $j = 1, 2, 3$, which corresponds to RO1, RO2 and US, respectively.

On the basis of the obtained data and the formulas (1), (2), the average level of SA of RO1 was 0.7785, the average level of SA of RO2 was 0.7782, and the average level of SA of US was 0.8554.

It can be seen from the above data that the SA level of RO1 is about the same to that of RO2, but the level of SA of US is higher than that of RO1 and RO2. The difference results from the differences of operator's role and task characteristic. When an incident/accident occurs, RO1 and RO2 implement incident/accident handling according to the state-oriented procedure (SOP), they mainly pay attention to the operation of primary loop (RO1) and secondary loop (RO2). There is little time to consider the deep questions about changes in system state parameters (for example, why and how to change). However, US is mainly responsible for collecting more information to understand the state of the unit and monitor/supervise the important operation of the primary loop by RO1 and operation of the secondary loop by RO2. Therefore, the SA level of US is higher than that of RO1 and RO2.

Furthermore, the mean level of SA of all operators in the whole experiment is shown as follows:

$$ISA_M = \frac{\sum_{i=1}^{m} ISA_i}{m} \tag{3}$$

Where m represents the number of operators participating in simulator experiments, and ISA_M represents the average level of SA of all operators.

The average level of SA of all operators is 0.8041. From the perspective of operator's performance, it means the average level of SA of all operator in the entire digitized NPP is good because their average performance is more than 0.8. But from the perspective of human reliability analysis (HRA), the operator's SA level is relatively low, so managers should continuously make the state of PSFs better to improve the SA level and safety level of NPPs in the future.

In addition, as for different typical risk scenarios including SGTR, SBLOCA, MSLB and LOOP, the levels of SA of operators with different roles are shown in Table 1.

Table 1. SA levels for different types of operators in different risk scenarios

Risk scenarios	Operators			Mean values
	RO1	RO2	US	
Steam generator tube rupture (SGTR)	0.7646	0.8348	0.8889	0.8295
Small break loss-of-coolant accident (SBLOCA)	0.6256	0.7138	0.7908	0.7101
Main steam line break (MSLB)	0.9040	0.8491	0.8762	0.8765
Loss of off-site power (LOOP)	0.6993	0.5898	0.7968	0.6953

It can be seen from Table 1 that the average SA level of the operator is most high in MSLB, the second one is SGTR, which is mainly because the accident symptoms of MSLB are more obvious and easy to identify than other accidents, but the symptoms of SBLOCA are not obvious, it is relatively difficult to identify and diagnose. The SA level of LOOP is relatively low, the causes are probably the complexity of the operation and less simulator training practice. Furthermore, as for different operator, the highest level of SA of RO1 is the MSLB, the lowest is SBLOCA. The highest level of SA of RO2 is BSLB, the lowest is LOOP. The highest SA level of US is SGTR, the lowest is SBLOCA, from which we can see that the operator's SA level is different in different risk scenarios because the complexity of the task and different state level of PSFs. As for the overall SA level in all experiments, the SA levels of US are higher than the ones of RO1 and RO2 except for MSLB. The SA levels of different operators are similar in MSLB, it is probably attributed to the workload of RO1 and RO2 is lower than other accidents and their training is more adequate.

4.2 Shared Situation Awareness

We can see shared SA is the common understanding of team members on the state of the system according to the definition of shared SA. Thus, it can be assumed that the shared SA is the common correct question items in the same questionnaire for a team. Because there are two or more team members to share their SA, so the shared SA level of a team has many possibilities. Therefore, we use the following three formulas to describe SSA level:

1. The minimum value of shared SA:

$$SSA_{Min} = \frac{X}{Y} \tag{4}$$

Where X indicates the total number of common correct question items, Y question indicates the total number of all question items in a questionnaire, SSA_{Min} indicates the common correct SA for the entire team members (or the proportion of the total number of common correct question items for the total number of all question items in a questionnaire).

2. The most probable value of shared SA. The above is the minimum value of the SSA level, but the fact is that one or some operators' answers are correct to the same question item, but the other is not correct, so there are some answers are partly correct for all operators. Regarding this situation, we think that the correct rate of a question item is the number of operators with correct answers for the same question item divided by the total number of team members. Therefore, the most probable value of SSA can be expressed as:

$$SSA_{MPV} = \frac{\frac{N_c}{n}}{T_k} = \frac{1}{n}\frac{N_c}{T_k} \tag{5}$$

Where SSA_{MPV} indicates the most likely value of the shared SA, N_c indicates the number of operators with correct answers for the same question item, n indicates the number of operators in a team.

3. The maximum value of SSA. There are communication and cooperation in a team, so they can correct the mistakes of other members through discussion etc. Therefore, if we consider the recovery of SA error, we can believe that if the item is correct only by one operator, then we think that the item will be correct throughout the entire team members. It can be expressed by the following formula:

$$SSA_{Max} = \frac{X + Z}{Y} \tag{6}$$

Where X indicates the total number of common correct question items for all the team members, Z indicates the number of partly correct items in a questionnaire. The minimum and maximum value of SSA can be used to determine their uncertainty boundaries.

Based on the data obtained by the simulator experiment, we can use the formulas (4)–(6) to calculate the most probable value, the maximum value and the minimum value of the shared SA of each team. For example, in the first SGTR experimental scenario, there were 31 question entries in the questionnaire, the wrong answers of RO1 are the items 6, 7, 9, 14, 15, 16, 17, 19, 21, the wrong answers of RO2 are 9, 12, 24, and the wrong answers of US are 6, 9, 14, 16, 17, 19, 21. Thus the minimum value can be obtained according to the formula (4):

$SSA_{Min} = (31 - 11)/31 = 0.6451$

According to the formula (5), the most probable value is:

$SSA_{MPV} = 0.6451 + (1+2+2+1+2+1+1+1+1+2)/3/31 = 0.7956$

According to the formula (6), the maximum value is:

$SSA_{Max} = (20+10)/31 = 0.9677$

The specific results are shown in Table 2. The average value of all SSA in simulator experiments can be calculated as shown in Table 2, including the minimum, maximum and most probable value.

Table 2. The related data of SSA level

Order	Minimum value	Most probable value	Maximum value
1 (SGTR)	0.6451	0.7956	0.9677
2 (SGTR)	0.8571	0.9286	1.000
3 (SGTR)	0.3333	0.6889	1.000
4 (SGTR)	0.7073	0.8049	0.9756
5 (SGTR)	0.7576	0.8182	0.9394
6 (SGTR)	0.8636	0.9394	1.0000
7 (MSLB)	0.8387	0.9247	1.0000
8 (MSLB)	0.7895	0.8772	1.0000
9 (MSLB)	0.7895	0.8246	0.8421
10 (MSLB)	0.5769	0.7436	0.9231
11 (MSLB)	0.7561	0.8933	1.000
12 (LOOP)	0.5517	0.6551	0.8965
13 (LOOP)	0.5789	0.7017	0.9808
14 (LOOP)	0.4872	0.7607	0.9744
15 (SBLOCA)	0.2800	0.6000	0.9600
16 (SBLOCA)	0.6585	0.7967	0.9756
Mean value	0.6544	0.7971	0.9647

Similarly, as for the state level of PSF impacting operator's SA level, we can also obtain the average rank/value of state level of different PSF based on the operators' evaluation results of the PSFs in these experiments, which are shown in Table 2. The average values of SSA levels for different risk scenarios are also shown in Table 3. According to the data in Table 2, the mean value of SSA level of all teams is 0.7971. If from the perspective of performance, the performance level of the operator is good, but from the perspective of HRA, the operator's SA level needs to be improved. However, from the viewpoint of the maximum value of SSA level, and the value is 0.9647, it belongs to very good SSA level. We think that if the team have a good communication and cooperation and correct each other's mistakes in time during an accident handling, then it is most likely to achieve the maximum value of SSA level. The best SSA level is the one of the teams in the MSLB accident according to Table 3, the second is the SGTR accident, and the worst is the SBLOCA. It shows that the symptom of accident is more obvious, the operator's workload is smaller, thus they have higher SSA.

Table 3. Comparison of SSA levels in different risk scenarios

Risk scenario	Mean value		
	Minimum value	Most probable value	Maximum value
SGTR	0.6940	0.8293	0.9805
SBLOCA	0.4693	0.69835	0.9678
MSLB	0.7501	0.8527	0.9530
LOOP	0.5393	0.7058	0.9506

5 Conclusion and Discussion

In order to identify operator's SA level and team SSA level, the SAGAT technique is used to measure the SA level of the operator and the team SSA in simulator experiments. The following conclusions are obtained:

1. In the same accident scenario, the different operators have different SA levels due to the different tasks and roles of operators. Generally speaking, RO1 and RO2 mainly operated the primary loop and secondary loop according to the procedures, respectively, US is responsible for grasping/understanding the state of the unit and supervising the important operation of RO1 and RO2. Therefore, the SA levels of RO1 and RO2 are lower than the one of US.
2. As for different accident scenarios, their occurrence mechanisms are different, the contextual environment impacting operator SA is different, so their SA level is different. The results of simulator experiments revealed that if the symptoms of the accident are more obvious, then operators have a higher SA level, for example, the operator's SA levels in MSLB and SGTR are significantly higher than the one in SBLOCA.
3. It is somewhat difficult to measure SSA level in a team, but the SSA level is expressed by the established calculation method and the most probable SSA is used to represent the SSA level of operator, which is more objective and close to the real value of SSA level. Generally speaking, If the operator's SA level is higher, then the SSA level of the team is also higher, so it is necessary to strengthen/enhance the individual SA level to improve the SSA level. However, although the individual SA

level is low, sometimes the SSA level of the team is high, which means that individual SA error can be corrected by good communication and discussion to enhance the team SSA level.

4. The experimental results show that the average value of SA level of the operator is 0.8041 and the average value of SSA level is 0.7971, which means that both the individual SA and team SSA level are relatively good from the perspective of human performance. However, the results of individual SA and team SSA are not better from the viewpoint of HRA. Therefore, we should try our best to improve the state level of PSFs to improve SA level.

The useful results on individual SA and team SSA are obtained by simulator experiments, but the number of simulator experiments is limited due to various objective reasons, so it is difficult to obtain a large amount of experimental data to validate the reliability of the results. Furthermore, it is also difficult to use the limited experimental data to establish reliable function expressions or quantitative assessment model between individual SA/SSA and PSFs. In addition, the importance of each question in each questionnaire is assumed as the same, but there may be some deviations in practice, so these limitations require be overcome, the experimental design require further be refined as well as the collection of more data in the future studies.

Acknowledgments. The financial support by the National Natural Science Foundation of China (No. 51674145, 71771084), Postdoctoral Science Foundation of China (No. 2016M600633), Natural Science Foundation of Hunan Province (No. 2017JJ2222), Humanities and Social Science Projects of Ministry of Education (No. 11YJC630207) and Social Science Projects of Hunan Province (No. 16YBA314) are gratefully acknowledged. We would like to express gratitude to the staff in a Chinese nuclear power plant (LingAo-II) for the interviews and investigations, which facilitate the research included in this paper.

References

1. Salas, E., Fiore, S.M.: Team Cognition: Understanding the Factors that Drive Process and Performance. American Psychological Association, Washington DC (2004)
2. Banbury, S., Tremblay, S.: A Cognitive Approach to Situation Awareness: Theory and Application. Routledge, London (2004)
3. Kim, S.K., Park, J.Y., Byun, S.N.: Crew resource management training for improving team performance of operators in Korean advanced nuclear power plant. In: IEEE International Conference on Industrial Engineering and Engineering Management, pp. 2055–2059 (2009)
4. Lin, C.J., Yenn, T.C., Yang, C.W.: Evaluation of operators' performance for automation design in the fully digital control room of nuclear power plants. Hum. Factors Ergon. Manuf. Ind. **20**, 10–23 (2010)
5. Salas, E., Prince, C., Baker, D.P., Shrestha, L.: Situation awareness in team performance: Implications for measurement and training. Hum. Factors **37**, 123–136 (1995)
6. Cooke, N.J., Gorman, J.C., Duran, J.L., Taylor, A.R.: Team cognition in experienced command-and-control teams. J. Exp. Psychol.: Appl. **13**(3), 146–157 (2007)
7. Gorman, J.C., Cooke, N.J., Winner, J.L.: Measuring team situation awareness in decentralized command and control environments. Ergonomics **49**(12–13), 1312–1325 (2006)

8. Saner, L.D., Bolstal, C.A., Gonzalez, C., Cuevas, H.M.: Measuring and predicting shared situation awareness in teams. J. Cogn. Eng. Decis. Mak. **3**, 280–308 (2009)
9. Salas, E., Cooke, N.J., Rosen, M.A.: On teams, teamwork, and team performance: discoveries and developments. Hum. Factors **50**, 540–547 (2008)
10. Endsley, M.R., Bolte, B., Jones, D.G.: Designing for Situation Awareness: An Approach to Human-Centered Design. Taylor & Francis, London (2003)
11. Lloyd, M., Alston, A.: Shared awareness and agile mission groups (2003). http://www.dodccrp.org/8thICCRTS/Pres/plenary/3_0830Lloyd.pdf
12. Nofi, A.: Defining and measuring situation awareness. Report CRM D0002895.A1, Center for Naval Analysis (2001)
13. Endsley, M.R.: Direct measurement of situation awareness: validity and use of SAGAT. In: Endsley, M.R., Garland, D.J. (eds.) Situation Awareness Analysis and Measurement, pp. 3–32. Erlbaum, Mahwah (2000)
14. Jones, D.G., Endsley, M.R.: Measurement of shared SA in teams: initial investigation. Report No. SATech-02-05, Marrietta, GA (2000)
15. Li P.-C., Zhang L., Dai L.-C., et al.: A validation research on fuzzy logic-AHP-based assessment method of operator's situation awareness reliability. Saf. Sci. (2018). https://www.sciencedirect.com/science/article/pii/S0925753518315741
16. Li, P.C., Zhang, L., Dai, L.C., Li, X.F.: Study on operator's SA reliability in digital NPPs. Part 2: a quantitative assessment method. Ann. Nucl. Energy **109**, 82–91 (2017)

Method for Crew Human Reliability Analysis Based on CREAM

Xiaoxia Wang, Wenjin Zhang, Xing Pan(✉), and Tun Liu

School of Reliability and System Engineering, Beihang University, Beijing, China
langshixuan@foxmail.com, {zwjok, panxing}@buaa.edu.cn, liu_tun@163.com

Abstract. The reliability of human being has been paid increasing attention by engineers in recent years, which significantly increases the success of aerospace mission. Many Human Reliability Analysis methods, which are referred to HRA have been proposed to quantify Human Error Probability (namely HEP) in the past few decades. This paper takes crew human reliability into account and present its concept. Then, an assessment method for crew human reliability is proposed basing on Cognitive Reliability and Error Analysis Method, referred to as CREAM, and this assessment method mainly focuses on how to make accurate quantification model on crew errors. Finally, an example about the crew of fueling process in space launch is evaluated to show the usage of the model.

Keywords: Human reliability analysis · Crew human reliability analysis · Crew · CREAM

1 Introduction

With the progress of aerospace industry, the reliability of human being, which significantly affects whether aerospace mission could success, has been greatly valued by engineers. HRA (Human Reliability Analysis) method, which originated in the 1950s, is the extension and development of human engineering. It applies information about human characteristics and behaviors to the design of objects, facilities and environments, and has gradually formed a relatively independent emerging discipline, widely being used in aerospace, nuclear power, ships and other fields.

Usually, human behaviors that cause accidents are called human errors [1]. In the past decades, many HRA methods have been proposed to quantify Human Error Probability (HEP). In the field of reliability engineering, the term used to describe human performance is Human Reliability (HR). It refers to the probability that a person will successfully complete a task at a specific stage of system operation, at a specific time, in a specified operating environment [2]. For the first time in his study, Williams [3] suggests that the rate of human error varies according to working environment, at 0.01 on the ground and 0.02 in the air, which opens the door to the quantification of human error. The HRA methods can be divided into three generation after a period of development.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 225–234, 2020.
https://doi.org/10.1007/978-3-030-20037-4_20

Great majority existing HRA methods focus on the reliability of a single person but neglect the analysis of the crew's overall reliability and its effect on individual reliability. In some methods, the factors of the crew are considered as a factor affecting the reliability of the individual. However, the execution process of most tasks requires a crew of multiple people to cooperate rather than a single person to complete alone.

Sasoua and Reason [4] point out that there are two points to be considered in crew missteps, one is the process of error generation, the other is the process of error correction. They classify the types of error correction into three categories: failure to discover, failure to instruct, failure to correct. Rouse et al. [5] consider the nature of crew performance in complex systems. It is found that the construction of psychological model may provide a basis for the principle explanation of crew performance, and it is also a way to improve crew performance. Sasoua et al. [6] proposed a task crew behavior simulation system that simulates the effects of operators' behaviors in unexpected circumstances, namely the concept of SYBORG. Shu et al. [7] presents a TBNM, which considers the cognitive processes of the crew and the interaction between operators and environment so as to simulate and analyze the responses of crews to dynamic time and situational sensitivity. Chang [8–10] analyzes the IDAC model for HRA in crew context, which includes a crew model consisting of operators of decision makers, action participants, consultants. In addition, this model is used to predict the response of operators in accidents for probabilistic risk assessment. environment, the behavior of each individual operator is modeled by a cognitive model that is influenced by many of the performance factors that are explicitly simulated.

From the above analysis, it can be seen that the analysis of the crew factors in the existing models is relatively simple. Therefore, we draw lessons from the previous research results and carry out the crew human factor reliability analysis.

The method of CREAM has been viewed as one of the most significant methods in the second-generation HRA methods. Basing of the method of CREAM, a new evaluation method from the prospective of the crew is proposed in this paper to evaluate the overall reliability of the crews in the task by evaluating Crew Performance Influence Factors (CPIFs). In addition, because of the fact that under different division of work and cooperation, the same factor can have different influence on different crews, this paper also considers the types of crew in the course of calculation to make the results more accurate. We also evaluate the crew of fueling process in space launch under the guidance of the method proposed in this paper, thus, the use of the model is illustrated. At the end of this paper, a summary is given.

2 Model Preparation

This chapter introduces the determination of CPIFs and the classification of the crew. It combines the crew with the CREAM method to prepare for the next chapter as well.

2.1 The Definition of CPIF

In the process of completing tasks, human reliability is mainly affected by two aspects: context and human nature. To make a quantitative assessment of it, Hollnagel [11] proposes eight Performance Influence Factors(PIFs) for the context in CREAM. In the IDAC model presented by Chang [10], six factors related to the crew are summarized and are showed as follows:

1. Crew Cohesiveness: it is also known as "crew morale" or "crew mood", is a symbol of crew integrity. It includes crew unity, crew harmony and how crew members get along with each other.
2. Crew Coordination: it refers to the effectiveness of a crew as a whole in terms of time, space, and responsibility, and control allocation also refers to the level of harmony of each member's contribution to the crew's tasks.
3. Communication Convenience: when the crew members are in different positions to perform tasks, the convenience of communication between members has an important impact on the reliability of the crew. And communication allows the crew to understand the common situation.
4. Communication Effect: it refers to the consistency of the message received by the recipient and the message sent by the dispatcher. Therefore, poor communication effect may be caused by communication equipment failure and signal interference.
5. Crew Composition: it involves size and uniformity, the heterogeneity of the staff, and the complementarity and redundancy required to complete the tasks. The number of crew members is usually determined by factors such as the complexity of the mission. The general task crew includes execution, supervision, and backup.
6. Leadership: it refers to the leader determines the direction of the crew and establishes a harmonious relationship with the subordinates so that they can perform their duties in the work and the ability to overcome difficulties with them.

It is reasonable to define crew cohesiveness, crew coordination, communication convenience, communication effect, crew composition and leadership as six of significance on crew human reliability, namely Crew Performance Influence Factors (CPIFs).

2.2 The Classification of the Crew

In the process of completing a task, the cooperation mode among the crew members is often different according to the needs of the task, yet the most important CPIFs that affect crew reliability are uncertain. Therefore, we' classification of the crews can not only simplify the calculation process, but also make it a targeted analysis.

In the light of the two CPIFs that are of the most importance to the reliability of the crew, the crew is divided into three categories: serial crew, parallel crew and circular crew. Of course, there are a large number of mixed crews in actual production activities, but in this paper, we only consider these three simple crews.

1. Serial Crew
   We define the crews that require the highest level of cohesion and cooperation as serial crews. Serial crew is the most common crew in production life, such as the assembly line. This kind of crew is a single line, operators are often in the form of pipeline and are arranged according to the production line, as shown in Fig. 1. Operators are arranged in the order that one person performs the task and passes it

to the next person. This type of crew is also fragile and any one of the problems in the middle usually leads to the paralysis of the entire crew.

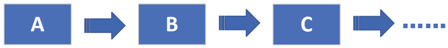


Fig. 1. Serial crew diagram

2. Parallel Crew
   The crew that has the highest requirements on the communication convenience and crew composition is defined as parallel crew. Yet operators of the parallel crew do not have a clear interaction with each other with each operator independently completing the assigned tasks, as shown in Fig. 2. The greatest feature of parallel crew is that individual operators are independent in one part of the task, so that the dependence of such crews on communication convenience and crew composition is the highest. These features increase the anti-jamming capability and reliability of this type of crew.

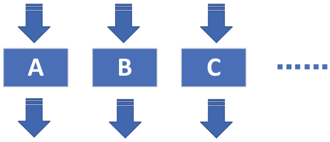


Fig. 2. Parallel crew diagram

3. Circular Crew
   We define the crew that requires the highest communication effect and leadership of the crew leader as circular crew. In circular crew, the crew members will not be too many. There is often a crew leader in charge of the overall management in circular crew as shown in Fig. 3. The biggest characteristic of the circle group is that the task is not independently determined, but under the unified command of the leader. And operators are required to follow the command. In the circular crew, the interaction of the crew member increases and the commanders play an important role in the crew.

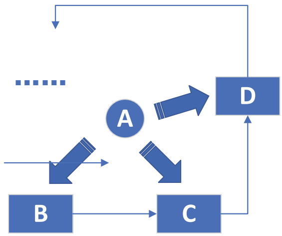


Fig. 3. Circular crew diagram

2.3 Crew Control Mode and Probability Interval

The main focus of CREAM is to make accurate quantification on human errors. It is feasible to combine the evaluation of CPIF with CREAM. The impact of six CPIFs on crew reliability were evaluated and divided into three levels: improved, not significant and reduced, as shown in Table 1. The total number of reduced CPIFs is denoted as Σ Reduced and the total number of improved CPIFs is denoted as Σ Improved. Crew human error probability (CHEP) is finally determined by $CHEP_0$ and the sum of improved, not significant and reduced. $CHEP_0$ is assumed to have the same value as HEP_0 with the value of 7.07×10^{-3}.

Table 1. The evaluations and influence of six CPIFs

CPIF	Evaluation	Influence		
		Improved	Not significance	Reduced
Crew cohesiveness	Good	✓		
	Medium		✓	
	Bad			✓
Crew coordination	Good	✓		
	Medium		✓	
	Bad			✓
Communication convenience	Good	✓		
	Medium		✓	
	Bad			✓
Communication effect	Good	✓		
	Medium		✓	
	Bad			✓
Crew composition	Good	✓		
	Medium		✓	
	Bad			✓
Leadership	Good	✓		
	Medium		✓	
	Bad			✓

In CREAM, four characteristic control modes, which are scrambled, opportunistic, tactical and strategic, are defined in line with these an operator or a crew to control this situation. This paper also divided the crew control mode into (1) Strategic (2) Tactical (3) Opportunistic (4) Scrambled four types, as shown in Fig. 4, the abscissa x represents the sum of reduced and the ordinate y represents the sum of improved. The region where (x, y) locates means the model is strategic, tactical, opportunistic and scramble respectively.

The probability interval of individual control mode is calculated in the nuclear field in CREAM, and the probability interval may be different among the crew. The probabilistic interval of the crew is assumed to be the same as that that of the original method.

CHEP interval Strategic (0.00005 ~ 0.00100), Tactical (0.00100 ~ 0.10000), Opportunistic (0.01000 ~ 0.50000), Scrambled (0.10000 ~ 1.00000).

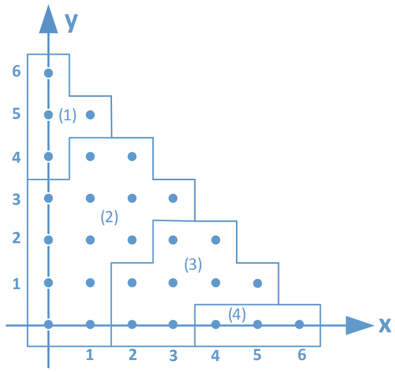


Fig. 4. Relationship between crew control mode and CPIF

3 The Quantification of the Model

There are several steps to evaluate the human reliability of a crew.

Firstly, according to the method in 2.2, a crew is evaluated to determine its type.

Then the six CPIFs that affect the crew's reliability are rated by professionals, and the fraction range is 1–10. If a CPIF's score ranges from [1, 4], this CPIF of the crew has poor performance and has a negative impact on the crew's overall reliability. If a CPIF's score ranges from [5, 7], this CPIF of the crew has general performance and has a not significant i on the crew's overall reliability. If a CPIF's score ranges from [8, 10], this CPIF of the crew has great performance and has an improved impact on the crew's overall reliability.

There are six CPIFs in the crew's evaluation method, max (Σ Improved) = 6 and max (Σ Reduced) = 6. If Σ Improved reach the maximum and Σ Reduced reach the minimum, then the crew has highest reliability and CHEP is at its minimum. Otherwise, if the minimum Σ Improved and the maximum Σ Reduced are reached at the same time, CHEP is then at its maximum.

Therefore, the authors have:

$$\begin{cases} CHEP_{min} = CHEP_0 \exp(\lambda) \\ CHEP_{max} = CHEP_0 \exp(-\lambda) \end{cases} \tag{1}$$

$CHEP_0$ and λ are two constant values. It can be found that $CHEP_{min} = 0.00005$ and $CHEP_{max} = 1$ from Sect. 2.3. So according to Eq. (1), we have: $CHEP_0 = 7.07 \times 10^{-3}$, $\lambda = -4.9517$. Thus, CHEP probability calculation formula is as shows:

$$CHEP = 7.07 \times 10^{-3} \exp\left[-4.9517\left(\frac{uIm1 + uIm2 + \cdots + uImy}{6} - \frac{uRe1 + uRe2 + \cdots + uRex}{6}\right)\right] \quad (2)$$

According to Eq. (2), CHEP is determined by the improvement and reduction of six CPIFs. Different types of the crew are affected by different CPIF levels, so we define a variable u to correct the number of improved and reduced. For example, in a parallel crew, communication convenience and crew composition are the two most significant CPIFs and the importance of other four CPIFs is relatively small. When evaluating a parallel crew, if the scores of crew cohesiveness, crew coordination, communication effect, leadership these four factors are excellent and the scores of communication convenience and crew composition are poor, then the sum of Improved equals four and the sum of Reduced equals two. From Fig. 4, it can be seen that the crew is evaluated as tactical. However, the evaluation of the two most influential CPIFs is poor, it is obviously unreasonable for the crew to be rated as a tactical crew (Fig. 5).

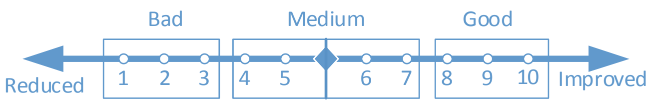


Fig. 5. Relationship between CPIF's score and evaluation

Therefore, it is very important to define the variable u to correct the evaluation. There are still some difficulties with the value of u, which we temporarily set as the values shown in Table 2.

Table 2. The value of variable u

Important CPIFs	Secondary importance CPIFs
1.0	0.3
1.2	0.5
1.4	0.7
1.6	0.9

The value of u of important CPIFs and secondary important CPIFs can be determined by professionals. This not only guarantees the rationality of the value of variable u, but also increases the weight of important CPIFs and reduces the weight of secondary important CPIFs. In addition, in the evaluation of CPIF with a Medium score, those with scores greater than 5 are equal to 0.5 improved to calculate, and those with scores less than or equal to 5 are equal to 0.5 reduced.

Finally, we get Eq. (2). In Sect. 4, an example will be given to illustrate the reliability evaluation method of the crew presented in this paper.

4 Case Analysis

Rocket launching is a complex task, including rocket fueling and other aspects. The rocket filling process consists of multiple steps such as replacement, pre-cooling, filling and so on. In the fueling processes of aircraft launching, to make the filling process more orderly and controllable, there is usually a commander who directs the entire crew. Thus, the fueling processes crew is a typical circular crew, as shown in Fig. 6.

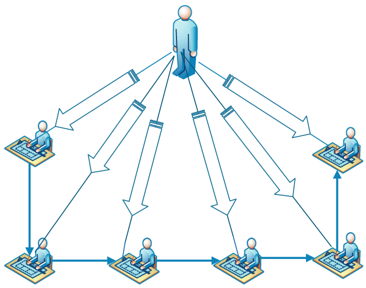


Fig. 6. The diagram of fueling processes crew

Assuming that there is a fueling processes crew, its six CPIF scores, as assessed by professional, are shown in Table 3. The value of u of important CPIFs is 1.2 and that of secondary important CPIFs is 0.9.

Table 3. The CPIF score table of an fueling processes crew

CPIF	Score	Evaluation	Influence		
			Improved	Not Significance	Reduced
Crew cohesiveness	5	Medium		✓	
Crew coordination	6	Medium		✓	
Communication convenience	3	Bad			✓
Communication effect	6	Medium		✓	
Crew composition	8	Good	✓		
Leadership	7	Medium		✓	

As can be seen from Table 3, the crew's assessment is as follows:

1. crew cohesiveness: 0.5 reduced
2. crew coordination: 0.5 improved
3. communication convenience: 1 reduced
4. communication effect: 0.5 improved
5. crew composition: 1 improved
6. leadership: 0.5 improved

Since communication effect and leadership are the two most important CPIFs for the fueling processes crew, their coefficient u are 1.2 and others' are 0.9. We can get Eq. (3).

$$\begin{array}{c} \Sigma \cdot \text{Improved} = 0.5 \times 0.9 + 0.5 \times 1.2 + 1 \times 0.9 + 0.5 \times 1.2 = 2.55 \\ \Sigma \cdot \text{Reduced} = 0.5 \times 0.9 + 1 \times 0.9 = 1.35 \end{array} \tag{3}$$

According to Eq. (2) and Eq. (3), we can calculate that the CHEP of the fueling processes crew is 7.07×10^{-3} exp (–4.9517 × 0.2) = 0.00263.

5 Summary

In this paper, we have made a complete presentation of the evaluation of Crew Human Reliability. In the proposed method, we summarize six CPIFs related to the crew from the literature. Then we take the types of the crew into account and classify the types of the crew according to the different emphases on the six CPIFs. An assessment method for crew human reliability has been proposed basing on CREAM. Finally, an example about the crew of fueling process in space launch has been evaluated to verify the method's effectiveness. However, there still are some areas where the method should be improved if it is to be applied.

The first aspect is the classification of the crew. The classification of the crew is based on different degrees of six CPIFs and has strong subjectivity and absoluteness. In addition, the applicability of different types of crews in different industries is still open to question. In the following studies, Bayesian Networks or Neural Networks will be used to classify crews.

The second aspect is the evaluation of the crew. There is still room for improvements in the accuracy of crew assessment, especially the determination of coefficients, which still lacks theoretical basis. Moreover, in the corresponding probability interval of crew control mode, the probability of cream method is used temporarily. The corresponding probability interval of control mode in cream method, which is obtained by statistics in the nuclear field, is definitely not universal. Therefore, the corresponding probability interval of control mode in this paper still needs to be improved.

In general, this paper classifies the crews and assesses the reliability of different crews from an overall point of view. Although there are many defects in this paper, it provides a good way of thinking for the study of human reliability. In the following study, we will further optimize the methods proposed in this paper to improve accuracy; and crew assessment combined with individual assessment may be a good direction.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grants No. 71571004.

References

1. Hollnagel, E.: Chapter 1 - The state of human reliability analysis. In: Hollnagel, E. (ed.) Cognitive Reliability and Error Analysis Method, pp. 1–21. Elsevier Science Ltd, Oxford (1998)
2. Balbir, S.: Human reliability with human factors. Microelectron. Reliab. **26**, 994–995 (1986)
3. Williams, H.L.: Reliability evaluation of the human component in man-machine systems. Electr. Manuf. **61**, (1958)
4. Sasou, K., Reason, J.: Team errors: definition and taxonomy. Reliab. Eng. Syst. Saf. **65**, 1–9 (1999)
5. Rouse, W.B., Cannon-Bowers, J.A., Salas, E.: The role of mental models in team performance in complex systems. IEEE Transact. Syst. Man Cybern. **22**, 1296–1308 (1992)
6. Sasou, K., Takano, K., Yoshimura, S.: Modeling of a team's decision-making process. Saf. Sci. **24**, 13–33 (1996)
7. Shu, Y., Furuta, K., Kondo, S.: Team performance modeling for HRA in dynamic situations. Reliab. Eng. Syst. Saf. **78**, 111–121 (2002)
8. Chang, Y.H.J., Mosleh, A.: Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Reliab. Eng. Syst. Saf. **92**, 997–1013 (2007)
9. Chang, Y.H.J., Mosleh, A.: Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. Reliab. Eng. Syst. Saf. **92**, 1014–1040 (2007)
10. Chang, Y.H.J., Mosleh, A.: Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 4: IDAC causal model of operator problem-solving response. Reliab. Eng. Syst. Saf. **92**, 1061–1075 (2007)
11. Hollnagel, E.: Chapter 6 - CREAM — A second generation HRA Method. In: Hollnagel, E. (ed.) Cognitive Reliability and Error Analysis Method, pp. 151–190. Elsevier Science Ltd, Oxford (1998)

Organizational Resilience Model in a Nuclear Power Plant

Dai Licao, Li Hu, Chen Jianhua, Lu Wenjie, and Li Pengcheng(✉)

Human Factor Institute, University of South China, Hengyang 421001, China
dailpc@sina.com

Abstract. Organizational resilience defines the capability of how a nuclear power plant organization manages disturbances, both outside and inside a nuclear power plant. The paper specifies the organizational resilience on the basis of analyzing the disturbances of a nuclear power plant. These disturbances include internal incidents, accidents and transients as well as those caused by external environment. The paper tries to reveal the mechanism of how organizational resilience functions on the safety of a nuclear power plant. The paper assumes the taking place of an accident (Small Loss of Coolant) and analyzes how the plant organization responds to the accident to illustrate the mechanism.

Keywords: Organizational resilience · Safety · Nuclear power plant

1 Introduction

A nuclear power plant is an energy-concentrated high risk industrial system. Accidents in a nuclear power plant may cause not only enormous casualties and huge economic losses but also great impact on communities and societies. Safe operation is of paramount importance.

Nuclear safety management theory advanced progressively since 1950s. At the beginning it focused on the reliability of system parts, then on human operators and operation organizations. Research work has been done from the equipment reliability analysis to a systematic Probabilistic Safety Assessment (PSA). All these techniques assume that risk happened in a form of "failure mode". Failure mode analysis takes a "post-accident" standpoint. It studies on in what failure modes how these assumed accidents go on and how root causes bring on the accident. Management measures then come in to prevent these root causes into play to avoid accidents. In 2011, a nuclear power plant accident happened in Fukushima, Japan. External events worked together with internal management failures making the nuclear radioactivity released into the environment. The accident implies that the industrial system is becoming more and more complicated and it is very difficult to predict the failure modes of possible risks. Many factors work together to make the accident and its consequence a reality. Accidents maybe caused by known risks and also those risks we even know their existence. Failure mode analysis seems to be obsolete. Safety management should assure the operation organization running the plant to maintain system safety margin despite any kind of failure modes. More and more researchers agree [1–4] that safety

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 235–243, 2020.
https://doi.org/10.1007/978-3-030-20037-4_21

research perspective should be "pre-accident" and "proactive" to study the inborn ability of organization, e.g. organizational resilience.

The term resilience is derived from the Latin word *resilio*, meaning "to jump back" [5]. When applied to social entities, resilience refers to "the ability to resist disorder" [6] to an organization's capacity " to continue its existence, or to remain more or less stable, in the face of surprise, either a deprivation of resources or a physical threat" [7]. From the perspective of safety management, resilience is defined as "the capacity of a social system to proactively adapt to and recover from disturbances that are perceived within the system to fall outside the range of normal and expected disturbances" [8]. For a nuclear power plant, disturbances include those events happening within a plant, the internal events and those outside, the external events. The organizational resilience of a nuclear power plant is the operating organization manages the plant to adapt to and recover from internal and external events. This paper is trying to build a model to analyze the resilience of a nuclear power plant on the basis of the above.

2 How Organizational Resilience Deal with the Internal and External Events

When the organization operates a nuclear power plant, it needs a constant adaptation to cope with the incidents within the system. This adaptation ability is intrinsic when the plant was designed and built. A nuclear power plant is resilient when it was born. When the plant is actually in operation, because of the constant conflict between the safety goal and production goal, the intrinsic resilience is always under pressure and gradually eroded. All these happen in a routine way and organization is not able to realize the erosion. Only when one day an accident takes place, the organization then takes measures to manage it.

The ability of an organization manages the system safety is determined by two factors, one is available resources, and another is control. Organizational resilience can be defined as the ability of employing available resources to manage the pressure from both the internal and external events. When events happen and plant deviate from normal situation, the system malfunctions at different levels. Different levels of management within an organization make use of available resources and manage to control the malfunctions. The plant is deemed safe when the functions of the systems are served. The mechanism of organizational resilience is defined as how organizational levels bounce back from the abnormalities and restore to normal. In other words, how different levels of organization use available resources to cope with the internal and external events to restore to normal become the critical part underlying the mechanism of organizational resilience.

On the basis of the above analysis, the five critical elements of organization dealing with events coming inside or outside a plant are Disruption events (E), System security goals (G), Control measures (M), Resources (R) and System Function Status (S) (as shown in Fig. 1).

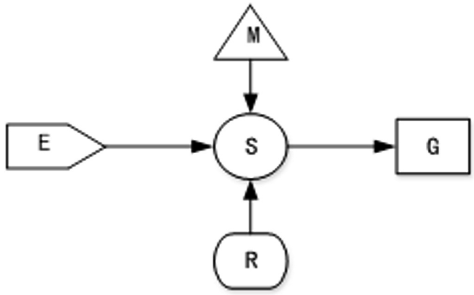


Fig. 1. Five elements of organizational resilience

Figure 2 is a more elaborate figure to illustrate the evolvement of the event scenario with five elements of organizational resilience. When events take place, the organization percepts and make some responses. When at t1, it has two paths. One is the organization uses Resources (R1) to Control (M1) to try to make System function status (S1) to reach a goal (G1). With the time goes on until tn, the plant finally ends up in a normal status.

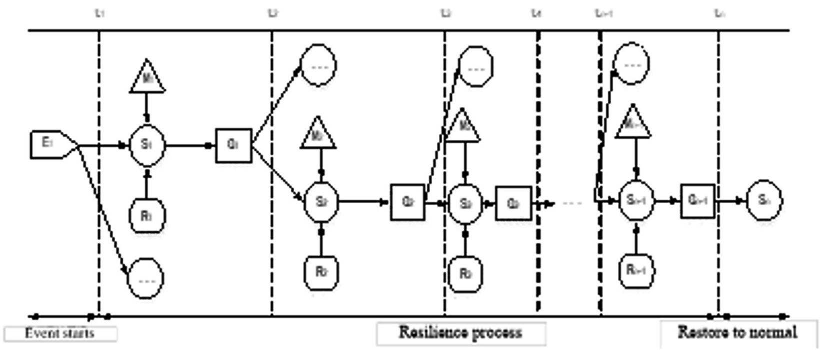


Fig. 2. Organizational resilience evolvement

3 Organizational Resilience Mechanism Simulation

3.1 Technique of Dynamic Bayesian Network

On the basis of the above analysis, combining with the Dynamic Bayesian Network theory, critical scenario to critical system function elements involved in a plant coping with surprising accidents and their corresponding control measures, resources and plant safety goal are considered as node variables. All the node variables dependent on the time are connected to form a dynamic network of a plant's organizational resilience managing surprising accidents.

Dynamic Bayesian Network is a system model describing time-dependent events [9–11]. In comparison with normal Bayesian Network, Dynamic Bayesian Network (DBN) can make the related information developed in the sequence of time and lower down the uncertainty of inference. DBN is appropriate to develop an event scenario and is good to construct and simulate organizational resilience mechanism.

DBN calculates the probability with the following equations:

1. Calculation of basic joint probabilities:

$$P(A,B) = p(A|B)p(B) \tag{1}$$

2. Bayesian equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

3. When information is from different resources, it is expressed by extended Bayes rule. P(Hi) is a prior probability that exists before any measuring status i. It is a posterior probability when status i is given e vector.

$$P\left(H_i|\overrightarrow{e}\right) = \frac{P\left(\overrightarrow{e}|\mathrm{H_i}\right)P(H_i)}{\sum_{i=1}^{s} P\left(\overrightarrow{e}|\mathrm{H_i}\right)P(H_i)} \tag{3}$$

4. Other network nodes' information is calculated on the basis of the given prior probability.

$$P(S_1, S_2, \ldots, S_n) = \prod_{1}^{n} P\left(S_i|P\left(\prod S_i\right)\right) \tag{4}$$

On the basis of the above methods, a DBN on organizational resilience working mechanism is constructed. Experts rated the node variables, their interrelationship and conditional probabilities. The joint probability equation is used to calculate the corresponding node probability and the status probability of next-time scenario is developed. Bayesian network software MSBNX is used to make the calculation and simulation.

3.2 Small Loss of Coolant Accident (SLOCA) Resilience Model

Scenario Definition. The scenario of small loss of coolant accident resilience is herein defined as a break of 2 cm^2 to 10 cm^2 on the primary circuit of a nuclear power plant. The break is caused probably by pipe rupture or leakage and failure of valves.

Evolvement of the Accident. Based on thermo-hydraulic calculation and accident analysis, the plant needs to make reactor trip system, auxiliary feed water system, reactor coolant system, and containment system work together to mitigate the accident.

SLOCA always involves plenty of scenarios and has diversified evolvement paths. In order to facilitate the calculation of the organizational resilience, the accident scenarios start from S1- reactor coolant pressure going down up to the shutdown of the reactor. Two event sequences are described below:

Sequence 1: reactor emergency shutdown succeeds __ second circuit succeeds to form heat sink __ pressurizer spray kicks in __ high pressure safety injection starts up __ containment spray starts up;

Sequence 2: If high pressure safety injection fails __ pressurizer safety valves forced open __ feed and bleed;

Table 1. Elements involved in SLOCA scenario

System function status (S)	System function goal (G)	Resources (R)	Control measures (M)
S1 reactor coolant system pressure going down	G1 reactivity control	R1 available automatic shutdown system	M1 automatic
S2 safety injection inflow is balanced with coolant outflow	G2 de-temperature and de-pressurize by reactor coolant system	R2 Available procedures; proper man-machine interface;	M2 operators suspend safety injection timely
S3 Primary circuit temperature and pressure going up	G3 Primary circuit pressure and temperature stabilized	R3 Available procedures; available pressurizer spray system;	M3 Operators use available procedures to start up pressurizer spray system
S4 Primary coolant pressure down to 10.78 MPa	G4 maintain pressurized vessel level and de-temperature and de-pressurize	R4 Available procedures	M4 operator observes the inflow of feed-bleed water
S5 hot shutdown of reactor			
S6 Second circuit heat sink fails	G5 de-temperature and de-pressurize the primary circuit	R6 available procedures; available feed pump and bleed pump; Proper man-machine interface	M6 operators use available procedures to feed and bleed the primary circuit
S7 shutdown of reactor			

According to these two sequences, SLOCA can be chronologically set into 7 function status (S) and 5 system safety goal (G), 6 available resources (R) and 6 control measures (M), as shown in Table 1. On the basis of Table 1, the accident evolvement process is formed.

SLOCA Resilience Model. Table 1 has listed the key scenarios while SLOCA accident happens. These scenarios are considered as the node variables of Bayesian net. These nodes are connected to form a dynamic Bayesian network of SLOCA accident in a nuclear power plant (Fig. 3).

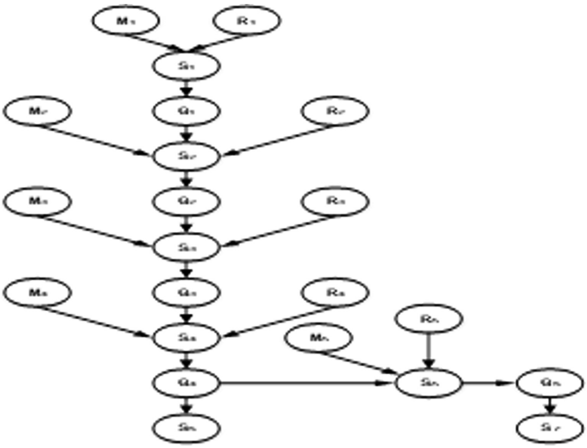


Fig. 3. Bayesian network of a nuclear power plant organization dealing with SLOCA accident

Node Probability Calculation. On the basis of Fig. 3, 22 operators in a reference pant were invited to fill in the Questionnaires to make assessment on the conditional probabilities of system function status (S), system safety goal (G), resources (R) and control measures (M) as shown in Table 2.

Table 2. Conditional probabilities of node variables

M_1	True	False	M_3	True	False
	0.65	0.35		0.75	0.25
R1	True	False	R3	True	False
	0.99	0.01		0.65	0.35
S1	M1 = True R1 = True	M1 = False R1 = False	S3	G2 = True M3 = True R3 = True	G2 = False M3 = False R3 = False
True	0.90	0.65	True	0.85	0.10
False	0.10	0.35	False	0.15	0.90
S1	M1 = True R1 = False	M1 = False R1 = True	S3	G2 = True M3 = True R3 = False	G2 = False M3 = False R3 = True
True	0.85	0.70	True	0.80	0.50
False	0.15	0.30	False	0.20	0.50
G1	S1 = True	S1 = False	S3	G2 = False M3 = True R3 = True	G2 = False M3 = True R3 = False
True	0.88	0.8	True	0.60	0.55
False	0.12	0.2	False	0.40	0.45

(*continued*)

Table 2. (*continued*)

M_1	True	False	M_3	True	False
M2	True	False	S2	G2 = True M3 = False R3 = True	G2 = True M3 = False R3 = False
	0.90	0.10	True	0.60	0.50
R2	True	False	False	0.40	0.50
	0.95	0.05	G3	S3 = True	S3 = False
S2	G1 = True M2 = True R2 = True	G1 = False M2 = False R2 = False	True	0.96	0.70
True	0.80	0.35	False	0.04	0.30
False	0.20	0.65	M4	True	False
S2	G1 = True M2 = True R2 = False	G1 = False M2 = False R2 = True		0.90	0.10
True	0.60	0.40	R4	True	False
False	0.40	0.60		0.99	0.01
S2	G1 = False M2 = True R2 = True	G1 = False M2 = True R2 = False	S4	G3 = True M4 = True R4 = True	G3 = False M4 = False R4 = False
True	0.60	0.40	True	0.70	0.20
False	0.40	0.60	False	0.30	0.80
S2	G1 = True M2 = False R2 = True	G1 = True M2 = False R2 = False	S4	G3 = True M4 = True R4 = False	G3 = False M4 = False R4 = True
True	0.75	0.70	True	0.55	0.40
False	0.25	0.30	False	0.45	0.60
G2	S2 = True	S2 = False	S4	G3 = False M4 = True R4 = True	G3 = False M4 = True R4 = False
True	0.97	0.30	True	0.60	0.35
False	0.03	0.70	False	0.40	0.65
S4	G3 = True M4 = False R4 = True	G3 = True M4 = False R4 = False	S6	G4 = False M6 = True R6 = True	G4 = False M6 = True R6 = False
True	0.58	0.30	True	0.80	0.70
False	0.42	0.70	False	0.20	0.30
G4	S4 = True	S4 = False	S6	G4 = True M6 = False R6 = True	G4 = True M6 = False R6 = False
True	0.70	0.30	True	0.75	0.70
False	0.30	0.70	False	0.25	0.30
S5	G4 = True	G4 = False	G5	S6 = True	S6 = False
True	0.90	0.20	True	0.60	0.40

(*continued*)

Table 2. (*continued*)

M_1	True	False	M_3	True	False
False	0.10	0.80	False	0.40	0.60
M6	True	False	S7	G5 = True	G5 = False
	0.90	0.10	True	0.85	0.20
R6	True	False	False	0.15	0.80
	0.60	0.40			
S6	G4 = True M6 = True R6 = True	G4 = False M6 = False R6 = False			
True	0.85	0.40			
False	0.15	0.60			
S6	G4 = True M6 = True R6 = False	G4 = False M6 = False R6 = True			
True	0.80	0.65			
False	0.20	0.35			

By means of Bayesian simulation software MSBNX, the status probabilities of node variables are calculated. The posterior probabilities of system function status (S) are listed in Table 3.

Table 3. Posterior probabilities of system function status after the accident of SLOCA in a nuclear power plant

System function status	Scenario status	Posterior probability
Coolant pressure going down S1	True/False	(82.9%, 17.1%)
Safety injection flow and bleed flow are balanced, reactor ΔTsat is proper S2	True/False	(75.7%, 24.3%)
Primary temperature and pressure up S3	True/False	(71.8%, 28.1%)
Primary coolant system lower than 10.78 MPa S4	True/False	(67.4%, 32.6%)
Hot shutdown S5	True/False	(59.8%, 40.2%)
Second circuit heat sink fails S6	True/False	(79.3%, 20.7%)
Hot shutdown S7	True/False	(56.3%, 43.7%)

4 Conclusions

From the above analysis, the following points maybe concluded:

1. When SLOCA takes place, the most probable paths are reactor coolant system pressure going down S1 and the normal heat sink of circuit two S6. The probabilities of them are respectively 82.9% and 79.3%.

2. There exists corresponding system safety goals, resources and control measures on every critical scenarios of SLOCA accident. These resources and control measures forming the most important part of organizational resilience when coping with certain kinds of accidents. The plant organization managing the accident has two final status probabilities, one is optimistic S5 (59.8%), another is pessimistic S7 (56.3%).
3. In the course of plant organization managing the accident, the failure probability of secondary circuit normally forming a heat sink (S 7) is 79.3%, subordinate to S1. It means that despite the functional goals, available resources and reasonable control measures, the probability of going to be pessimistic remains higher. It does not mean that resources and control measures have no influence upon the evolvement of the accident. A comparison between S7 and S5 shows that the organizational emergency response raises the probability of leading nuclear power plant to a safe status.

References

1. Zhongtang, W., Guohang, C.: Fukushima Accident [M]. Atomic Energy Press, Beijing (2014)
2. Adger, W.N.: Social and ecological resilience: are they related? [J]. Prog. Hum. Geogr. **24** (3), 347–364 (2000)
3. Keck, M., Sakdapolrak, P.: What is social resilience? lessons learned and ways forward[J]. Erdkunde **67**(1), 5–18 (2013)
4. Pfefferbaum, B.J., Reissman, D.B., Pfefferbaum, R.L., et al.: Building Resilience to Mass Trauma Events[M]// Handbook of Injury and Violence Prevention, pp. 347–358. Springer, US (2007)
5. Klein, K.R., Nagel, N.E.: Resilience to natural hazards: How useful is the concept? Environ. Hazards **5**(1–2), 35–45 (2003)
6. Fiksel, J.: Designing resilient, sustainable systems. Environ. Sci. Technol. **37**(23), 5330–5339 (2003)
7. Longstaff, P.H.: Security, resilience and communication in unpredictable environments such as terrorism, natural disasters and complex technology. Program on Information Resources Policy, Center for Information Policy Research, Harvard University (2005)
8. Comfort, L.K., Boin, A., Demchak, C.C.: Designing Resilience. University of Pittsburgh Press, Pittsburgh (2010)
9. Weber, P.: Dynamic bayesian networks in system reliability analysis[J]. Fault Detect. Superv. Saf. Tech. Process. **6**(1), 444–449 (2007)
10. Klein, G.A., Orasanu, J., Calderwood, R., et al.: Decision Making in Action Models and Methods.[J], vol. 38(12). Psyccritiques, River Edge (1993)
11. Murphy, K.P.: Dynamic bayesian networks: representation, inference and learning[J]. Probab. Graph. Models **13**, 303–306 (2002)

Tasks and Errors

An Estimation and Comparison of Human Abilities to Communicate Information Through Pursuit Tracking vs. Pointing on a Single Axis

Michael Blandino[1(✉)], Edgar Berdahl[1], and R. William Soukoreff[2]

[1] Louisiana State University, Baton Rouge, LA, USA
{mblandi, edgarberdahl}@lsu.edu
[2] Toronto General Hospital Research Institute, Toronto, ON, Canada
will@dgp.toronto.edu

Abstract. This paper describes a human subject study that compared the limits at which humans could communicate information through pursuit tracking gestures versus pointing (i.e. tapping) gestures. These limits were measured by estimating the channel capacity of the human motor-control system for pursuit tracking versus pointing along a single axis. A human-computer interface was built for this purpose, consisting of a touch strip sensor co-located with a visual display. Bandwidth-limited Gaussian noise signals were used to create targets for subjects to follow, enabling estimation of the channel capacity at bandwidth limits ranging from 0.12 Hz to 12 Hz. Results indicate that for lower frequencies of movement (from 0.12 Hz to 1 Hz or 1.5 Hz), pointing gestures with such a sensor may tend to convey more information, whereas at higher frequencies (from 2.3 Hz or 2.9 Hz to as high as 12 Hz), pursuit tracking gestures will afford higher channel capacities. In this work, the direct comparison between pursuit tracking and pointing was made possible through application of the Nyquist sampling theorem. This study forms a methodological basis for comparing a wide range of continuous sensors and human capacities for controlling them. In this manner, the authors are aiming to eventually create knowledge useful for theorizing about and creating new kinds of computer-based musical instruments using diverse, ergonomic arrangements of continuous sensors.

Keywords: Pursuit tracking · Pointing accuracy · Shannon-Hartley theorem · Information theory · HCI · Continuous control · Analog sensor

1 Introduction

Musical practice can demand the performance of complex gestures accurately and repeatably in order to realize sound with composed attributes. Technical systems incorporated into new interfaces for musical performance often include sensors in order to afford continuous control of a parameter or a combined array of parameters that are mapped to that of musical synthesis systems. Accordingly, design of these interfaces will require a consideration of what demands of musical composition and performance can be accommodated with the sensors of the system.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 247–257, 2020.
https://doi.org/10.1007/978-3-030-20037-4_22

Human-computer interaction (HCI) literature reflects decades of investigation into the pointing gesture for communicating information and into the relationships of target characteristics to human capability. Fitts' Law and extensions within information theory have developed knowledge of the limits of information throughput using a pointing gesture, even informing international standards for pointing devices [1, 2].

Fewer investigations of pursuit tracking with continuous control have been conducted using information theory [3–7]. It does not appear that any studies have directly compared these modes of communicating information using a common human computer interface with a continuous control sensor.

The ability to convey information through such a sensor is an essential part of the utility of its afforded interaction. A quantitative measure of the upper limit of what amount of information may be conveyed through a sensor is pertinent to musical performance limitations of the sensor and, further, may be important to the design of its use in this application and in others.

Beyond applications in music, it is believed that this work can be informative for design of flight control systems, video games, assistive devices, other human-computer interactions, and ergonomics.

A prior pilot study of pursuit tracking using four continuous control sensors of different modes that were not co-located with their target signals showed that channel capacities as high as 4–5 bits per second were achieved with adequate training [8]. Of those four sensors, the system including the touch strip was found to have the highest channel capacity. The human subject study of this paper furthers this work by including a higher level of training of multiple subjects and a comparison to pointing/tapping in an equivalent model and target set.

2 Model

2.1 Fundamentals

In this work, it is assumed that the subjects are aware of some target signals $X(t)$ that they want to input into a computer. Due to various effects, somewhat different gesture signals $Y(t)$ are actually registered in the computer. It is decided to model this as a communications channel as shown in Fig. 1.

This model is for example suggested by prior research into human performance with airplane and related control systems [3, 5, 6]. Accordingly, the noise in the human motor control system is modeled with the signal $Z(t)$. This noise is understood to be approximately independent of the gestures being performed [5]. Such a model is suggested by research into neuromotor noise theory [9, 10]. Moreover, such signals are biomechanically filtered by the human body, which will also tend to make the noise signals look Gaussian distributed due to the Central Limit Theorem [11]. (Finally, further evidence along this vein includes the fact that errors in the endpoints of pointing tasks tend to be Gaussian distributed as well [12].)

However, the authors believe that the model requires an additional filter with impulse response $h(t)$ to model observed human behavior. Consider if it did not and imagine the case in which a subject is performing a gesture signal $Y(t)$ that approximates $X(t)$ albeit with some noise included. Due to independence of *Z(t)* and $X(t), E(Y^2(t)) = E(X^2(t)) + E(Z^2(t))$, implying that $E(Y^2(t)) > E(X^2(t))$, which will however not be the case if the user is following the target signal *X(t)* with the same power level. Therefore, a model component *h(t)* is needed to model how the subject's input signal component is attenuated to make room for the noise power *Z(t)*. *h(t)* could also in some situations potentially model other dynamic effects in the subject's performance [3].

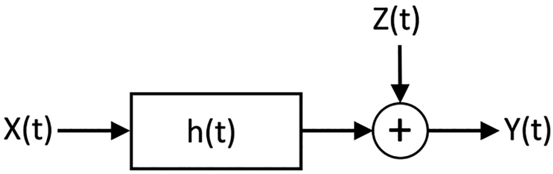


Fig. 1. A model of the user's performance in which *h(t)* is a filter's impulse response that models the deterministic component of a user's performance, and *Z(t)* models the random motor noise.

For the recordings made in the present study, not enough data was present to be able to robustly model *h(t)* in detail. Therefore, using Occam's razor, and in the case of the present application, it was decided to model *h(t)* with the constant h_0; in other words, the authors set $h(t) = h_0$.

For a given trial, if a subject is performing a gesture signal *Y(t)* that is very close to the target signal *X(t)*, then h_0 will be close to 1.0 and the noise *Z(t)* will have a low power. In contrast, if a subject is performing a gesture signal *Y(t)* that is not very precisely tracking a target signal, then h_0 will be significantly closer to zero, and the noise *Z(t)* will have a relatively larger power.

According to this model then, h_0 can be robustly estimated given even only small amounts of data. From the model, one can derive that

$$E(X(t)Y(t)) = E(X(t)(h_0X(t) + Z(t))) = E(h_0X^2(t)) + E(X(t)Z(t)). \tag{1}$$

Since the target signal *X(t)* and the motor noise *Z(t)* are uncorrelated, then $E(X(t)Z(t)) = 0$, which leads to the following:

$$E(X(t)Y(t)) = h_0E(X^2(t)). \tag{2}$$

$$h_0 = \frac{E(X(t)Y(t))}{E(X^2(t))}. \tag{3}$$

So finally, given some example data, the estimate of h_0 can be obtained by averaging as follows:

$$\hat{h}_0 = \frac{\text{avg}(X(t)Y(t))}{\text{avg}(X^2(t))}. \tag{4}$$

2.2 Channel Capacity for Pursuit Tracking (for Continuous Inputs)

Consider the case where the analysis is being performed on a single trial with bandwidth f_X. For pursuit tracking of continuous inputs, the channel capacity can then be estimated using the Shannon-Hartley theorem [11, 13]. For systems where the signal-to-noise ratio is constant across the bandwidth of the channel, the channel capacity at bandwidth f_X is then

$$C(f_x) = f_x \cdot \log_2(1 + \frac{S}{N}), \tag{5}$$

where $\frac{S}{N}$ is the signal-to-noise ratio, which can be estimated as follows:

$$\frac{S}{N} = \frac{E((h_0 X(t))^2)}{E(Z^2(t))} = \frac{E((h_0 X(t))^2)}{E((Y(t) - h_0 X(t))^2)} \approx \frac{\text{avg}\left((h_0 X(t))^2\right)}{\text{avg}\left((Y(t) - h_0 X(t))^2\right)}. \tag{6}$$

2.3 Channel Capacity for Pointing (for Discrete-Time Inputs)

For pointing, the signal-to-noise ratio can be estimated in essentially the same way. A single pointing gesture operates with the channel capacity of the discrete Gaussian channel [11]:

$$C_{pointingonce} = \frac{1}{2}\log_2(1 + \frac{S}{N}). \tag{7}$$

If sampled at the Nyquist rate (e.g. 2 f_x pointing gestures per second for a bandwidth of f_x), then the same expression as in (5) is obtained for the net channel capacity:

$$C(f_x) = 2f_x \cdot C_{pointingonce} = f_x \cdot \log_2(1 + \frac{S}{N}). \tag{8}$$

This correspondence, which is enabled by the sampling theorem, motivated the experimental design for the following subject test [14].

3 Subject Experiment

3.1 Apparatus

An experimental apparatus was assembled in order to compare pursuit tracking and pointing gestures using a common interface to match a co-located target signal (see Fig. 2). The apparatus was comprised of a flat screen high-definition monitor of 30 cm by 47.3 cm, a Spectra Symbol 200 mm "soft potentiometer" (also known as a touch strip), an Arduino Micro microcontroller, and a 5 V power adapter for reference voltage. As shown in Fig. 2, the touch strip was mounted to the display surface and placed 11 cm from one short side and centered evenly between the long sides of the display.

Fig. 2. The experimental apparatus provides a display with colocated sensor for target performance.

To achieve a higher accuracy of microcontroller sampling of the sensor output, an external reference voltage was maintained through a separate 5 V adapter connected to the reference pin of the Arduino Micro.

A program realized in the Cycling'74 Max application assembled and displayed the target signals onto the display and recorded the performed gesture data from the sensor as audio file data at 4410 samples per second. The application also provided instructions and control to progress through phases of the experiment.

3.2 Stimuli

Target signals were generated as bandwidth-limited Gaussian noise in two modes: pursuit tracking and pointing. For pursuit tracking gesture targets, a continuous curve with a length of 20 s at 4410 samples per second formed the target shape (see Fig. 3). These curves were formed by taking Gaussian-distributed noise sampled at 4410 Hz and filtering it by a fourth-order low-pass Butterworth filter. This filter was applied twice forwards and twice backwards, resulting in a zero-phase filter of net order sixteen [15].

For pointing gesture targets, 13 mm diamonds were presented at values sampled from the pursuit tracking curve. The signal was always sampled at twice the frequency of the bandwidth limit in an evenly spaced time interval (see Fig. 4). Sampling at twice the frequency bandwidth serves to meet the requirements of the Nyquist frequency sampling rate for reproducing the original signal [14].

These Gaussian target signals were generated for 12 frequency limits that were spaced logarithmically from 0.12 Hz to 12 Hz. Two signals were prepared for each bandwidth and in the two forms of pursuit tracking and pointing. Therefore, the total number of target gestures for each participant totalled 48 gestures.

Fig. 3. For pursuit tracking, the gesture waveform moved down the screen toward the touch strip, and the subject was asked to move her or his finger along the touch strip in synchrony with the waveform as it moved downward.

3.3 Procedure

Participants were seated in a chair of appropriate height to allow comfortable movement and free range of motion to interact with the interface. The apparatus was laid upon a work-station surface with display and attached sensor facing up, oriented with the side closest to the sensor immediately before the subject. Subject participants used an interface on the laptop device to navigate the study options and continue through its phases. There, they were directed to follow target signals of 20 s duration on the sensor apparatus.

In presentation, the two types of signals moved at the same rate from the top to the bottom of the screen to approach and travel below the sensor, crossing its axis. Targets moved at a rate of 23.6 cm per second with a total preview visibility of 2.94 s and post-view visibility of 0.97 s. The feature of implementing post-view visibility was believed to be novel, but the authors believed that it may have enabled subjects to more accurately see and follow the gesture. The range of display for the target gesture amplitude was 190 mm from a maximum value of +1.0 at the left to a minimum value at the right of −1.0.

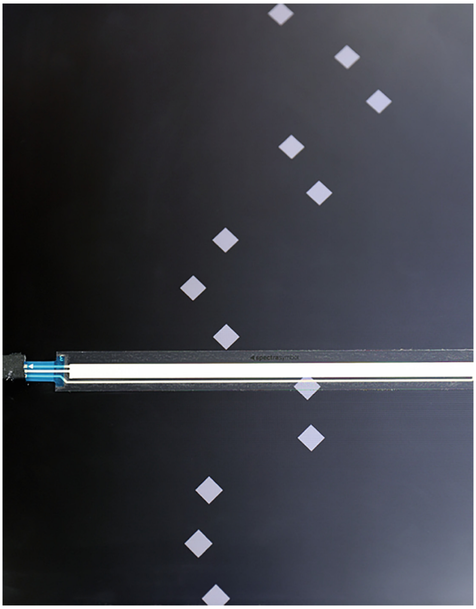


Fig. 4. For pointing gestures, diamond shapes were sampled from Gaussian targets at $2fx$ Hz, where fx is the bandwidth limit.

In order to ensure a measurement of the channel capacity for participants familiar with the interface, a training phase introduced the types of gestures to the subjects in three escalating levels of difficulty. Subjects were offered the opportunity to repeat gestures in training and also to request additional gestures until they felt satisfied with their command of and familiarity with the interface.

Instructions were provided to describe the type of movements and to characterize the training difficulty levels. Three levels were provided in training for both pursuit tracking and pointing/tapping. The 0.7 Hz, 1.5 Hz, and 7 Hz bandwidth limits were presented as easy, medium, and difficult levels, respectively. For the difficult level, subjects were encouraged to make their best effort to perform the target gestures with as much accuracy as possible.

During the recorded portion of the study, the order of the 48 gestures was randomized throughout the trial in order to avoid factors that may result from learned agility or developed fatigue of participants. Participants were given the opportunity to rest, if requested.

Upon completion of each gesture trial, the guiding interface presented the option of retrying the completed gesture in case the subject felt, in their own estimation, that they could improve their performance. No performance feedback or error estimate was provided. The gesture could be repeated an unlimited number of times. When satisfied with their performance, the subject would then elect the option to accept the last performed gesture and continue to the next one.

The duration of subject trials was 35 to 40 min of continuous participation.

3.4 Analysis

Before conducting analysis using an information-theoretic approach, some adjustments to the data were made. First, in instances where a participant was not touching the control strip, either due to error in their use of the sensor or due to exceeding its effective sensor area, a value of −1.0 was recorded by the sensor apparatus (its rest value). Second, to compensate for errors of anticipation or delay while pointing, the beginning and ending samples for each point instance were located and extended to a midpoint between neighboring point instances, with the rationale that the modified signal was still communicating the same information inputted by the user, just transformed into a slightly different format.

Third, to account for instances where subjects were consistently late or early in the performance of the gestures, an iterative calculation of the mean-squared- error from −200 ms to 200 ms was conducted in relation to the target signal at 1 ms intervals. In the interest of finding maximum channel capacities, the most favorable delay interval within the resolution described above was tabulated and accepted as the representative value for each trial. With these adjustments, a best representation of the performed gesture is prepared for the channel capacity calculation.

Using the signal-to-noise ratio as calculated in the time domain, the channel capacity may be calculated, utilizing the bandwidth limits and the limits of human performance speeds as observed in this study. The bandwidth of the signal in the case of the human computer system is limited not only by the target design, but also by the capability of movement in time by the human participant. Where the target signal exceeded this capacity of movement, the upper limit is applied within the bandwidth component to calculate the channel capacity.

To wit, upon analysis of pursuit tracking results using the Fast Fourier Trans- form (FFT), the highest sustained frequency rate of movement observed was 5.6 Hz. An upper limit of 5.6 Hz was therefore applied as input to the bandwidth of the Shannon-Hartley equation for the 7 Hz and 12 Hz target results for pursuit tracking gestures. For pointing gestures, a maximum of 7.0 Hz was observed for a sustained pointing movement rate. Accordingly, a maximum of 7.0 Hz was applied to the channel capacity calculation for the 12 Hz target results for pointing.

3.5 Results

Main Subject Pool Eight subjects (1:7, female:male) from the main subject pool participated in the study. All subjects were musicians enrolled in either undergraduate or graduate music study at a research university. Subjects performed gestures with their dominant hand.

As shown in Fig. 5, the mean observed channel capacity for pointing attained levels as high as 6 bits per second, representing the highest overall capacity for the subject pool. This peak channel capacity for pointing was at bandwidth limit 1.0 Hz, following a steady curve to that level and descending to the next highest capacity found near that level at 1.5 Hz.

The channel capacity of pursuit tracking similarly followed a discernible curve, clearly exceeding that of pointing capacities at 2.9 Hz and higher. Peak channel capacity for pursuit tracking was around 4 bits per second on average at bandwidth limit 2.3 Hz.

Analysis using Welch's t-test with Bonferroni correction identified any significance of differences across bandwidths between the two gesture types. It appears from these results that, with subjects having a very minimal amount of training, pointing at a lower frequency of movement allows communication of more information than pursuit tracking at such rates of movement. At 1.0 Hz, a mean of 2.6 bits/sec more information was communicated than with pursuit tracking (95% CI:1.53, 3.65; $p < 0.01$).

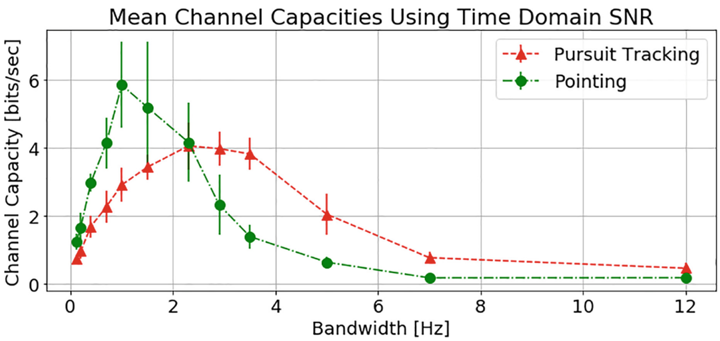


Fig. 5. Main subject pool: Estimated channel capacity across bandwidth limits *fx* of target signals for pursuit tracking and pointing gestures.

Under these conditions, at higher rates of movement, pursuit tracking appears to offer a higher capacity to communicate information. At 3.5 Hz, 2.4 bits/sec more information was communicated than with pointing (95% CI:1.8, 2.99, $p < 0.01$).

A varying delay was observed for all subjects. There are several factors that could contribute to this delay. Screen refresh rates in relation to the recording of input gestures present information to the subject later than the recording. Simple visibility of the target beneath the transparent sensor and estimation of its position under the opaque portion of the sensor could lead to some inaccuracy either before or after the recording moment. The delay of reaction to the previewed signal and delayed contact after the impulse to follow or touch the signal target point is a likely contributor to this observed delay as well.

A slackening of movement intensity was observed at the higher bandwidth limits for most participants, despite instructions of encouragement to try to follow as closely as possible or touch as many targets as possible. The seeming impossibility of following such a complex target or touching so many shapes at the rate presented was perhaps dispiriting. Fatigue could also be a factor here.

Author Data Two of the authors also participated in the study. Their data was treated separately as they had considerably more training gained during preparation of the study and apparatus design, although not as a controlled condition to prove a performance plateau. They also repeated their trials more frequently, in order to try to achieve even higher capacities. Their data is shown in Fig. 6. Overall, these two authors were able to achieve higher capacities both for pointing and for pursuit tracking. The additional training appeared to provide more benefit for the pursuit tracking condition, under which the authors almost managed to catch up with their maximum channel capacities for pointing (see Fig. 6).

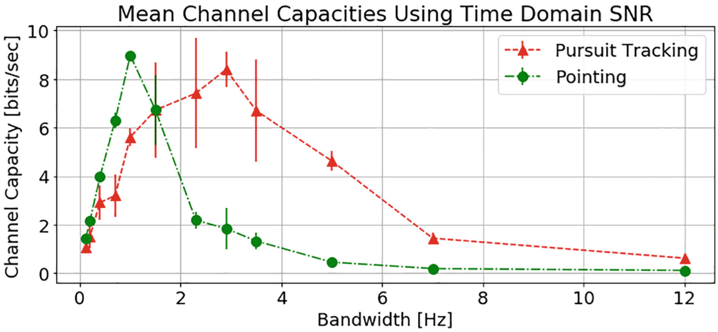


Fig. 6. Author data: Estimated channel capacity across bandwidth limits fX of target signals for pursuit tracking and pointing gestures.

4 Discussion

In general, even with a training session component to the study design, the subjects performed as novice users compared to the authors in using the interface. Therefore, the channel capacity results should be considered maxima only for such a class of users. A more intensive training protocol, perhaps combined with a competition paradigm, could improve results and demonstrate a higher channel capacity for an advanced performer with significant practice on the interface.

Factors that could differentiate the novice from the experienced user could include a residual uncertainty due to novelty, inattentiveness during the session, and a lack of learned adaptive behavior that would assist with anticipating movement. These latter could include strategic thinking about how to best perform in light of high frequency signal components.

5 Conclusions

In summary, a comparison of pursuit tracking and pointing gestures was observed on a single analog sensor interface that was co-located with visual target stimuli. Application analysis based in information theory shows a straightforward means for evaluation of subject performance using the interface in these two ways.

In utilizing systems for applications that require higher throughput rates, composer/designers or performers can ensure that capacity is available by arranging their gestures to include pointing at a rate of 2.0 Hz to 3 Hz. Conversely, where movement of 5 Hz to 10 Hz is desired, it is clear that a higher throughput is available via a continuous control movement than via pointing.

Further investigation along these lines should include more ambitious training with interface use by subjects to seek limits beyond the novice level. Indeed, analysis of performances after memorization of the target gestures as would be the case with the performance of a composed musical work would be informative. Virtuosic levels of pointing or pursuit tracking may differ from the results found here. No feedback other than the benefits of co-location with the target stimuli were provided. Investigation of haptic, sonic, or visual feedback on the performance accuracy for subjects may demonstrate that higher capacities are possible when such information is incorporated into the human computer system.

References

1. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. **47**(6), 381 (1954)
2. Soukoreff, R.W., MacKenzie, I.S.: Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' Law research in HCI. Int. J. Hum.-Comput. Stud. **61**(6), 751–789 (2004)
3. Elkind, J., Forgie, C.: Characteristics of the human operator in simple manual control systems. IRE Transact. Autom. Control **1**, 44–55 (1959)
4. Crossman, E.R.F.W.: The information-capacity of the human motor-system in pursuit tracking. Q. J. Exp. Psychol. **12**(1), 01–16 (1960)
5. Elkind, J., Sprague, L.: Transmission of information in simple manual control systems. IRE Transact. Hum. Factors in Electron. **1**, 58–60 (1961)
6. Baty, D.L.: Effects of display gain on human operator information processing rate in a rate control tracking task. IEEE Transact. Man Mach. Syst. **10**(4), 123–131 (1969)
7. Doubler, J.A., Childress, D.S.: An analysis of extended physiological proprioception as a prosthesis-control technique. J. Rehabil. Res. Dev. **21**(1), 5–18 (1984)
8. Berdahl, E., Blandino, M., Baker, D., Shanahan, D.: An approach for using information theory to investigate continuous control of analog sensors by humans. Proc. Audio Most. **2016**, 85–90 (2016)
9. van Galen, G.P., de Jong, W.P.: Fitts' law as the outcome of a dynamic noise filtering model of motor control. Hum. Mov. Sci. **14**(4), 539–571 (1995)
10. Van Gemmert, A.W.A.: The Effects of Mental Load and Stress on The Dynamics of Fine Motor Tasks. NICI, Nijmeegs Instituut voor Cognitie en Informatie, Nijmegen (1997)
11. Cover, T., Thomas, J.: Elements of Information Theory, 2nd edn. Wiley, Hoboken (2006)
12. Wobbrock, J.O., Cutrell, E., Harada, S., MacKenzie, I.S.: An error model for pointing based on Fitts' Law. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1613–1622 (2008)
13. Shannon, C.: Communication in the presence of noise. Proc. IRE **86**(1), 10–21 (1949)
14. Smith, J.O.: Mathematics of the Discrete Fourier Transform (DFT): With Audio Applications. W3 K Publishing (2007)
15. Smith, J.O.: Introduction to Digital Filters: With Audio Applications, vol. 2. W3 K Publishing (2007)

Linguistic Approach to Support Human Reliability Analysis and Validation Work in Advanced Radiotherapy Technologies

Mariarosa Giardina[1(✉)], Pietro Buffa[1], Pablo Cirrone[2], Luigi Raffaele[2,3], and Vincenzo Salamone[3]

[1] Department of Energy, Information Engineering and Mathematical Models, University of Palermo, Palermo, Italy
{mariarosa.giardina, pietro.buffa}@unipa.it
[2] Laboratori Nazionali del Sud, Istituto Nazionale di Fisica Nucleare (INFN-LNS), Catania, Italy
{pablo.cirrone, raffaele}@lns.infn.it
[3] Azienda Ospedaliero - Universitaria "Policlinico - Vittorio Emanuele" U.O.C. di Radiodiagnostica e Radioterapia, Catania, Italy
vincenzosalamone@hotmail.com

Abstract. An integrated methodology, that combines Hierarchical Task Analysis (HTA), Cognitive Task Analysis (CTA) and a modified Human Error Assessment and Reduction Technique (HEART), is proposed to identify safety-relevant human actions in innovative and advanced facilities. It is suggested to use in HEART the concept of linguistic expressions for proportion assessment factors. To prove its applicability, a case study is presented. The validation work has related to safety analyses of accidental events that can involve radiological over-exposure of patients during proton therapy treatments carried out by CATANA (Centro di AdroTerapia Applicazioni Nucleari Avanzate) centre, Italy. CATANA is the first Italian clinical facility that since 2002 used proton beams for ocular melanoma treatments. The results demonstrate the validity of the proposed integrated methodology as well as the versatility of HEART linguistic approach.

Keywords: Hierarchical Task Analysis · HEART · Operator human error · Adrotherapy · Centro di AdroTerapia Applicazioni Nucleari Avanzate

1 Introduction

Human Reliability Analysis (HRA) can provide valuable insights into identification human failures most critical to safety and so support Probabilistic Safety Assessment (PSA).

However, in innovative and advanced facilities, as particle accelerator for medical applications, HRA methods involve several issues, mainly related to the extensive use of one-of-a kind components, the lack of right know-how in the failure modes, or the singleness of some of operator's tasks [1–6].

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 258–270, 2020.
https://doi.org/10.1007/978-3-030-20037-4_23

Tools such as Failure Mode and Effect Analysis (FMEA) and HAZard and OPerability (HAZOP) analysis have been widely used [7–10] in the field of innovative technologies, but the importance of complementing these considering the human element is recognized for particle accelerators [11]. In this context, HRA can provide valuable insights concerning the identification of the human failures most critical to safety and their contributing factors, thus complementing the application of human factors design, verification and validation requirements and best practices.

However, when considering innovative installations, the application of PSA/HRA methods faces several challenges, mainly related to lack of experiences in failure modes and data collection. Moreover, the uniqueness of some operator tasks and of performance conditions limits the HRA applications, as well as the possibility to accumulate experience data across different facilities.

In this paper, above challenges are addressed by combining Hierarchical Task Analysis (HTA) [12], a systematic method to identify safety-relevant human actions, Cognitive task analysis (CTA) [13], aimed at understanding tasks that require a lot of cognitive activities from the user such as decision-making, problem-solving, memory, attention and judgement, and a modified Human Error Assessment and Reduction Technique (HEART), recognized as a flexible and resource-efficient method to quantify the human error probability, extensively used in the UK nuclear chemical, aviation, rail industries [14, 15].

HTA and CTA are first performed to identify sequences and hierarchies of tasks, from the top-level goals down to the level of individual operations, and to capture information about the expert's task performance. The application of HEART, that requires knowing operator performance conditions (e.g. timing involved in the operator action, plant operating procedures, human-machine interface, etc.), makes active use of collected data in HTA.

HEART provides reference Generic Task Type (GTT) values and support the analysts in assessing the influence of context-related factors, that are used to modify the reference value to obtain task and context specific estimates. These factors are referred to Error Producing Conditions (EPCs) that are the possible conditions having negative influence on task performances. The maximum extent of EPC is given by the Maximum Affect (MA) factor. A sensitive issue is to assess a proportion factor AP by means of which it is possible to adjust EPC's maximum affect in function of the real influence on the task in the examined context. The AP assessment is subjective (based on analyst judgement), furthermore limited guidance is available in current HEART documentations [16–18]. This aspect has led some researchers to develop 'anchoring schemes' to support the assessors, as well to try and improve consistency among different users, however the problem has still not been resolved. To assist the analyst in choice of AP weight for the task performance, it is suggested to use the concept of linguistic expressions.

In the paper, a case study is presented to prove the applicability of the proposed methodology.

The results are relevant to safety analyses of accidental events that can involve radiological over-exposure of patients during proton therapy treatments in CATANA (Centro di AdroTerapia Applicazioni Nucleari Avanzate) centre, at Laboratori Nazionali del Sud (LNS), National Institute for Nuclear Physics (INFN), Italy.

CATANA was the first Italian clinical facility that used 62 meV proton beams for ocular melanomas treatments. It was built thanks to the collaboration between INFN-LNS and Public Health Policlinic named AOUP-Vittorio Emanuele at Catania, and is operational since 2002. The proton beams are accelerated by the INFN-LNS superconducting cyclotron and more than 400 patients have been successfully treated.

The validation work has allowed to demonstrate the validity of the proposed integrated methodology.

2 Combination of HTA, CTA and HEART Techniques

HRA methods require in advance knowing operator performance conditions, consequently, HTA is first performed to identify sequences and hierarchies of tasks, from the top-level goals down to the level of individual operations.

For this purpose, the tree-like diagram is used to represent the sequence of various decisions and actions that the team is expected to perform when confronted with a particular process event.

HTA guidelines adopted in [19] have been used to perform data representation. The difference, compared to what is used in practice, is consisted in CTA applications by using interview and observation strategies to capture a description of the knowledge that the operators use to perform complex tasks.

Large amounts of knowledge have been obtained through collections of data acquired on the field, e.g. photos taken during the course of tasks, informative material provided by operators, analysis of the various storage and calculation tools used during the process. Subsequently, Worksheets, characterized by a table structure appropriately formulated in the context of this research, has been compiled.

HEART technique used to analyze the human factors is described in the next paragraph.

2.1 A Modified HEART Approach

HEART is recognized as being flexible and resource-efficient technique to quantify the human error probability, extensively used in the UK nuclear industry and also in most other industries (chemical, aviation, rail, medical etc.). The use of HEART in UK NPP (Nuclear Power Plant) PSAs has been reviewed by the UK nuclear regulatory body the Nuclear Installations Inspectorate (NII).

As above said, the technique is based, for different type of tasks, on the set of generic error probabilities, GTT, reported in Table 1. EPCs reported in Table 2 are the possible conditions having a negative influence on task performances, the maximum extent of which is given by MA factors.

The Human Error Probability (HEP) of a generic task is assessed by using the following relationship:

$$\mathrm{HEP} = \mathrm{NHEP} \prod_{\mathrm{i}} [(\mathrm{MA_i} - 1)\mathrm{AP_i} + 1]. \quad (1)$$

Table 1. NHEPs in HEART method.

Generic Task Type (GTT)		NHEP
A	Totally unfamiliar, performed at speed with no real idea of the likely consequences	0.55
B	Shift or restore system to a new or original state on a single attempt without procedures	0.26
C	Complex task requiring a high level of comprehension and skill	0.16
D	Fairly simple task performed rapidly or given scant attention	0.09
E	Routine, highly practiced, rapid task involving a relatively low level of skill	0.02
F	Restore or shift a system to an new state following procedures, with some checking	0.003
G	Completely familiar, well-designed, highly practiced, and routine task occurring several times per hour, performed to the highest possible standards by highly motivated, highly trained and experienced persons, totally aware of the implications of failure and having the time to correct potential errors but without the benefits provided by significant job aids	0.0004
H	Respond correctly to system commands even when there is an augmented or automated supervisory system providing an accurate interpretation of the system state	0.000002
M	Miscellaneous task for which no description can be found	0.03

where NHEP is the nominal HEP value for the selected GTT; MAi is the maximum affect for the ith EPC; and APi is the Assessed Proportion factor (from 0 to 1) for the ith EPC.

A main step is to assess a proportion factor AP by means of which it is possible to adjust EPC's maximum affect in function of the real influence of the selected EPC on the task of interest.

HEART has the following shortcomings:

- the method's focus is on generation of human error probabilities and does not provide guidance for the qualitative analysis aspects of HRA such as task analysis and the identification of human errors to be modelled;
- the technique provides a set of EPCs which can be used to take into account conditions having a negative influence on tasks, but it does not provide a practical systematic tool to support the analyst in making his decision;
- the suggested strategies for EPCs choice cannot take into account explicitly the context in which the tasks are performed (environmental conditions related to working conditions, ergonomics aspects, etc.);
- the use of the method is extremely subjective and relies heavily on the experience of the analyst.

The preliminary application of HTA-CTA analysis allows to overcome some of these issues because the related results can help the analyst to model in a more consistent way the dynamic of an error, highlighting several interacting factors on the task performance.

Table 2. EPCs in HEART method.

EPC	Error-producing condition	MA
1	Unfamiliarity with a novel or infrequent situation that is potentially important	17
2	Shortage of time for error detection and correction	11
3	Noisy or confused signals	10
4	A means of suppressing or overriding information or features that is too easily accessible	9
5	No means of conveying spatial and functional information to an operator in a form they can readily assimilate	9
6	Poor system or human user interface. A mismatch between an operator's model of the world and that imagined by a designer	8
7	No obvious means of reversing an unintended action	8
8	Information overload	6
9	Technique unlearning or one which requires application of an opposing philosophy	6
10	Transfer knowledge from one task to another	5
11	Ambiguity in the required performance standard	5
12	Mismatch between perceived and actual risk	4
13	Poor, ambiguous or ill-matched feed-back	4
14	No clear/direct/timely confirmation of an intended action from the portion of the system over which control is to be exerted	4
15	Inexperience (newly qualified but not expert)	3
16	Poor instructions or procedures	3
17	An impoverished quality of information conveyed by procedures and person-person interaction	3

Moreover, it is suggested to use in HEART method the concept of linguistic expressions in the representation of the proportion assessment factor, APi. Phrases or sentences in a natural language are very useful when one deals with situations too complex or ill-defined to be reasonably described in conventional quantitative expressions.

The linguistic classification reported in Table 3 is proposed to facilitate the management of AP_i that is evaluated by using the following relationship:

$$AP_i = AP_{PFi} \times AP_{MFi}. \tag{2}$$

where AP_{PFi} is the rate of AP_i, named Promoting Factors (PF), representing how much the negative circumstances/factors of EPC can influence the task, and AP_{MFi} is the rate of AP_i, named Mitigating Factors (MF), which introduces the influence of those favorable elements that somewhat mitigate the impact of error producing conditions.

This conceptual decomposition of APi, used in Eq. (1), into two factors, by using Eq. (2), allows to provide a more systematic and effective support for the required of 'anchoring schemes' and, consequently, help the analyst in assessing actual influences of EPC on examined tasks.

Table 3. Ranking scale of proportion assessment factors AP_{PFi} and AP_{MFi}.

AP_{PF} (Promoting Factor)			AP_{MF} (Mitigating Factor)		
Linguistic expression		Value	Linguistic expression		Value
Circumstance/factor increases EPC impact	Extremely Low Promotion (ELP)	0.1	Circumstance/factor reduces EPC impact	Best Mitigation (BM)	0.1
	Very Low Promotion (VLP)	0.2		Very Good Mitigation (VGM)	0.2
	Low Promotion (LP)	0.3		Good Mitigation (GM)	0.3
	Slight Promotion (SP)	0.4		Appreciably-Favorable Mitigation (AFM)	0.4
	Moderate Promotion (MP)	0.5		Moderate Mitigation (MM)	0.5
	Appreciable Promotion (AP)	0.6		Slightly-Favorable Mitigation (SFM)	0.6
	High Promotion (HP)	0.7		Poor Mitigation (PM)	0.7
	Very High Promotion (VHP)	0.8		Very Poor Mitigation (VPM)	0.8
	Extremely High Promotion (EHP)	0.9		Extremely Poor Mitigation (EPM)	0.9
	Full Promoting Factor (FPF)	1.0		Lack of Mitigating Factor (LMF)	1.0

Moreover, this approach can make the expert judgements easier, reducing to some extent the subjective aspects connected to these.

In Table 4, where it is shown an example from the analysis performed in this paper. AP_{PFi} and AP_{MFi} follow the same order and abbreviation reported in Table 3.

For the task ‘Acquisition of temperature values of water phantom, used for radiation absorbed dose measurements, and pressure and humidity in CATANA room’, the condition EPC17 ‘Little or no independent checking or testing of output’ is related to absence of pre-printed forms adequately formulated for data recording by the operator.

Table 4. Worksheet used for HEART application.

Plant: CATANA, **Task n.** 4
Operator Action description: Acquisition of temperature values of water phantom, used for radiation absorbed dose measurements, and pressure and humidity in CATANA room.
GTT: E - Routine, highly practiced, rapid task involving a relatively low level of skill
Nominal HEP: 0.02

EPC	MA	AP_{PF}	AP_{MF}	(MA -1) (AP_{PF} AP_{MF}) + 1	Assessor's note
5. No means of conveying spatial and functional information to an operator in a form they can readily assimilate	9	ELP = 0.1 ☐ VLP = 0.2 LP = 0.3 ☐ SP = 0.4 ☐ MP = 0.5 ☐ AP = 0.6 ☐ HP = 0.7 ☐ VHP = 0.8 ■ EHP = 0.9 ☐ FPF = 1.0 ☐	BM = 0.1 ☐ VGM = 0.2 ☐ GM = 0.3 ☐ AFM = 0.4 ☐ MM = 0.5 ☐ SFM = 0.6 ■ PM = 0.7 ☐ VPM = 0.8 ☐ EPM = 0.9 ☐ LMF = 1.0 ☐	4.84	**AP_{PF}:** data recording is performed by using one time sheet (sheet is not regularly updated) **AP_{MF}:** More operators perform this task. There are two measurement instruments (analogue and digital) to reduce conditions of read error.
17. Little or no independent checking or testing of output	3	ELP = 0.1 ☐ VLP = 0.2 ☐ LP = 0.3 ☐ SP = 0.4 ☐ MP = 0.5 ☐ AP = 0.6 ☐ HP = 0.7 ■ VHP = 0.8 ☐ EHP = 0.9 ☐ FPF = 1.0 ☐	BM = 0.1 ☐ VGM = 0.2 ☐ GM = 0.3 ☐ AFM = 0.4 ☐ MM = 0.5 ☐ SFM = 0.6 ☐ PM = 0.7 ☐ VPM = 0.8 ☐ EPM = 0.9 ☐ LMF = 1.0 ■	2.4	**AP_{PF}:** Absence of pre-printed forms adequately formulated for data recording. **AP_{MF}:** No mitigating elements are present.

Assessed HEP: 0.02 x 4.84 x 2.4 = **0.23**
Assessor's comment: use of pre-printed work sheets to store data read from measurement instruments are suggested.

It is deemed to have 70% of chance that could be significant in the examined context (i.e. High Promotion condition with AP_{PF} = 0.7 in Table 3).

AP_{MF} equal to 1 allows to take into consideration that there is no further recovery action to highlight error or negligence (i.e. Lack of Mitigating Factor in Table 3).

3 CATANA Description

CATANA passive proton beam line has been entirely built at INFN-LNS of Catania (Fig. 1). The proton beam exits in air through 50 μm kapton window placed at about 3 meters from isocenter. Just before the exit window the first scattering foil made of 15 μm tantalum is placed in vacuum.

The first element of the beam in air is a second tantalum foil 25 μm thick provided with a central brass stopper of 4 mm in diameter. The double foils scattering system is optimized to obtain a good homogeneity in terms of lateral off-axis dose distribution, minimizing the energy loss. Beam data are acquired in a water phantom with a Hi-pSi

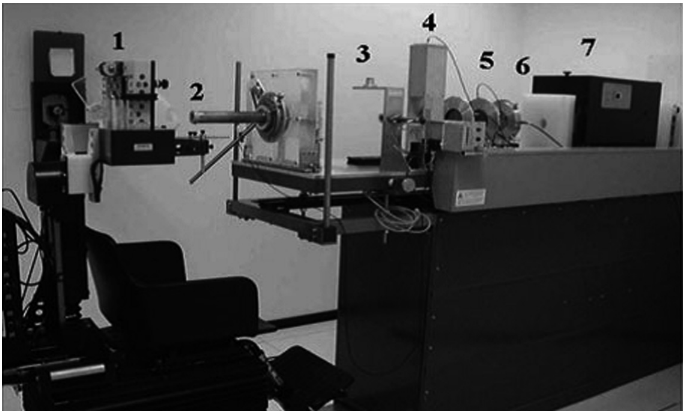


Fig. 1. View of CATANA beam line: 1. Treatment chair for patient immobilization; 2. Final collimator; 3. Positioning laser; 4. Light field simulator; 5. Monitor chambers; 6. Intermediate collimator; 7. Box for the location of modulator wheel and range shifter.

diode (0.6 mm detector diameter) at a depth of 12 mm, corresponding to the middle of a Spread-Out Bragg Peak (SOBP).

Range shifter and range modulator are placed inside a box, downstream of the scattering system. Two laser diodes, located orthogonally and coaxially to beam line, provide a system for the isocenter identification and for patient centering.

The radiation field is simulated by a light field with cross-air indicating the principal axis. A key element of the treatment line is represented by two transmission monitor ionization chambers, which provide on-line control of the dose delivered to the patient. The last element before isocenter is the brass shaped patient collimator located 83 mm upstream of the isocenter [20].

In order to optimize the geometrical reconstruction of the eye and the tumor, 5 pairs of X-ray images for five different fixation angles of the eye are acquired. All data are then elaborated by 3-D treatment planning system called EYEPLAN [21].

4 Application of HTA-CTA and HRA Analysis

CATANA procedures are articulated on the basis of a large number of well-defined activities, carried out according to time-framed targets, sometimes realized by one or more operators. The analysis was performed after a direct observation of planned tasks for various therapy cycles.

The General Objective (GO), i.e. proton therapy for eye melanoma treatment, has been divided into five main processes (Fig. 2):

Task.1 Verification of CATANA safety systems and equipment control procedure;

Task.2 Planning of radiotherapy treatment, first simulation (eight days before effective treatment);

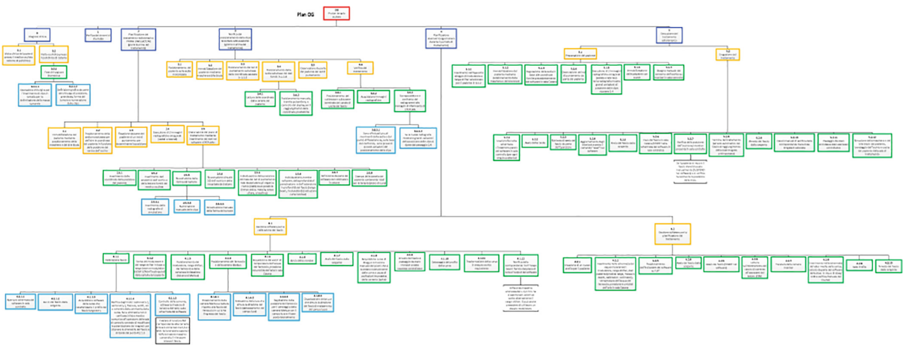


Fig. 2. HTA tree obtained for the proton therapy procedure in CATANA.

Task.3 Verification of procedures defined in planning treatment, second simulation (one day before effective treatment);

Task.4 Dosimetric planning (daily for the entire treatment period);

Task.5 Execution of the treatment.

For the sake of brevity, only some above tasks are described with some more details as follows:

"*Task 2"* is performed a first simulation, for each patient, eight days before the effective treatment. The aim is to obtain 3-D model of the eye and to plan the treatment.

The patient is placed on a motorized chair in CATANA room and immobilized by using a thermoplastic mask and a bite block.

The medical physicist, hereafter referred to as "medical physicist 1", regulates the motorized chair, defines the patient's coordinates and fixes a luminous led, located on a graduated scale, to determine the patient's optimal position for the therapy, minimizing the dose to healthy organs.

Finally, the medical physicist in the control room, indicated as "medical physicist 2", elaborates the treatment plan with the software EYEPLAN, developed at the Massachusetts General Hospital for the therapy of ocular tumors with protons, updated with new features provided from the Paul Scherrer Institute (PSI), Switzerland.

"*Task 3"* provides a second procedure that reviews all activities performed in Task 2. Some parameters of the chair's position are modified, if necessary.

"*Task 4"* executes the activities necessary to proton beam calibration. It involves a medical team, LNS operators and "medical physicist 2".

"Medical physicist 2" contacts by phone the operator in the accelerator control room and asks to start the equipment needed to produce the proton beam. Subsequently, he chooses the modulator wheel and the range shifter and places them in the special housings of the beam line system (Fig. 1). Other operators position the water phantom and ionization chamber in CATANA room.

Before leaving the CATANA room, "medical physicist 2" records on a sheet of paper phantom water temperature, measured by using both analog and digital instrument, and air pressure and humidity values of the room.

In the control room, he creates a new patient profile in the Treatment Planning Software (TPS) and inserts the data previously recorded in the sheet, i.e.: patient data; range shifter and modulator wheel parameters; values of phantom temperature; air pressure and humidity.

Afterwards, he delivers the proton beam through “START” command, present in the screen of TPS. Finally, the proton beam calibration curves are evaluated.

“*Task 5”* is related to the treatment execution that involves several important steps.

The modulator and the range shifter, chosen for the patient, are placed in the beam line system. A second operator checks the correct insertion of the two components.

At the same time, the patient is prepared for the treatment and “medical physicist 1”, on basis of the coordinates evaluated during Task 3, adjusts the position of the motorized chair. A camera located in CATANA room allows to visualize the patient’s eye in a monitor placed in the control room, the radiotherapist traces the outline of the eye on the screen to monitor any eye unwanted movements during the treatment.

“Medical physicist 2” loads patient profile created in TPS during Task 4. Patient positioning procedure on the motorized chair is performed. “Medical physicist 1” carries out the patrol procedure in CATANA room and verbally communicates to “medical physicist 2” that is possible to proceed with the treatment.

“Medical physicist 2” resets the safety system interlocks and asks the start of proton beam, calling by phone the operator in the accelerator control room.

During the treatment, if a macroscopic eye movement is detected in the monitor located in the control room, the beam is immediately stopped via the “SUSPEND” button in the “TREATMENT PANEL” screen of TPS.

When the radiation dose value is reached, the software automatically stops the therapy.

Finally, “medical physicist 1” and “medical physicist 2” perform calculations to verify the correspondence between the delivered radiation dose and that was calculated during the planning phase.

4.1 Critical Tasks Highlighted by HTA/CTA Analysis and Human Errors Probability Evaluation

HTA/CTA analysis highlighted the following main issues:

Task 2.

- Error conditions in eye anatomic data entry in EYEPLAN software;

Task 4.

- no warning light signal to alert the operators that the proton beam is available at the entrance gate of the CATANA room (i.e. proton beam is up, but not delivered inside the treatment room);
- absence of barcode to identify the range shifter. If the range shifter is not the right for the patient, the barcode could activate via software a proton beam interlock. The mask, the bite block and the collimator should report the same barcode in order to avoid misunderstandings;

- lack of pre-printed work sheet form, to store data read from the measurement instruments (e.g. the sheet should include: date of acquisition, name of the operator who performed the task, clear indication of recorded data);

Task 5.

- lack of procedures to avoid patient homonymy errors;
- errors in loading of patient data from TPS.

HEART technique was applied by using the scheme reported in Tables 1 through 4. It is worth noting that CTA analysis allowed to characterize the tasks performed by the various involved operators and the conditions that could have a negative influence on the task performance.

In HEART application to avoid the choice of EPCs to be considered similar, guidelines suggested in [22], that allowed to use EPCs compatible among them, were used.

Table 5 shows some results of HEP values together with suggestions useful for increasing the human reliability during the various steps of the examined procedures.

Table 5. Results obtained by HEART method.

Task number	Description of the work undertaken	HEP	Suggestions
Task 2	Eye anatomic data entry in EYEPLAN software	0.055	Redundancy of “control/supervision” by a second operator, qualified for this task
Task 4	Data read from the measurement instruments in CATANA room	0.23	Use of pre-printed work sheet to store the data read from measurement instruments (e.g. sheet form should include: date of acquisition, name of the operator who performed the task, clear indication of recorded data)
Task 4	Inserting the following data in the TPS: patient information, modulator, range shifter, calibration collimator….	0.11	redundancy of “control/supervision” by a second operator qualified for this task
Task 5	Loading patient data in TPS	0.033	Use of two lists of daily treated patients, the first list should be placed at the entrance to the CATANA room and the second in the control room. The medical physicist, in the treatment room, should require at the operator, in the control room, to compare the patient’s name is present in the program list

5 Conclusion

An integrated methodology, that combines Hierarchical Task Analysis, Cognitive task analysis and a modified Human Error Assessment and Reduction Technique, is proposed to safety analyses of accidental events that can involve radiological overexposure of patients during proton therapy treatments carried out in CATANA center at Laboratori Nazionali del Sud (LNS), National Institute for Nuclear Physics (INFN), Italy. This Center is first Italian clinical facility that since 2002 uses proton beams for ocular melanomas treatments.

The results have allowed to demonstrate the validity of the proposed integrated approach as well as the versatility of HEART linguistic method.

The suggestions taken out from the application of the proposed methodology were critically discussed and examined with LNS's experts. This made it possible to verify the effectiveness and validity of the risk analysis adopted in advanced and innovative facilities as the one being examined in this research work.

References

1. Castiglia, F., Giardina, M.: Fuzzy risk analysis of a modern γ-ray industrial irradiator. Health phys. **100**(6), 622–631 (2011)
2. Castiglia, F., Giardina, M.: Analysis of operator human errors in hydrogen refuelling stations: comparison between human rate assessment techniques. Int. J. Hydrogen Energy **38** (2), 1166–1176 (2013)
3. Giardina, M., Castiglia, F., Tomarchio, E.: Risk assessment of component failure modes and human errors using a new FMECA approach: application in the safety analysis of HDR brachytherapy. J. Radiol. Prot. **34**(4), 891–914 (2014)
4. Castiglia, F., Giardina, M., Tomarchio, E.: THERP and HEART integrated methodology for human error assessment. Radiat. Phys. Chem. **116**, 262–266 (2015)
5. Buffa, P., Giardina, M., Greco, S.F., Palermo, G., Dang, V.N., Podofillini, L., Esposito, J., Prete, G.: Human reliability analysis to support operational planning of an experimental facility. In: 25th European Safety and Reliability Conference, pp. 2901–2908 (2015)
6. Giardina, M., Buffa, P, Dang, V., Greco, S.F., Podofillini, L., Prete, G.: Early-design improvement of human reliability in an experimental facility: A combined approach and application on SPES. Saf. Sci. (2018). https://doi.org/10.1016/j.ssci.2018.08.008
7. Casamirra, M., Castiglia, F., Giardina, M., Lombardo, C.: Safety studies of a hydrogen refuelling station: determination of the occurrence frequency of the accidental scenarios. Int. J. Hydrogen Energy **34**(14), 5846–5854 (2009)
8. Morale, M., Giardina, M.: Safety study of an LNG regasification plant using an FMECA and HAZOP integrated methodology. J. Loss Prev. Process Ind. **35**, 35–45 (2015)
9. Tomarchio, E., Giardina, M., Veronese, I., Cantone, M.C.: A review of healthcare failure mode and effects analysis (HFMEA) in radiotherapy. Health Phys. **111**(4), 317–326 (2017)
10. Burgazzi, L.: Probabilistic safety analysis of an accelerator-lithium target based experimental facility. Nucl. Eng. Des. **236**, 1264–1274 (2006)
11. Sarkar, P.K.: Deterministic and probabilistic analysis of safety in particle accelerators. Radiat. Prot. Environ. **35**, 164–174 (2012)
12. Shepherd, A.: Hierarchical Task Analysis. Taylor & Francis, London (2001)

13. Schraagen, J.M., Chipman, S.F., Shute, V.J.: Cognitive Task Analysis. Lawrence Erlbaum Associates, Mahwah (2000)
14. Williams, J.C.: HEART – a proposed method for assessing and reducing human error. In: 9th Advance in Reliability Technology Symposium, University of Bradford (1986)
15. Williams, J.C.: HEART – a proposed method for achieving high reliability in process operation by means of human factors engineering technology. Saf. Reliab. **35**(3), 5–25 (2015)
16. Casamirra, M., Castiglia, F., Giardina, M., Tomarchio, E.: Fuzzy modelling of HEART methodology: application in safety analyses of accidental exposure in irradiation plants. Radiat. Eff. Defects Solids **164**(5–6), 291–296 (2009)
17. Castiglia, F., Giardina, M., Tomarchio, E.: Risk analysis using fuzzy set theory of the accidental exposure of medical staff during brachytherapy procedures. J. Radiol. Prot. **30**(1), 49–62 (2010)
18. Castiglia, F., Giardina, M., Caravello, F.P.: Fuzzy fault tree analysis in modern γ-ray industrial irradiator: use of fuzzy version of HEART and CREAM techniques for human error evaluation. In: 9th International Conference on Probabilistic Safety Assessment and Management, vol. 2, no. 2008, pp. 1377–1384 (2008)
19. Kirwan, B.: A Guide to Practical Human Reliability Assessment. CRC Press Book, Boca Raton (1994)
20. Cirrone, G.A.P., Cuttone, G., et al.: IEEE Trans. Nucl. Sci. **51**, 860 (2004)
21. Damato, B., Kacperek, A., Chopra, M., Sheen, M., Campbell, I., Errington, R.D.: Int. J. Radiat. Oncol. Biol. Phys. **63**, 109 (2005)
22. Gibson, W.H.: Railway action reliability assessment user manual. A technique for the quantification of human error in the rail industry, Great Britain. Rail Safety & Standards Board, RSSB (2012)

Proposal of Information Analysis System on Latent Human Error

Tatsuya Shimada[1(✉)], Riku Tsunori[2], Kasumi Koike[2], and Yusaku Okada[2]

[1] Graduate School of Science Technology, Keio University, Kanagawa, Japan
shimada1046@keio.jp
[2] Department of Administration Engineering, Faculty of Science and Technology, Keio University, Kanagawa, Japan
{rikustunori,k.koike.9021}@keio.jp, okada@ae.keio.ac.jp

Abstract. Latent Human Error information is gathered in companies, but there is a problem that it cannot be used well. The problem behind it is lack of knowledge on trouble analysis and insufficient analysis on text data. Therefore, in this research, we propose an information analysis system on latent human error. In the analysis system, the following three important elements are included. The first point is knowledge of Human Factors which is useful for trouble analysis. The second point is natural language processing technology which processes text information on trouble. The third point is statistics and DeepLearning technology using processed text information. Based on these factors, we aim to build an analysis support system that can be used by people who are not familiar with trouble analysis at the company site.

Keywords: Latent Human Error · Human factors · Machine learning

1 Introduction

In this research, among trouble information and potential incident information collected by companies, potential risk information with possibility of leading to human error is defined as Latent Human Error Information. The trouble information currently collected in companies includes text data describing occurrence events, and metadata such as date and time, place and department. The company's safety management department uses this information for risk assessment and aims to improve the safety management system. Specifically, the department is trying to make use of information by simple analysis of metadata, or sensory evaluation by trusted veterans. However, there are three problems. The first point is the limit of analysis by simple tabulation of metadata only (Problem 1). Depending on the person, Latent Human Error Information tends to be ambiguously judged whether it is risk to be reported. For this reason, the information that should be collected is judged not to be a risk, and it is not reported, and the relationship that the risk is higher as the number of events is larger does not hold. That is considered to be the cause of the limit of analysis by simple tabulation. The second point is the problem of the structure of the safety management department (Problem 2).

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 271–275, 2020.
https://doi.org/10.1007/978-3-030-20037-4_24

Most of this department, except for several experts, consists of people coming from the site with job rotation of two or three year periods. The reason why the safety management department does not secure new graduate talent is because there is a philosophy that people who do not know the site can not consider the safety of the site. If the person who came to that department stayed for many years, the knowledge of the site will be out of date, so that person will return to the original site in two or three year period. As a result, there is insufficient time to learn and grow knowledge on safety, and there is a not low hurdle for that person to make good use of the Latent Human Error Information. The third point is the problem of subjective evaluation reliability (Problem 3). Subjective evaluation is not bad. An experienced and professional worker has experiences of learning from failure in times when safety measures were inadequate. This experience is helpful in evaluating the risk of trouble. However, as safety measures progress, opportunities to learn from mistakes on site have decreased, and the number of young people who are beginning to work just as they are manual is increasing. Therefore, it is difficult to train an experienced and professional worker like the present. Dependence on subjective evaluation only has a potential danger from a long-term perspective. In this research, we propose a system concept and planning that solves these problems and helps appropriate safety management activities.

2 Proposal System

2.1 Latent Human Error Factors Extraction Support Function

The proposed system mainly has three functions. The first function is to support factor extraction leading to human errors. As mentioned in Problem 2, those who utilize the collected Latent Human Error Information lack the skill to appropriately extract factors leading to human errors from the information. With reference to past research [1] systematically summarizing factors leading to human errors, we provide rule-based factors extraction support function (Fig. 1).

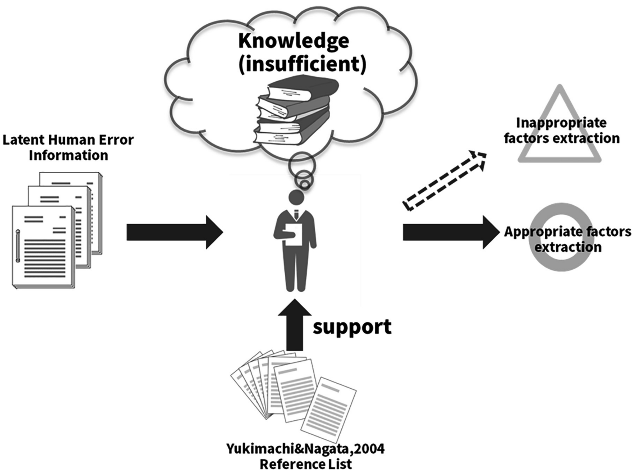


Fig. 1. Image of Latent Human Error factors extraction support function

2.2 Analysis Support Function by Natural Language Processing Technology

The second function is analysis support using natural language processing technology for text data in Latent Human Error Information. As mentioned in Problem 1, there is a limit to the analysis by simple tabulation of metadata. There is research on analysis support utilizing text data [2–4]. However, for less-experienced people, there are not low hurdles when applying analysis methods and interpreting results. Complex processing is mechanically processed using word2vec [5] and LSTM (Long Short-Term Memory) [6], and results are presented in a form that is easy to interpret.

2.3 Analysis Support Function by Statistical Processing and Deep Learning

The third function is to utilize text data by statistical processing and deep learning technology. The appearance frequency and co-occurrence expression of words are visualized by statistical processing, and the tendency of text data is captured. Also, as mentioned problem 3, it is important not to rely on veteran subjective assessment. Supervised learning is conducted by using veteran evaluated results as to what kind of information is judged to be risky information (level0 ~ level5). We aim to make models learning veteran evaluation criteria by utilizing DeepLearning technology together with the second function (Fig. 2).

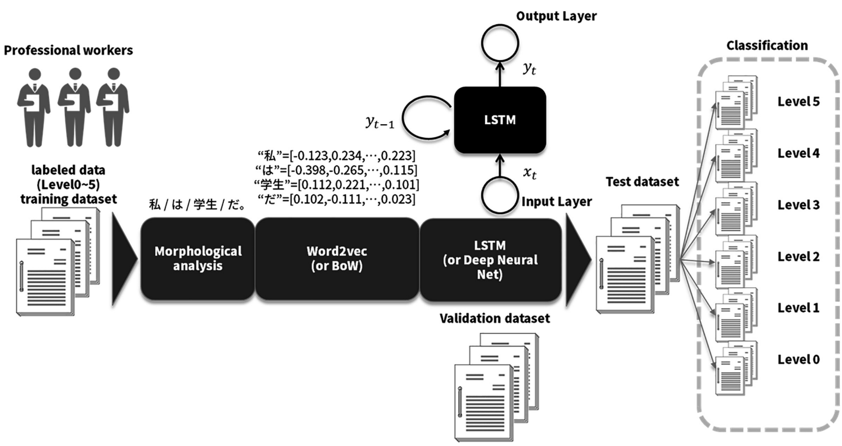


Fig. 2. Image of function 2 and function 3

3 Partial Verification of Proposed System

In the proposed system functions, we verify the function of utilizing text data by statistical processing using KH coder [7, 8].

Figure 3 is a simple summary of feature word frequencies, and Fig. 4 is a diagram showing the features of usage phrases of latent human error information for each employee's length of service by correspondence analysis. Companies collecting information used for analysis focus on safety education for young people with less than 10 years of service. On the other hand, veterans of more than 10 years do only the minimum necessary education. Based on that fact, it can be judged from the result of Fig. 4 that the educational effect appeared in the way of writing Latent Human Error Information. The possibility that the function based on statistical processing can be used effectively is shown.

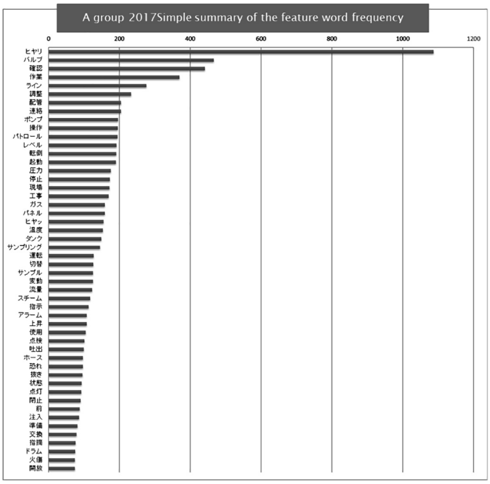


Fig. 3. Simple summary of the feature word frequency

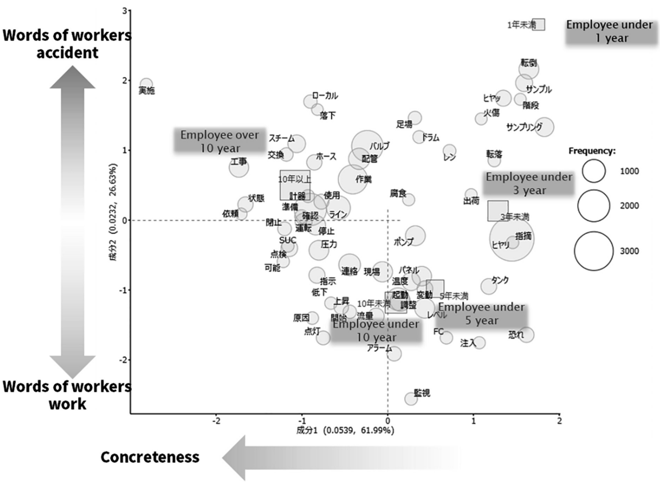


Fig. 4. Correspondence analysis (text data and each employee's length of service)

4 Conclusion

In this research, we proposed the concept and planning of a system with the function to effectively utilize latent human error information. We also checked the effectiveness of some of the functions. We plan to have the safety administrator use the entire prototype system and verify its effectiveness.

References

1. Yukimachi, T., Nagata, M.: Reference list on basis of GAP-W concept and a case study. Hum. Factors **9**(1), 7–14 (2004)
2. Kimura, M., Ohkura, M., Tsuchiya, F.: Text mining analysis of medical near-miss reports. Japan. J. Ergon. **43**, 232–233 (2006)
3. Takeda, Y., Nagao, Y., Sato, S., Hayashi, H.: Development of an analysis method to the incident report from the Kyoto University hospital using TRENDREADER. J. Soc. Saf. Sci. **13**, 305–315 (2010)
4. Otani, Y.: A Study on Knowledge Extraction from Free Form Incident Report. Mie University (2007)
5. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space. In: Proceedings of Workshop at ICLR (2013)
6. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
7. Higuchi, K.: A two-step approach to quantitative content analysis: KH Coder tutorial using Anne of Green Gables (Part I). Ritsumeikan Soc. Sci. Rev. **52**(3), 77–91 (2016)
8. Higuchi, K.: A two-step approach to quantitative content analysis: KH Coder tutorial using Anne of Green Gables (Part II). Ritsumeikan Soc. Sci. Rev. **53**(1), 137–147 (2017)

Mining Human Error Incident Patterns with Text Classification

Toru Nakata(✉)

Artificial Intelligence Research Center, National Institute of Advanced Science and Technology (AIST), Aomi 2-4-7, Koto-ku, Tokyo 135-00064, Japan
torunakata@gmail.com

Abstract. Reports on human error incidents contain crucial information to understand why and how incidents happened. There are huge numbers of documents that report industrial accidents or incidents. Instead of reading by humans, we can use document classification technique to find valuable knowledge hidden in the incident reports. Using the technique of document classification, we can detect group of frequent incident. First, the computer observes a set of words (so-called bag-of-word) which appear in each report. Second, the computer calculates similarities among reports. The report similarity is evaluated as agreement ratio of frequency of term appearances. Third, the computer builds the tree of report similarity: this tree holds group of similar reports on a branch. We find the typical patterns of incidents as branches of the tree. Fourth, the computer now calculates similarities of words, which are evaluated as ratio of word co-occurrence.

Keywords: Human factors · Human error · Text mining · Document classification · Big data

1 Introduction

Most of major industrial accidents are caused and/or escalated by human errors. Analysis of human mistakes has the highest priority for accident prevention.

Mechanism of human error, however, is hard to analyze in general. Other causes of accident, such as mechanical failures, are suitable for physical or chemical analysis, since they behave particular natural laws. In contrast, human behavior does not well fit to mathematical law in general, because human activity is structured with many factors and generated through consideration. Especially erroneous actions are very hard to explain by statistical method.

We usually study written documents (book, newspaper articles, worker's reports and so forth) to understand how and why human errors happened in industrial accidents in past. Actually, we already have many large datasets about industrial incidents; most of them are collected by governmental or public organization of many countries monitoring industry safety. Those records are opened and accessible for anyone on the internet.

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 276–285, 2020.
https://doi.org/10.1007/978-3-030-20037-4_25

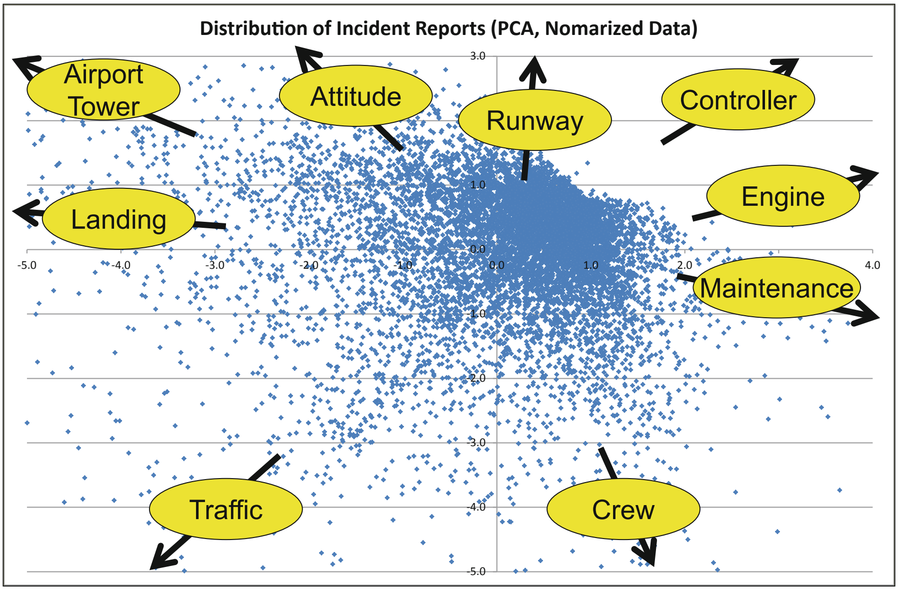


Fig. 1. An example result of conventional document classification with principal component analysis (PCA) on Bag-of-Word. Each blue dot represents a certain incident report of aviation. The position is determined in respect to the contents of words in the report, so that similar reports are putted close together. Even though we can find some correlations between certain keywords (in the yellow ovals) and particular directions on the plot, we cannot distinguish explicit patterns of their stories, since majority of the blue dots are concentrated on the center.

In general, those databases of industrial incident reports have large size of text data, because laws on industrial safety often require companies to upload reports on their incidents mandatory. It is natural that many databases have over thousands reports.

Even though we have accessible and large datasets, we have trouble to analyze them. Reading the reports is the best way for deep analysis, but we do not have enough number of experts nor time to do that.

Natural language processing (NLP) technology should owe this text-mining problem on industrial incident analysis. Bioinformatics is an early adopter of text mining to process huge amount of text data (namely, biological papers) [1]. Some of their techniques are being developed and getting some success to detect useful facts on causes of illness and materials for medicines. Each paper reports only a simple finding that is not enough to be a useful fact. Beneficial facts can be derived from not a single paper but combination of many sources of information.

We, therefore, should use NLP techniques to analyze industrial incident reports, especially about human errors.

Some primitive NLP techniques are not enough to analyze deeply.

For instance, Principle Component Analysis (PCA) is one of simplest NLP method to detect tendencies among text dataset. We applied this method for 4 469 reports about aviation safety collected NASA Aviation Safety Reporting System (ASRS) [2]. The result about distribution of report contents was summarized with Fig. 1. If the reports have explicit patterns on their contents, the distribution has particular shape to reflect them. The distribution was strongly concentrated around the center, and it means the method failed to catch the differences among the reports and recognized most of reports are more or less same stories. Each report contains many common phrases, so they look similar to each other.

Text data, in general, is high-dimensional data; if a text consists of 500 difference words, it should be dealt as 500 dimensional data to express existences of each word. PCA and other primitive methods give up when they face such high-dimensional data.

In this paper, the author proposes a technique of text-classification tree to analyze large set of human error incident reports. This method is one of well-known method in NLP, but it is new for safety engineering. The purpose of this paper also covers practical tips to use this method.

2 Detecting Meaning from Reports with Neighbor-Joining Tree Method

2.1 Classification of Incident Reports with Neighbor-Joining Classification Tree

Document Classification is one of basic technology to observe and to detect hidden patterns and tendencies among a set of text information [3]. Classification gives us the following knowledges:

- patterns of documents, and divergence of the document dataset,
- representative words of each cluster detected as the center of word distribution, (so those words are convenient to summarize meaning of each pattern), and
- switching words, of which existence in document strongly effect classification of the document.

There are many clustering method, and many of them require us to designate the number of clusters/categories beforehand. The proper number of cluster is, however, very difficult to determine. We use hierarchical classification method, which does not require such decision on the cluster number and give a result as a hierarchical tree. Observing the hierarchical tree, we can find proper number of categories.

We explain our method to generate Neighbor-Joining Classification Tree (NJCT) by taking simple example of document dataset shown in Fig. 2. Suppose we have four incident reports.

First, we ignore grammatical and syntactic information of the documents, because the strict analysis of sentences is still difficult to the computer and the quality of result is not good enough. Here, we deal each document by watching only existence of words; i.e. each document is regarded as a set of words (‘Bag-of-Words’).

Second, amount of difference (that is to say ‘distance’) among the documents are measured. In the example of Fig. 2, we use number of different words between two documents as the distance. (There are many other ways to define the distance).

Third, we build the tree from its terminals. Find the pair of documents with the smallest distance, and tie up them first. In this example, the pair of Report A and B with one distance is tied up first. The pair of Report C and D should be tied also. Then, we observe the distances among the remainder documents and tied pairs. Tie them in order of distance. Finally, we get the tree as Fig. 3.

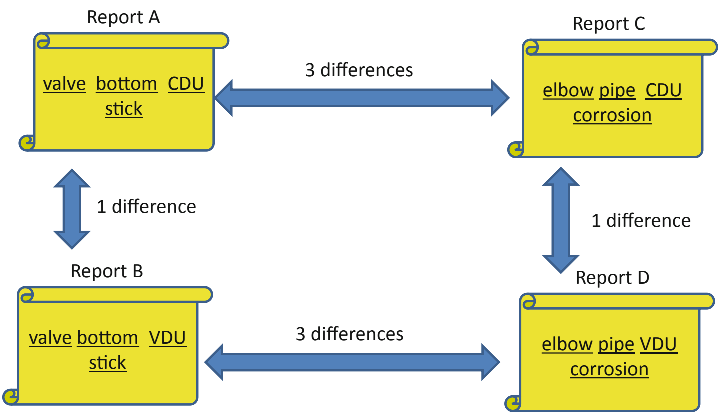


Fig. 2. Distances among incident reports. Four incident reports are described as Bag-of-Word. To keep simplicity, the distance in this example is defined as number of different words.

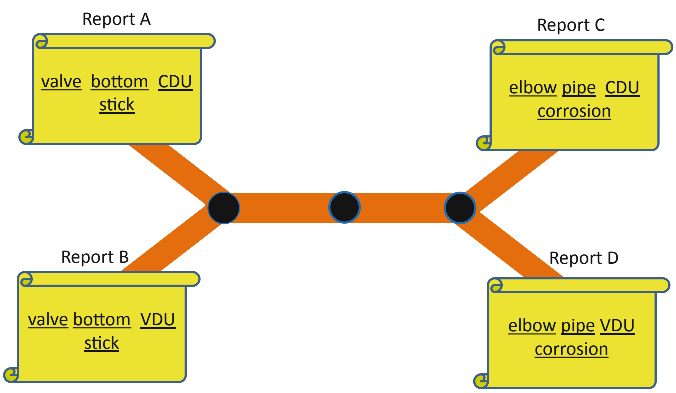


Fig. 3. A neighbor-joining unrooted tree reflecting distances of the reports shown in Fig. 2. Since Report A and B are closest, they are tied with only one knot. Contrary, report A and D do not resemble each other, so the route connecting them has more knots. (To keep simplicity and compactness of the graph, the length of each branch does not quantitatively reflect the distance calculated as Fig. 1.)

Let us see more realistic example. Figure 4 is the NJCT result of 4 469 ASRS incident reports of aviation. We stretched the three manually to make the shape easy to observe.

The tree has many major and minor branches. Each of them is a cluster of documents sharing common words. To some extent, we can regard this similarity of words as similarity of story on incident.

The most representative document of each pattern is attached at a top of each branch. This document is the closest one to the average of word distribution of each cluster. Reading such documents at the tops, we can understand the contents and meaning of the pattern braches. Figure 5 shows interpretation made by the author.

Without this document classification, we need unpractically large effort to read all reports. Otherwise, we get only dull result as Fig. 1.

The tree has five major branches, but the branches also have sub-branches inside of them. It is hard to explain that five is the only proper number of categories. Thanks to hierarchical clustering, we can decide clustering structure after observing the tree.

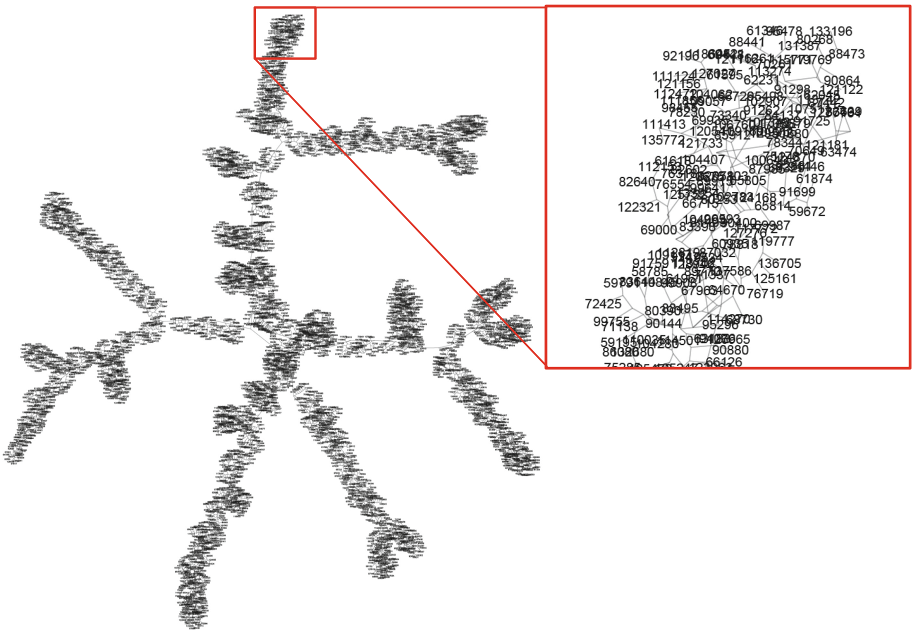


Fig. 4. The neighbor-joining classification tree of all 4 469 incident reports of Aviation Safety Reporting System (ASRS) in 2013. Each leaf is ID number of the report.

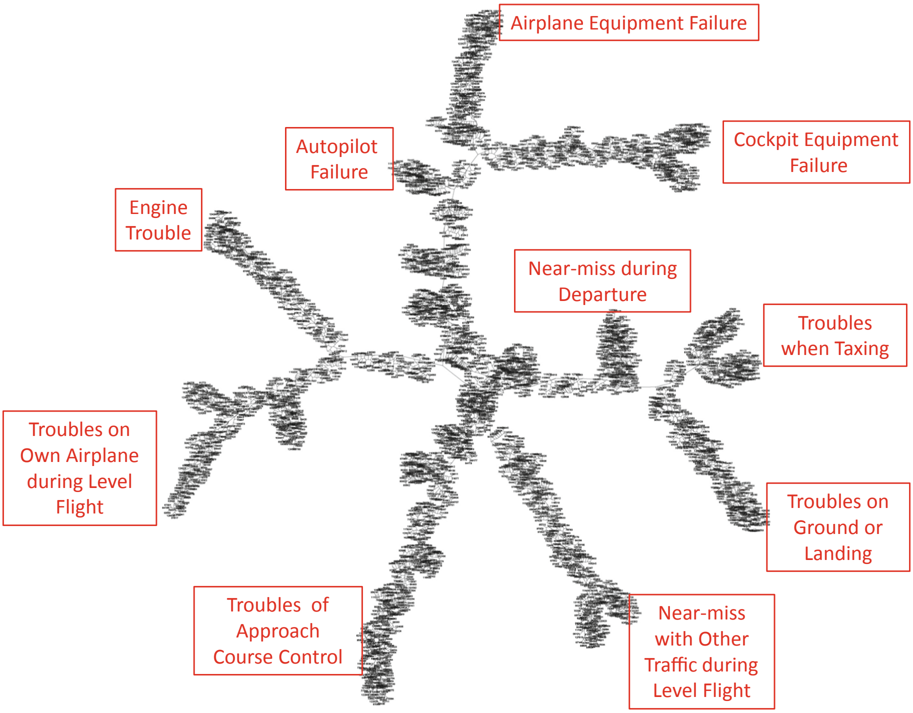


Fig. 5. Interpretations of major branches of the ASRS report tree. Reports in the same braches have mutual similarity in used words, so we can make the interpretations as written in the red boxes.

2.2 Classification of Words with Neighbor-Joining Classification Tree

Similar to classification of documents, words in the dataset can be classified in the same way of the neighbor-joining tree.

Distances among terms can be defined based on co-occurrences in the dataset. We use the example shown above again; here are four incident reports formatted as Bag-of-word shown in Fig. 6. This time we count numbers of co-occurrences among the words. For instance, the term of ‘valve’ and ‘bottom’ appear in two reports together. Observing such co-occurrences of words, we get a matrix data of the numbers shown in Table 1.

In general, co-occurrence implies similarity of meaning and/or existence of particular relationship. Therefore, distance between words should be defined as the inverse number of co-occurrence count.

In the same algorithm to build report classification tree, we can generate the tree of words shown in Fig. 7 based on the distance matrix of Table 1.

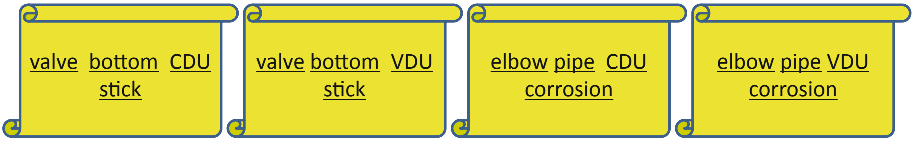


Fig. 6. Example of four incident reports described in Bag-of-Word form.

Table 1. Font sizes of headings. Table captions should always be positioned *above* the tables.

	valve	*bottom*	*stick*	*CDU*	*VDU*	*elbow*	*pipe*	*corrosion*
valve		2	2	1	1	0	0	0
bottom	2		2	1	1	0	0	0
stick	2	2		1	1	0	0	0
CDU	1	1	1		0	1	1	1
VDU	1	1	1	0		1	1	1
elbow	0	0	0	1	1		2	2
pipe	0	0	0	1	1	2		2
corrosion	0	0	0	1	1	2	2	

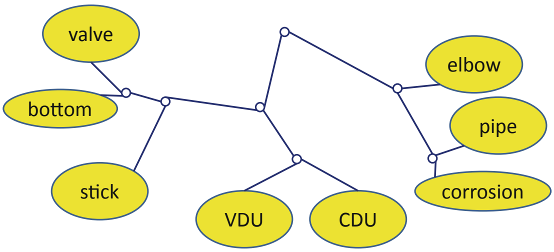


Fig. 7. A neighbor-joining unrooted tree reflecting word co-occurrences of Table 1.

Figure 8 is the word tree result generated from the real data of 4 469 ASRS reports.

Each branch has particular set of terms. For example, we find a group of words in a branch located on right and upper position (Fig. 9); it has terms of *'TCAS' (Traffic Collision Avoidance System), 'ATC' (Air Traffic Control), 'traffic', 'other aircraft', 'clearance', 'course', 'altitude', 'climb',* and *'descend'*. Those are obviously words describing the scene of collision avoidance. Likewise, other branches of the tree have more or less consistency of meaning among their words, so we can spot typical scenes in the stories of document dataset.

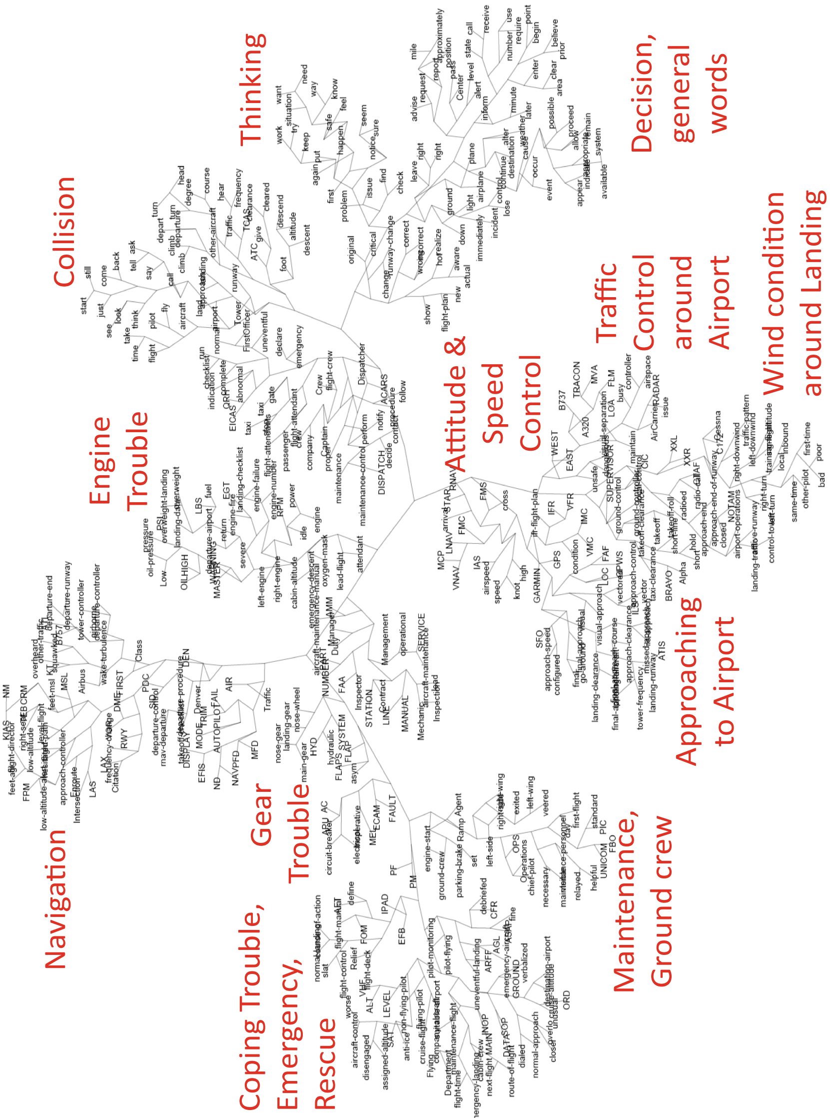


Fig. 8. The neighbor-joining unrooted tree of words generated from ASRS incident reports. Read labels are interpretations of major branches.

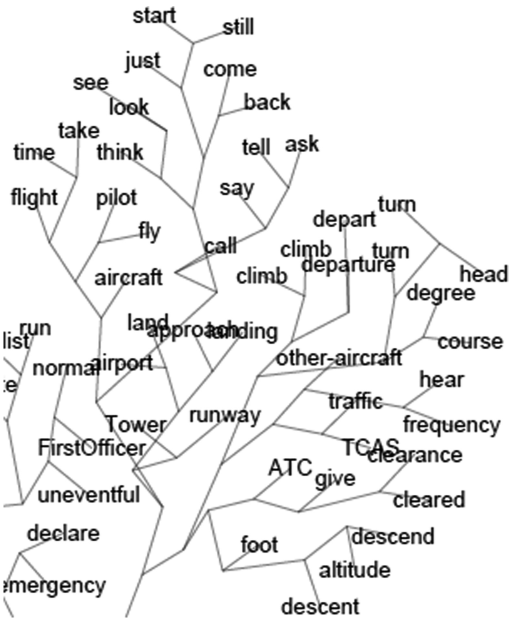


Fig. 9. A branch of the word tree (Fig. 8) with terms about risk of collision of aircrafts.

3 Conclusions: Understanding Mechanism of Human Error

The purpose of this paper is to show the effectiveness of the document classification methods to study incident report databases, which are very important to learn human errors. Although this method has been famous as a basic document classification method, application for safety engineering has been left untouched for long years.

In future work, refining ontology is at the first priority. During the analysis, we often find wrong classification; sometimes reports sharing very similar meaning are putted with long distance. We can detect cause of the failure by revising co-occurrence matrix; perhaps we failed on handling synonyms or hypernyms whose similarity of meaning should be taken into account even if the words are different. We may have counted meaningless or too general words also, and those numbers effected as noise to result classification failure. Revising the words to be observed, we can improve the classification result and find important terms to understand incidents.

Moreover, we should consider about meaning relationship among the term. For example, sentences of *"A sold X to B."* and *"B bought X from A."* have the same meaning. To understand this coincidence, we have to analyze grammatical information. This is difficult for present state-of-art of natural language processing in general, because our language has ambiguity greatly.

Everybody wants to know causes of human errors to understand mechanism of failure. If the text dataset has information about temporal order of events, we can have some clues to find the causality.

Most of ASRS reports consist of series of sentences aligned in temporal order; the first sentence usually states about the beginning, and the last sentence describes the situation at the last. Using this advantage, we can detect causality to some extent [4].

ASRS reports are exceptionally convenient, and sentences of text database in general may not be aligned in such causality order. In such case, we have to pay attention of meaning of terms. Some words stand for result of incident, and other represent preconditions of mistake. Watching roles of words, we can get some information on causality and mechanism of human error accidents.

References

1. Rak, R., Rowley, A., Black, W., Ananiadou, S.: Argo: an integrative, interactive, text mining-based workbench supporting curation. Database **2012** (2012)
2. Connell, L.: NASA Aviation reporting system, NASA (2011)
3. Feldman, R., Sanger, J.: The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, Cambridge (2007)
4. Nakata, T., Sohrab, M.G.: Detection of typical progress patterns of industrial incidents by text mining technique. In: Advances in Human Error, Reliability, Resilience, and Performance, pp. 221–229. Springer, Heidelberg (2017)

Human Resilience in Aviation Systems

Human Reliability Research Needs for Long-Duration Spaceflight

Ronald L. Boring[1(✉)], David I. Gertman[2], and Thomas A. Ulrich[1]

[1] Idaho National Laboratory, PO Box 1625, Idaho Falls, ID 83415, USA
{ronald.boring, thomas.ulrich}@inl.gov
[2] Boise, ID 83716, USA
david.gertman@gmail.com

Abstract. This paper considers human reliability analysis (HRA) for long-duration space missions. HRA has previously been used in the space program to support ground-based operations, Space Shuttle and International Space Station missions. New missions, such as a prospective mission to Mars, present new contexts for human spaceflight, including longer distances and durations, and different gravity levels. For HRA to be used to inform and prevent hazards, new research and data gathering are needed to understand the psychological and physiological aspects of astronautics. This paper outlines several areas of research that would support HRA for long-duration spaceflight. HRA methods must be adapted for space, which requires new data to inform and validate human error categorization and quantification.

Keywords: Human reliability analysis · Spaceflight · Mission to Mars · Research needs · Astronautics

1 Introduction

Human reliability analysis (HRA) is the systematic process of identifying, modeling, and quantifying human error. HRA is typically used as part of a comprehensive probabilistic risk assessment (PRA) to consider the implications of the human to the overall risk, including harm to humans, systems, or the environment [1].

The National Aeronautics and Space Administration (NASA) has for many years considered HRA as part of its risk analyses for the Space Shuttle and International Space Station. HRA is considered as part of the Loss of Crew (LOC) and Loss of Mission (LOM) analyses, using a variety of HRA methods [2]. Methods like the Technique for Human Error Rate Prediction (THERP) [3] and Cognitive Reliability and Error Analysis Method (CREAM) [4], respectively, model the action/execution and cognitive components of spaceflight as well as ground support activities.

HRA methods like THERP and CREAM were developed primarily for nuclear power applications. There are considerable and obvious differences between the main control rooms of nuclear power plants and the dynamic phases of spaceflight. Nuclear

Independent Consultant

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 289–297, 2020.
https://doi.org/10.1007/978-3-030-20037-4_26

power represents a largely stationary process control activity, whereas spaceflight encompasses different types of activities ranging from launch, to orbiting, to docking, to landing activities (among many more) and includes significant changes in the environment (e.g., microgravity and isolation) that may physiologically or psychologically impact the astronauts' well-being and ability to perform required tasks.

Such differences between ground process control in a control room and space-based activities are amplified as NASA undertakes long-duration space missions like the mission to Mars. New types of spacecraft are being created like the Orion capsule, Gateway Lunar Orbiter, the Crew Transport Vehicle, and the Mars Lander. Each of these presents new types of environments and contexts to consider—the impact of which is accentuated by the long duration of such a mission. The long duration will potentially impact astronaut performance beyond currently modeled elements. Additional hazards are introduced simply due to the distance between Earth and the astronauts, which brings delays in communication channels, logistic concerns with supply chain and longevity of essentials like medicine and food, and unavailability of backup crews and emergency return-to-earth vehicles. Success at such missions poses significant engineering challenges, but it also requires human-centered and risk-informed engineering to mitigate any risks to the crew. HRA is not simply a tool to ensure acceptable margins on astronaut performance; rather, it is also a necessity to the design of environments, vehicles, and missions that will ensure the well-being of the crew.

Long-duration spaceflight poses two particular challenges:

1. The increased gap between the purposes for which HRA methods were originally developed, and
2. The unknown nature of many performance changes over time and distance in space.

To address these challenges, additional research is necessary. This paper outlines activities that may improve the suitability of HRA for long-duration spaceflight. HRA is a requirement that ultimately serves to ensure the protection of crew, systems, and mission. To meet this requirement, HRA must be refined and adapted to reflect our emerging understanding of human hazards and risks in deep space.

2 Adapting HRA to Long-Duration Space Missions

Space-based HRA, despite being used for many years in space programs, is still very much in its infancy. To provide a glimpse of the work required to adapt HRA to long-duration space missions, the following list provides some of the challenges for HRA still to solve:

- The relative complexity and ambiguity for scenarios, actions, and equipment reliability after prolonged reduced gravity and higher-than-Earth radiation levels may initially make space HRA quantification difficult, because data to match these contexts is not available. Data must be extrapolated to these first-of-a-kind situations, or it must still be collected.

- There is a challenge in understanding the full range of effects of musculoskeletal, visual, vestibular, psychomotor, psychological and neurological effects upon performance and translating that information to qualitative and quantitative HRA. Current HRA methods lack the scope to account for these factors. Are current performance shaping factors (PSFs) adequate, or do new ones need to be developed?
- The ability of the analyst to extrapolate performance findings including reliability data from low earth orbit missions (e.g., International Space Station) to long duration missions (e.g., greater than one year) may prove difficult.
- Repairing equipment, including metallurgical tasks such as working with molten materials at low gravity, represent risk important tasks not currently modeled and quantified by HRA, and existing HRA data and insights may simply not fit.
- The expected introduction of advanced robotics will require that HRA address human-automation interaction for hazardous operations, including medical intervention and robotic assisted maintenance of life support systems in radiological or physically dirty environments. HRA will need to consider the roles, requirements and responsibilities for crew members and automation and their associated failure modes. To date, the effects of automation on performance are not well captured in HRA methods and practice.
- The medical system deployed in low Earth orbit for the International Space Station is designed to enable a "stabilize and transport" concept of operation. In this paradigm, an ill or injured crewmember would be rapidly evacuated to a definitive medical care facility on Earth, rather than being treated for a protracted period in orbit. Mars exploration class missions are quite different in that they will significantly delay or prevent the return of an ill or injured crewmember to Earth. It represents new modeling challenges for HRA to consider long-term incapacitated crew members due to injury. While it is presumed such an occurrence would be a low probability event, the possibility requires potential duplication in staffing or increased use of automated systems to ensure mission continuity.
- Long-duration missions will encounter novel situations, if for no other reason than travel to Mars will be first-of-a-kind undertaking. These unprecedented situations present challenges for HRA. By definition, response for novel situations is almost under-proceduralized because being overly prescriptive can result in brittle situations where performance degrades at the edges. The question is to what extent HRA can and should model novel, unexampled events.
- HRA for maintenance during long duration missions should reflect the type of maintenance to be performed and crew involvement. For example, there will be some degree of autonomous maintenance, some degree of manual corrective maintenance by the crew, and some combination of preventative maintenance as performed by autonomous and human processes.

The bottom line is that HRA has not been used for such applications because such domains are unprecedented. Extrapolating HRA to account for these applications certainly goes beyond the scope of all existing HRA methods. HRA methods must therefore be adapted. Adaptation requires data, some of which has not yet been gathered and other of which has not been disseminated to a form suitable to inform

HRA. Even HRA methods that are flexible enough to consider novel domains must still be validated, which again requires research and data. Evolving HRA for long-duration space missions requires both research to understand and model novel human performance domains and refinement of the method and supporting guidance for analysts.

3 Research Needs

3.1 Additional HRA Methods

NASA previously decided to downselect to four HRA methods [2]: THERP and CREAM, as previously discussed, and Standard Plant Analysis Risk-Human (SPAR-H) [5] and Nuclear Action Reliability Assessment (NARA) [6]. It may be worth revisiting this downselection in light of newer HRA methods, refinements and new insights on existing HRA methods, and the changing scope of NASA missions. It would be fruitful to consider the extension of HRA methods to meet different mission needs. For example, THERP may benefit from refinements to its tables of execution errors, CREAM and SPAR-H may appreciate a revised list of PSFs, and NARA may improve with space-specific task types. For NASA applications, it needs to be made clear when and why to use specific HRA methods vs. using a one-size-fits-all HRA method. Multiple HRA methods will benefit from refinements, including all four of the NASA-accepted methods. Additionally, it is reasonable to assume other HRA methods may have particular suitability to emerging missions.

3.2 Bayesian Approaches for PSF Selection

Research by Groth and Mosleh [7] introduced a method to develop a model for building Bayesian Belief Networks (BBNs) of interactions between PSFs, or Performance Influencing Factors (PIFs) as they are also referred to within their study. These networks offer extensive possibilities for combinations of PSFs linked to human error events. The demonstrated method uses factor analysis to discover patterns of variance and suggests Bayesian techniques to link these patterns to human error. The researchers provided an example model with commercial nuclear power plant applications, but predicted that the generic nature of most PSFs afforded the opportunity for potential applications to other industries relying upon human-machine interaction tasks. For applications such as space exploration, the researchers suggested that a set of PSFs could be adapted to contain factors related to reduced gravity. Their research provides guidance for adapting the models to these specific applications.

Further, Groth and Mosleh highlighted the lack of standard terminology used in HRA methods, asserting that various HRA methods "may use different terminology for similar concepts or similar technology for different concepts." Within the method, the researchers established a set of central guidelines for creating a PSF set as well as a hierarchy of PSFs developed solely for causal modeling. The tiered PSF classification provides an "application neutral, clearly defined, non-overlapping set of factors for modeling use." The fully expanded set of PSFs can be used to carry out qualitative analysis, while the collapsed version can be used for quantitative analysis.

This work was instrumental in Mindock's research on space PSFs [8]. Mindock developed a contributing factors map (CFM) that attempts to comprehensively describe factors influencing human performance and reliability. The factors are organized into a five-layer hierarchy that broadly segregates socio-technical domains (i.e., operations, vehicle design, and human) and then domain categories. For example, the *operations* domain has categories of *organization*, *training*, *team*, and *task specific characteristics*. Each category is segregated into functional groupings, contributing factors, and underlying factors. For example, within the *team* socio-technical domain category, there is a functional grouping for *crew collaboration quality*, and this grouping contains a contributing factor for *cooperation* with underlying factors of *interpersonal leadership*, *collective efficacy*, *interpersonal conflict level*, and *team cohesion*. Mindock's CFM not only hierarchically and graphically organizes over 270 factors but describes and provides references for each. The goal of her research was to use this map to as an early step toward defining a set of PSFs for spaceflight applications. Mindock's initial work must now be expanded to arrive at the appropriate list of PSFs for long-duration space missions.

3.3 Phased PSFs

The effects of PSFs like microgravity will vary over the duration and phases of the missions. As such, a single set of PSF multipliers to account for universal effects on performance is likely not appropriate across the phases of a space mission. Especially long-duration missions may accentuate the deleterious effects of some PSFs. To account for these varying effects, it is necessary to consider PSFs by phase. Three approaches should be considered for treating the PSF multipliers by phase:

- *Correction factors*, in which the calculated HEP is adjusted according to the phase. For example, a multiplier greater than 1 might be applied to the calculated HEP for descent phases, thereby increasing the HEP. Currently, no guidance or precedence is available for such a correction factor, and values would need to be validated. Moreover, a concern is that not all PSFs change uniformly across phases. However, the correction factor could take the form of weighting table that acts differently on each PSF multiplier.
- *Different PSF multiplier tables* could be used according to phase. This approach, similar to the different PSF worksheets for SPAR-H depending on At Power vs. Low Power and Shutdown [5], would simply change the multipliers and accompanying guidance for each phase. Like with correction factors, these multipliers would need to be carefully validated.
- *PSF distributions* could be provided instead of single-point multiplier values. In this approach, no single value would be provided, but rather the value would vary depending on where in the mission the analysis was centered. This approach has not, to date, been employed in HRA methods and would likely necessitate the use of software tools instead of worksheets to properly use the distribution. Note that a distribution assumes a degree of continuity, often linked by intervals of time. It is not clear if the mission phases are continuous or rather represent categorical changes. Or, perhaps there are categories of distributions corresponding to changes

in PSFs over time once in a mission phase. For example, the negative effects of gravity during the landing phases of a mission will gradually decrease. Once in the category of descent and landing, a specific distribution would be applied to the astronauts' performance. This distribution might be different than one for orbit.

The optimal approach to treating PSFs across phases needs to be established, and the multipliers for those phases need to be derived and validated.

3.4 Additional Data to Inform Space HRA

NASA conducts human performance studies as part of the human research program. NASA's Human Health and Performance (HH&P) Directorate is dedicated to developing methods and technologies to support safe, productive human space travel. To date, the human performance data and studies span NASA space experience, which includes low-earth orbit and lunar mission environments. Sending humans to Mars represents a number of considerable and diverse challenges for sustained human space travel, including the long-duration spaceflight required to reach Mars and the complicated landing and exploration phases. The various human research program efforts all aim to support human performance data collection to enable this Mars mission with minimal risks of loss of crew or mission.

Three prongs of research are foreseen to support the need for additional data to inform space HRA:

- Data mining of existing HH&P data to determine overall effects of space PSFs on astronautic performance. This approach involves a systematic exploration of data through targeted data extraction to answer specific research questions and populate needed human error probabilities. Data mining would involve use of a broad range of data extraction tools to perform meta analyses on available operational and research data.
- Identification of HRA data needs to be answered by the human research program. While an initial focus on extracting available data will surely prove productive, it is also likely that HRA can identify vulnerabilities for which current research does not provide answers. This effort might entail working closely with HH&P researchers to develop new lines of research that could better inform HRA quantification and redress vulnerabilities.
- Finally, data are needed to model dependence better. In simplest terms, dependence suggests that error begets error. In reality, a more complex relationship is typically at play, in which a vulnerability serves as a common cause opportunity for error across a protracted period of time. HRA methods like CREAM do not directly treat dependence, whereas THERP provides the most commonly implemented dependence correction. Dependence can significantly affect the outcome of an event and of the performance of tasks in sequence. Yet it is little understood, even in ground-based HRA. HRA research might use available historic data from simulator runs and spaceflights to construct realistic models of dependence that may be extrapolated to new domains.

3.5 Consideration of Advanced Technologies

Long-duration spaceflight will require the development of new technologies, potentially ranging from new requirements for medicines with longer shelf lives, to new interstellar launch technologies, to new types of sustainable sustenance, to self-repairing systems, to new forms of autonomous control, to augmenting human performance through robotics. Ultimately, as these technologies are developed, they will feature some form of interaction with astronauts. Not all present new hazards, of course. For example, the diminished potency of medications over time may not present as a vulnerability if new forms for stabilizing pharmaceuticals can be developed. Other technologies such as robotic exoskeletons to overcome gravitational forces upon entry into Mars' atmosphere present considerably new modes of interaction. HRA has historically not addressed space-specific applications, but it has also been married to a particular generation of control room equipment in nuclear power plants. New technologies are equally as important to performance as space-specific PSFs, and the effects of such equipment must be considered in the overall risk context. In many cases, new technologies will actually serve to enhance the abilities of the astronauts, and these performance enhancements must find a way to be credited in the HRA. However, new technologies also often introduce new error modes that must not be overlooked. Technologies present tradeoffs to performance, and HRA—whether space or ground based—must adequately consider them.

3.6 Dynamic HRA Modeling

The four preferred NASA HRA methods are static HRA methods, meaning they refer to worksheet or paper-and-pencil methods. In fact, none of these methods features a ready-made software implementation, although various aftermarket software tools have since been developed to assist analysts in using the methods. Even with software, the methods remain fundamentally static, meaning they analyze a fixed set of human failure events, typically as prescribed and defined in the static event tree used in the probabilistic risk assessment. An alternate approach involves creating a dynamic representation of the event in which the human error types and probabilities are automatically calculated by the software. Such a tool might be thought of as creating a virtual analyst or virtual astronaut [9], depending on perspective. Creating a virtual representation of the HRA affords a greater deal of possible risk exploration such as might be seen using Monte Carlo or Markov chain style iterations to explore various what-if scenarios. Such tools, when coupled with realistic physical world and system models, allow extrapolation to unexampled events. Given known data, it becomes possible to extrapolate how the performance might be expected to play out. Importantly, such models allow bounding of event outcomes by simulating worst, normal, and best case outcomes. Such modeling can, with the right tools, be automated, which would allow the dynamic HRA model to be used to test different outcomes even at design time to make better risk-informed design decisions.

Dynamic modeling tools are still research tools, but they are reaching a point of sufficient maturity and flexibility to allow them to be used for practical HRA applications. An effort to incorporate dynamic HRA modeling into space HRA would involve determining the best modeling environment to use, coupling those modeling tools to any physical and system simulations available for deep space missions, and running simulations with the combined models. These runs should ideally be validated against empirical data derived from astronaut-in-the-loop studies. Fortunately, modeled data can make very specific predictions that can be tested empirically. Dynamic models can identify potential red flags in predicted performance, which can then be investigated as part of the human research program.

4 Conclusions

HRA has historically been used for as-built systems, such as when applied to existing nuclear power plant control rooms as part of plant risk determination. The prospect of long-duration spaceflight opens up new domains for HRA. It is important to note that these domains represent equipment that hasn't yet been built and scenarios that haven't yet been specified. As such, much of the utility of HRA remains as a risk screening tool to help make design decisions. Space-based HRA is a tool not so much to capture inherent risk in a system but rather to determine the safest mission parameters. Research to support space HRA ensures mission success and safety by properly accounting for human considerations across the lifespan of the space mission.

Finally, it must be noted that referencing NASA in this paper simply serves as a convenient shorthand for any long-duration space missions. It does not imply any endorsement by NASA regarding the content or research direction presented in this paper, nor does it suppose any alignment to NASA research. Long-duration spaceflight undertaken by any governmental or private carrier must heed reasonable risk to the passengers of those space vehicles. The research outlined here should not be seen as a caution or discouragement to possible long-duration space missions. The data to support mission safety will certainly be gathered. This paper simply begins to identify some research that is possible to help adapt HRA to such applications.

Disclaimer
The opinions expressed in this paper are entirely those of the authors 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.

References

1. National Aeronautics and Space Administration: Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. Washington, DC (2011)
2. Chandler, F.T., Chang, Y.H.J., Mosleh, A., Marble, J.L., Boring, R.L., Gertman, D.I.: Human Reliability Analysis Methods: Selection Guidance for NASA. NASA Office of Safety and Mission Assurance, Washington, DC (2006)
3. Swain, A.D., Guttmann, H.E.: Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, NUREG/CR-1278. U.S. Nuclear Regulatory Commission, Washington, DC (1983)
4. Hollnagel, E.: Cognitive Reliability and Error Analysis Method (CREAM). Elsevier, Amsterdam (1998)
5. Gertman, D., Blackman, H., Marble, J., Byers, J., Smith, C.: The SPAR-H Human Reliability Analysis Method, NUREG/CR-6883. U.S. Nuclear Regulatory Commission, Washington, DC (2005)
6. Kirwan, B., Gibson, H., Kennedy, R., Edmunds, J., Cooksley, G., Umbers, I.: Nuclear action reliability assessment (NARA): a data-based HRA tool. Saf. Reliab. J. **25**(2), 38–45 (2005)
7. Groth, K., Mosleh, A.: A data-informed model of performance shaping factors for use in human reliability analysis. Ph.D. Dissertation. Center for Risk and Reliability, University of Maryland (2009)
8. Mindock, J.: Development and application of spaceflight performance shaping factors for human reliability analysis. Ph.D. Dissertation, Department of Aerospace Engineering Sciences, University of Colorado at Boulder (2012)
9. Boring, R.L., Gertman, D.I., Tran, T.Q., Gore, B.F.: Framework for the application of MIDAS for modeling control room crew performance at nuclear power plants. In: Proceedings of the 52nd Annual Meeting of the Human Factors and Ergonomics Society, pp. 930–934 (2008)

A Management Science Perspective on Resilience

Kimberly Stowers[1(✉)], Yash Mehta[2], Youjeong Huh[1], Lisa Brady[1], and Nicholas Kasdaglis[3]

[1] Department of Management, Culverhouse College of Business, University of Alabama, Tuscaloosa, AL, USA
kstowers@cba.ua.edu, yhuhl@crimson.ua.edu, llbrady@crimson.ua.edu

[2] Department of Computer Engineering and Sciences, College of Engineering and Science, Florida Institute of Technology, Melbourne, FL, USA
ymehta2016@my.fit.edu

[3] MITRE Corporation, McLean, VA, USA
nkasdaglis@mitre.org

Abstract. The practice of resilience within many industries, especially aviation, oil, and nuclear power, focuses on practice at the individual, system, and organizational level. At the individual and system levels, near-miss events, mishaps, and incidents are investigated to determine causal factors to proactively rectify the problems. At the organizational level, resilient behavior begins with a strategic model that encompasses management commitment to safety. The field of resilience engineering has made great strides in understanding resilience at each of these levels. However, work in this field sometimes fails to consider additional perspectives provided by other fields that may fill gaps in our knowledge. The purpose of this paper is to consider how theories from management science may augment ideas in resilience engineering to allow for a more complete understanding of resilience as a multi-level phenomenon.

Keywords: Systems engineering · Resilience engineering · Management

1 Introduction

On October 31st, 2014, the VSS Enterprise, Scaled Composite/Virgin Galactic SpaceShipTwo suborbital spaceplane was launched from WhiteKnightTwo carrier aircraft at 50,000 feet to fly over the Mojave Desert in California. Approximately 30 s after the release and the ignition of the rocket motor, the Enterprise broke apart resulting in the death of the co-pilot. Thus began an 8-month investigation by the National Transportation Safety Board [1], which showed that the accident occurred due to the co-pilot unlocking the wing feather mechanism (which folds part of the wings during reentry to reduce drag) prematurely at Mach 0.8 instead of Mach 1.4. By unlocking the feather mechanism before the Enterprise was completely supersonic, the pilot subjected the feather mechanism to aerodynamic forces so great that the vehicle itself bent and broke apart [1].

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 298–309, 2020.
https://doi.org/10.1007/978-3-030-20037-4_27

The investigation highlighted a wide-range of issues including gaps in government safety regulations, organizational safety, human error, and human-systems integration. Ultimately, Scaled Composite failed to address issues related to resilience, or "the ability of an organization (system) to keep, or recover quickly to, a stable state, [and] continue operations during and after a major mishap or in the presence of continuous significant stresses" ([2], p. 275). A system, within the context of this paper, is the collection of hardware, software, humans, and procedures that engage in specific functions and operations within an organization [3]. An organization, within the context of this paper, is a social grouping or construct with a high degree of structure, formalism, and specific goals, such as businesses and institutions [4]. When viewing resilience at both system and organizational levels, Scaled Composite failed to build a resilient organization that could withstand and adapt to lapses in safety and mitigate the effects of a hazard experienced by the organization and the system. At the system level, designers missed opportunities to consider human-systems integration and develop cockpit interfaces that could reduce chances of human error. At the organizational level, human factors and human errors were not prioritized in safety plans and hazard analysis.

The gaps and failures leading to this accident highlight the complexity of resilience, including its existence as a multi-level phenomenon across individuals, systems, and organizations [5]. A great deal of work has been done to understand resilience at each of these levels, including how they are connected [6, 7]. The field of resilience engineering in particular has led this effort, with increasingly effective models and frameworks of resilience that take into account the temporal complexity of resilience. Work in resilience engineering has necessarily deviated from sociological and organizational perspectives (e.g., "normal accidents" [8, 9] and "high reliability organizations" [10]) which were touted as confusing and ambiguous [11]. However, as a result, work in resilience engineering has sometimes failed to consider evolving perspectives provided by these other fields that may fill gaps in our knowledge. Thus, the purpose of this paper is to expand the field of resilience engineering by providing a discussion of resilience from the perspective of management science, which considers resilience primarily at the individual and organizational level. First, we will review the evolution of resilience engineering in human factors. Next, we will examine how resilience is characterized in management science. Finally, we will draw parallels between the work done in these fields, while delineating areas in which management science can provide unique perspectives to augment the field of resilience engineering.

2 Resilience: A Human Factors Problem

The concept of resilience engineering evolved in parallel and due to the evolution of safety thinking over the generations. Whereas ideas from reliability engineering and management theory initially allowed for the development of safe systems, resilience engineering took this to the next level by developing the ability of such systems to adapt [2]. As concepts of safety analysis and accident causation theories changed, so did resilience engineering. Initial views of resilience were impacted by Jens Rasmussen's work in safety thinking and its placement within the larger context of a system [12] as well as James Reason's model of why accidents occur [13]. This foundation paved the way for newer models in system-oriented safety analysis influenced by the works of scholars such as Hollnagel and Leveson. Resilience thus evolved from being linear and sequential to a more complex system-based process [14–16]. In system-based concepts of safety and resilience, simple hazards do not pave a linear path to an accident. As such, a static safety structure that ensures adherence to safety rules does not predicate the resilience of a system or organization. Rather, resilience is a dynamic process that adapts to variances in *how safety is perceived and performed* within an organization to ensure that an organization:

1. Has the necessary layers of safety to absorb hazards and mitigate its impact
2. Has the means to adapt to the shift in safety perception and practice to minimize adverse impact of how safety practices vary.

Rasmussen's work on safety introduced human performance classification in the form of Skills, Rules, and Knowledge-based tasks/errors (*'SRK'* model [12]). This human performance model was used in Reason's *Swiss Cheese* model [13] to explain human error, as well as in Shappell and Wiegmann's Human Factors Analysis and Classification System (*HFACS*) model [17], which essentially expanded the *Swiss Cheese* model to include more subcategories, including human error classification based on the SRK model. Since these latter two models are static and view accidents as direct sequences of events, they may not adequately address the complexities and emergent phenomena associated with accidents.

In contrast, both Hollnagel and Leveson approached safety as a systemic model where there are non-linear causations for accidents and where hazards may be indirectly linked to possible events. Within Hollnagel's work on safety is the concept of *emergence* and *variability* [18]. Emergence is the concept that accidents can occur due to an ephemeral event or hazard arising from complex interactions in socio-technical systems, whereas variability is the variation in performance and safety practices from one (socio-technical) function or entity to another. If the variability is sufficiently large such that safety constraints are violated, then an accident may occur. Critical to dampening variability is the adaptation and control of variability itself. Leveson asserts that safety is a control problem and accidents occur when controls over hazards are not sufficient [19]. Control problems arise when:

1. Control actions are not provided or followed
2. Incorrect or unsafe control action is provided
3. Correct control action is provided too early or late
4. Correct control action is stopped too soon.

In line with viewing resilience as a more complex phenomenon, researchers have likewise sought to examine resilience with a layered approach. Thus, a key aspect of resilience is the *defence-in-depth* concept [14] where there are multiple layers in an organization to dampen the impact of error on the organization or a system within the organization. This concept can be expanded to encompass multiple layers of defense spanning across the organization, industry bodies, government, and related safety regulations to ensure resilience at each level can reduce the impact of hazards or accidents. This concept was previously addressed by Rasmussen [20]—and later Leveson [16]—by considering the socio-technical hierarchy of a system beginning at the top with the government and precipitating down to operators, agents, components, etc. Connecting each layer of the socio-technical system allows for some control of input and output to ensure that a system is safe and resilient.

To this end, each organization or system requires a highly nuanced approach to making the system more resilient. Furthermore, the practice of resilience within many industries, especially aviation, oil, and nuclear power, focuses on the individual and system level as well as at the organizational level [2]. At the individual and system levels, incidents are often investigated to determine causal factors and prevent future mishaps. At the organizational level, resilient behavior begins with a strategic model that guides organization behavior. Hollnagel, Woods, and Leveson [2] define these as seven themes in a highly resilient organization:

1. Top-level commitment
2. Just culture
3. Learning culture
4. Awareness
5. Preparedness
6. Flexibility
7. Opacity

How these themes are structured in the model of organizational resilience and how they are linked to system and individual-level resilience is not always clear. Although a great deal of work has been done to link organizational resilience to system and individual-level resilience [6, 7], efforts to apply such work to real-world systems are often (necessarily) reductionist. As a result, gaps in our understanding of individual resilience remain. Such gaps may ultimately contribute to failures, especially in tightly coupled, complex systems where organizations may struggle to accommodate all risks that emerge. Although failures in such systems may be conceptualized as unavoidable or "normal" [8, 9], it is ultimately up to the organization to account for limitations in the individual, technology, and system-level characteristics. In this vein, we argue that it may be helpful to consider individuals in a more holistic way than has been done previously in the field of resilience engineering. In particular,

1. How does the development of resilience occur in individuals separate from the socio-technical system?
2. How can organizations account for these differences to maximize individual and system-level resilience?

We will attempt to answer these questions by considering resilience from the view of management science, as discussed next.

3 Resilience: A Management Science Perspective

A recent review article on resilience in business and management research indicates that resilience has been conceptualized and operationalized differently across studies, possibly because resilience research has been highly context-dependent [21]. Discussion of different usage of the concept across disciplines is beyond the scope of this review. However, we will consider resilience in terms of micro-organizational behavior (at the individual and system level) and macro-organizational behavior (at the organization and strategy level). These perspectives may augment each other, allowing for the connection of individual and system-level resilience to organizational resilience. We explore each of these perspectives in depth, and then discuss how they relate to each other, as well as how they relate to perspectives in resilience engineering.

3.1 Micro-organizational Resilience

Within the micro-organizational behavior literature, resilience is an individual-level phenomenon with two distinct streams of research stemming from differences regarding its definition. While some researchers in this area view resilience as an individual *trait* that someone does or does not possess, others refer to resilience as a *behavioral process* that individuals demonstrate in their reactions to adverse events (e.g., stress). From the trait perspective, resilience is the ability to resist being damaged or deformed by traumas or destructive forces. From the behavioral process perspective, there are generally three acknowledged patterns that reflect resilience: (1) positive functioning under stressful conditions, (2) recovering from stressful conditions, and (3) personal development in the face of stressful conditions [22–24]. Researchers studying resilience as a behavioral process examine how individuals handle adverse events as well as the patterns of well-being over time that determine whether or not someone will be resilient when faced with stress [25, 26].

Developing Resilience. An individual's developmental origins and previous life experiences can play a role in developing one's level of resilience. Research indicates that events in both childhood [27] and adulthood [28], as well as the ways in which individuals cope with and respond to those events, can contribute to resilient behaviors. However, there is a lack of consensus regarding the role of prior traumatic or adverse experiences. Do they weaken or strengthen an individual? For some, a stressful event or hardship can be an opportunity to learn, grow, and better address the next adverse event. For those without previous experience successfully responding to stressors, however, they might be unable to maintain positive outcomes when an adverse event arises [29].

In addition to the development of resilience through one's life experiences, an individual's level of resilience may be enhanced through intervention and training programs. Resilience training programs have been implemented in educational, corporate, and government settings to enhance individual resilience. These training programs have utilized several different approaches [30], including teaching coping strategies, providing social support, encouraging a growth mindset, and developing emotion awareness skills. Although the results of these programs vary significantly [31], findings indicate that face-to-face training is more effective than computer-based programs. Additionally, because social factors interact with individual factors to affect resilience, multi-level intervention programs are most effective.

Protective Factors. The extent to which an individual is resilient is associated with specific risk and protective factors. Risk factors are characteristics associated with a decline in one's ability to recover from adversity. In contrast, protective factors are viewed as resources that an individual is able to draw upon in the face of adversity [32], and they include a wide range of types of resources, including psychological, social, and contextual resources [23]. Psychological factors, for example, may include those associated with recovery in facing adverse conditions. Social factors are those associated with an individual's social relationships (e.g., social support or social network). Contextual factors refer to social and physical environmental characteristics, such as institutional and economic capacities, which extend beyond an individual's capacities but affect individual-level resilience [33].

While research in micro-organizational behavior generally focuses on resilience as a person-level construct, resilience can be viewed as a property of an individual, a dyad, a team or group, an organization, or a broader collective (e.g., a culture or nation). Additionally, these different levels of resilience (e.g., individual, community, and national) are associated with one another [33]. Next, we discuss how management science examines and plans for resilience at the organizational (macro) level.

3.2 Macro-organizational Resilience

Resilience at the macro or organizational level can be considered an organization's positive adjustment under challenging conditions [21]. Organizations have structure, culture, and routine that have been developed, solidified, and deeply embedded since conception. These together act as a frame of reference for individuals; members of organizations learn and gradually internalize a series of interlocking routines and habituated behavioral patterns based on an organizational frame of reference (e.g., knowing who to report to, how to complete a task, and how to talk and behave). Therefore, when an unexpected event happens that deviates from organizational routine, individuals engage in *sensemaking* based on what they have learned about their organization. In this context, sensemaking can be defined as making sense of unexpected and ambiguous organizational events in a prompt and adaptive manner [34]. When the routine is severely disrupted (e.g., a series of unaccountable events lead to a disastrous accident or "*cosmology episode*" [35]), individuals suddenly and deeply feel that the universe is no longer a rational, orderly system. This unsettling feeling of incoherency elicits panic during the event of cosmology episode because both the sense of what is happening and the means to rebuild that sense collapse together. Since individuals may regress to their most habituated ways of responding when they are under pressure [36], they may not be able to promptly adapt their existing skills or knowledge in a creative way to respond to the unforeseen crisis. For this reason, previous research indicates that individuals' ability to engage in sensemaking may play a key role in their ability to respond to novel, ambiguous, and stressful situations under pressure [37, 38]. Thus, it is important to unpack macro-organizational factors that may serve as antecedents of individuals' ability to make sense of unexpected organizational breakdowns and to respond promptly. Individuals' ability to improvise solutions using their discretion is valuable because it can prevent small failures from escalating into high-consequence disasters [35]. To this end, two macro-organizational factors we would like consider are (1) *organizational culture* and (2) *organizational identity*. Although these constructs are closely related, we explore them separately to delineate the specific ways in which they contribute to organizational resilience.

Organizational Identity. Organizational identity defines what an organization is and stands for [39]. Organizations consider their identity when they are pressed to make difficult strategic choices with limited information. For example, when organizations have to choose specific products to market or use but cannot reach a conclusion due to lack of reliable information, discussion of organizational goals and values is likely to take place [39]. Questions about organizational identity may revolve around defining "who are we?" as a collective. A clear, deeply ingrained organizational identity serves as a guide for individuals faced with unexpected disruptions in organizational routine because identity guides and activates individuals' interpretations of an issue and motivations for action [40]. The process by which organizational members respond to events is equivalent to sensemaking in that the members attempt to look at the "organizational mirror (identity)" when they interpret, react, and commit to organizational actions. In the event of severe disruptions followed by devastating system failure, individuals need to improvise action plans in order to keep the system from falling apart. They also need to restore the system to order. In doing so, individuals use an "organizational mirror" as a frame of reference in making further action plans during the chaos where organizational routines (e.g., rules, report system, leadership) collapse [37]. Because of this, it is prudent for organizations to tune their organizational identity to emphasize that safety is the defining feature of their organizational identity. If organizations commit to establishing their identity as safety-prioritizing entities and the identity permeates into individuals' work routines, individuals also may start to consider safety as priority in carrying out tasks. Such organizations who effectively prioritize and plan safety to reduce the likelihood of failure are considered "high reliability organizations" [10].

Organizational Culture. Organizational identity is tied to organizational culture as cognitive maps like identity are closely aligned with traditions that make up culture [41]. Organizational culture can be broadly defined as a set of shared mental assumptions that guide interpretation and action in organizations by defining appropriate behavior for various situations [42]. For this reason, organizational culture can be considered a platform for organizations to alter and influence the way individuals think and act [42]. This process is often referred to as *sensegiving* [43].

According to seminal work on sensemaking by Weick [34, 35], resilient organizations may have specific practices ingrained in their culture: (1) *improvisation*, (2) *an attitude of wisdom*, (3) *respectful interaction*, and (4) *leadership*. In the spirit of improvisation, organizations may benefit from forming a culture in which individuals learn to think outside the box and use their own discretion when organizational routine is severely disturbed. Likewise, exercising the wisdom to deviate from outdated routines that may otherwise threaten organizational adaptation [44] can offer an opportunity for growth. This process begins with encouraging individuals to stay open-minded about questioning the "status quo" and finding new solutions in a changing environment. Additionally, fostering a culture where teams engage in respectful interaction by building trust, honesty, and self-respect [45] can increase the likelihood that such teams will engage in mutual adaptation and form creative solutions in the face of disruption. Finally, leaders can contribute to resilience in organizations by acting as *sensegivers* (see sensegiving above [43]; also [35]). Specifically, leaders that seem to have legitimate authority are more persuasive and therefore able to rebuild a failing system during crisis. For this reason, organizations need to make it clear that their leaders possess legitimate authority to command and that they are trustworthy. Organizations should also make sure they are developing leadership by providing quality training to reinforce this legitimacy.

3.3 Linking Micro- and Macro-organizational Resilience

There are a number of notable similarities between how management science conceptualizes resilience at the organizational (macro) level and at the individual or system (micro) level. Just as organizational resilience is a dynamic process that adapts to variances in how safety is perceived and performed at the organizational level, individual resilience is a within-person process that allows individuals to adapt to variances in their unique work environment. In a safety context, resilience at the individual level ensures that an individual has (1) the necessary knowledge and abilities related to safety to absorb hazards and mitigate adverse effects, and (2) the means to adapt to the shift in safety perception and practice to minimize negative outcomes.

Even with these similarities, macro- and micro-organizational perspectives on resilience have one stark contrast in the ideas they offer for developing and maintaining resilience. Macro-organizational perspectives, in focusing on resilience as a larger—and generally system-based—phenomenon, focus on the collective (organizational) performance involved in resilience. Although it is imperative to consider resilience at this level in order to account for all of its complexities, research that focuses on resilience primarily as a top-down process fails to take into account individual differences that may stand in the way. In particular, it is prudent to consider the following ideas of micro-organizational resilience (as reviewed in Sect. 3.1) and how they are linked to macro-organizational perspectives:

1. Individuals in an organization may have varying levels of capability in their abilities to express resilience
2. Individual resilience can be enhanced through training that improves capability in skills such as coping strategies, growth mindset, and similar skills
3. Organizations bolster protective factors of resilience by providing psychological, social, and contextual resources for individuals

From a macro-organizational perspective, resilience in organizations is linked to the complex process of sensemaking, which is facilitated by the development of a stable (1) organizational identity and (2) organizational culture. Within organizational identity, safety can (and should) become a focus. Within organizational culture, focus should be given to (1) improvisation, (2) an attitude of wisdom, (3) respectful interaction, and (4) leadership. It is in organizational culture that additional links to micro-organizational perspectives can be made. For example, to create a culture in which individuals have the courage and wisdom to improvise and embrace change, it might be helpful to consider individual differences in the capability to "think on one's feet" and "be open" to change. Likewise, when developing training programs to improve resilience as a safety construct, it might be helpful to also incorporate training that allows individuals to overcome any limitations (e.g., not having coping skills or a growth mindset) that might inhibit their abilities to respond adequately to system-level risks. Finally, leaders in organizations can instill respectful interaction through the implementation of psychological, social, and contextual resources for individuals, thus increasing the likelihood of successful system-level reaction to threats against resilience. The connections outlined here may also be useful to consider in resilience engineering, as discussed next.

4 Key Takeaways for Resilience Engineering

There are multiple points on which the fields of management science and resilience engineering agree, particularly when considering macro-organizational management science. For example, the importance of creating an organizational identity around safety is integral to the development of resilience in safety systems [2]. Additionally, regarding organizational culture, a focus on improvisation and wisdom to improve operational habits is similar to what has been argued in resilience engineering [2, 44].

However, management science offers unique perspective when considering the questions posed in the introduction of this paper:

1. How does the development of resilience occur in individuals separate from the socio-technical system?
2. How can organizations account for these differences to maximize individual and system-level resilience?

Two key macro-organizational management constructs (as reviewed in Sect. 3.2) can be considered as a starting point for answering question #2: (1) respectful interaction, and (2) leadership development. In particular, beyond the development of effective performance between individuals or individuals and technology, work can be done to create a culture in which individuals want to and enjoy working together effectively either to prevent a crisis or respond to one. Additionally, beyond developing leaders who are committed and can implement robust safety protocols, work can be done to develop leaders who will be appropriately respected when enforcing these protocols.

In supporting respectful interaction and leadership development, question #1 must additionally be answered. Here, ideas from micro-organizational management (as reviewed in Sect. 3.1) should be considered. First, the encouragement of qualities such as growth mindset and coping skills in individuals is just as important as the encouragement of safety practice, as it allows individuals and systems to think more critically and work together more effectively. Second, training individuals' ability to actively engage such skills (i.e., growth mindset and coping) is integral to their ability to effectively carry out safety protocols learned in training, especially if a system is experiencing a type of risk or failure not previously anticipated. Third, training interventions (especially those focused on safety) have the highest likelihood of success when implemented at multiple levels (individual, system, and organization). Finally, if it is true that organizations are the final arbiters of resilience [8, 9], then organizations should take the responsibility of providing psychological, social, and contextual resources that can assist such resilience.

To support this final point of providing resources that assist resilience, organizations can implement active control loops and organizational memory systems that remove complacency and create opportunities to reflect on past incidents and improve safety practices. Organizations can furthermore implement precedents that encourage safety training and consistent practices, such as rewarding safety practices. However, ultimately, it is important for organizations to provide an environment of psychological safety in which individuals feel that they are able to take interpersonal risks (e.g., reporting a safety incident [46, 47]). Individual contributions to system and organizational resilience will be effective only when the necessary resources are provided and adequate incentives and rewards systems are in place.

5 Future Work

The ideas presented herein offer a foundation for jumpstarting additional work in resilience engineering by considering research done in management science. We encourage readers to consider whether there are perspectives from additional fields that can further augment the field of resilience engineering. Additionally, we advocate for the development of additional research to consider whether the perspectives pulled from management science and other fields are indeed compatible with the ideas developed in the field of resilience engineering. By doing this, we can work toward an increasingly safe future.

References

1. National Transportation Safety Board (NTSB): In-Flight Breakup During Test Flight Scaled Composites SpaceShipTwo, N339SS. Washington, DC (2015)
2. Hollnagel, E., Woods, D.D., Leveson, N.: Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot (2006)
3. Maier, W.M., Rechtin, E.: The Art of Systems Architecture, 2nd edn. CRC Press, Boca Raton (2000)
4. Scott, W.R.: Organizations: Rational, Natural, and Open Systems, 5th edn. Prentice Hall, Upper Saddle River (2003)
5. Woods, D.: Engineering organizational resilience to enhance safety: a progress report on the emerging field of resilience engineering. In: Proceedings of the Human Factors and Ergonomics Society, vol. 50, no. 19, pp. 2237–2241. Sage, Los Angeles (2006)
6. Annarelli, A., Nonino, F.: Strategic and operational management of organizational resilience: current state of research and future directions. Omega **62**, 1–18 (2016)
7. Tillement, S., Cholez, C., Reverdy, T.: Assessing organizational resilience: an interactionist approach. Management **12**, 230–264 (2009)
8. Perrow, C.: Normal Accidents: Living with High-Risk Technologies New York. Basic Books, New York (1984)
9. Perrow, C.: Normal Accidents: Living with High Risk Technologies-Updated Edition. Princeton University Press, Princeton (2011)
10. Roberts, K.H.: New challenges in organizational research: high reliability organizations. Organ. Environ. **3**, 111–125 (1989)

11. Leveson, N., Dulac, N., Marais, K., Carroll, J.: Moving beyond normal accidents and high reliability organizations: a systems approach to safety in complex systems. Org. Stud. **30**, 227–249 (2009)
12. Rasmussen, J.: Models of mental strategies in process plant diagnosis. In: Human Detection and Diagnosis of System Failures, pp. 241–258. Springer, Boston (1981)
13. Reason, J.: Human Error. Cambridge University Press, Cambridge (1990)
14. Dekker, S., Hollnagel, E., Woods, D., Cook, R.: Resilience engineering: new directions for measuring and maintaining safety in complex systems, pp. 1–67. Final Report, Lund University School of Aviation (2008)
15. Hollnagel, E.: Flight decks and free flight: where are the system boundaries? Appl. Ergon. **38**, 409–416 (2017)
16. Leveson, N.: Engineering a Safer World: Systems Thinking Applied to Safety. MIT Press, Cambridge (2011)
17. Shappell, S.A., Wiegmann, D.A.: The Human Factors Analysis and Classification System—HFACS, Report DOT/FAA/AM-00/7, Department of Transportation, FAA, Washington, DC (2000)
18. Hollnagel, E.: An Application of the Functional Resonance Analysis Method (FRAM) to Risk Assessment of Organisational Change. No. SSM–2013-09, Swedish Radiation Safety Authority (2012)
19. Leveson, N.G.: Complexity and safety. In: Complex Systems Design & Management, pp. 27–39. Springer, Heidelberg (2012)
20. Rasmussen, J.: Risk management in a dynamic society: a modeling problem. Saf. Sci. **27**, 183–213 (1997)
21. Linnenluecke, M.K.: Resilience in business and management research: a review of influential publications and a research agenda. Int. J. Manag. Rev. **19**, 4–30 (2017)
22. Bonanno, G.A.: Loss, trauma, and human resilience: have we underestimated the human capacity to thrive after extremely aversive events? Am. Psychol. **59**, 20–28 (2004)
23. Bonanno, G.A.: Resilience in the face of potential trauma. Curr. Dir. Psychol. Sci. **14**, 135–138 (2005)
24. Masten, A., Best, K., Garmezy, N.: Resilience and development: contributions from the study of children who have overcome adversity. Dev. Psychopathol. **2**, 425–444 (1990)
25. Luthar, S.S., Cicchetti, D., Becker, B.: The construct of resilience: a critical evaluation and guidelines for future work. Child Dev. **71**, 543–562 (2000)
26. Becker, T.E., Ferry, D.L.: Profiles as a way of learning more about resilience. Ind. Organ. Psychol. **9**, 503–508 (2016)
27. Masten, A.: Ordinary magic. Am. Psychol. **56**, 227–238 (2001)
28. Ong, A.D., Bergeman, C.S., Boker, S.M.: Resilience comes of age: defining features in later adulthood. J. Person. **77**, 1777–1804 (2009)
29. Davydov, D.M., Stewart, R., Ritchie, K., Chaudieu, I.: Resilience and mental health. Clin. Psychol. Rev. **30**, 479–495 (2010)
30. Yost, P.: Resilience practices. Ind. Organ. Psychol. **9**, 475–479 (2016)
31. Vanhove, A.J., Herian, M.N., Perez, A.L., Harms, P.D., Lester, P.B.: Can Resilience be developed at work? A meta-analytic review of resilience-building programme effectiveness. J. Occup. Organ. Psychol. **89**, 278–307 (2016)
32. Hobfoll, S.E., Stevens, N.R., Zalta, A.K.: Expanding the science of resilience: conserving resources in the aid of adaptation. Psychol. Inq. **26**, 174–180 (2015)
33. Kimhi, S.: Levels of resilience: associations among individual, community, and national resilience. J Health Psychol. **21**, 164–170 (2016)
34. Weick, K.E.: Sensemaking in Organizations. Sage, London (1995)

35. Weick, K.E.: The collapse of sensemaking in organizations: the Mann Gulch disaster. Admin. Sci. Quart. **38**, 628–652 (1993)
36. Barthol, R.P., Ku, N.D.: Regression under stress to first learned behavior. J. Abnorm. Soc. Psychol. **59**, 134–136 (1959)
37. Beck, T.E., Lengnick-Hall, C.A.: Resilience capacity and strategic agility: prerequisites for thriving in a dynamic environment. In: Resilience Engineering Perspectives, vol. 2, pp. 61–92. CRC Press (2016)
38. Linnenluecke, M.K., Griffiths, A.: The 2009 Victorian bushfires: a multilevel perspective on organizational risk and resilience. Organ. Environ. **26**, 386–411 (2013)
39. Albert, S., Whetten, D.: Organizational identity. In: Cummings, L.L., Staw, B.M. (eds.) Research in Organizational Behavior, vol. 7, pp. 263–295. JAI Press, Greenwich (1985)
40. Dutton, J., Dukerich, J.: Keeping an eye on the mirror: image and identity in organizational adaptation. Acad. Manag. J. **34**, 517–554 (2001)
41. Child, J., Smith, C.: The context and process of organizational transformation-Cadbury limited in its sector. J. Manag. Stud. **24**, 565–593 (1987)
42. Ravasi, D., Schultz, M.: Responding to organizational identity threats: exploring the role of organizational culture. Acad. Manag. J. **49**, 433–458 (2006)
43. Gioia, D.A., Chittipeddi, K.: Sensemaking and sensegiving in strategic change initiation. Strat. Manag. J. **12**, 433–448 (1991)
44. Woods, D.D., Branlat, M.: Basic patterns in how adaptive systems fail. In: Resilience Engineering in Practice, pp. 127–144 (2011)
45. Campbell, D.T.: Asch's moral epistemology for socially shared knowledge. In: Rock, I. (ed.) The Legacy of Solomon Asch: Essays in Cognition and Social Psychology, pp. 39–52. Erlbaum, Hillsdale (1990)
46. Edmondson, A.C.: Psychological safety and learning behavior in work teams. Admin. Sci. Q. **44**, 350–383 (1999)
47. Edmondson, A.C.: Psychological safety, trust, and learning in organizations: a group-level lens. In: Kramer, R.M., Cook, K.S. (eds.) Trust and Distrust in Organizations: Dilemmas and Approaches, pp. 239–272. Russell Sage, New York (2004)

Alternatives for Scheduling Departures for Efficient Surface Metering in ATD-2: Exploration in a Human-in-the-Loop Simulation

Bonny K. Parke[1(✉)], Lindsay K. S. Stevens[2], William J. Coupe[2], Hanbong Lee[2], Yoon C. Jung[2], Deborah L. Bakowski[1], and Kimberly Jobe[1]

[1] San Jose State University Research Foundation, NASA Ames Research Center, P. O. Box 1 Moffett Field, Mountain View, CA 94035-0001, USA
Bonny.Parke@nasa.gov

[2] NASA Ames Research Center, P. O. Box 1 Moffett Field, Mountain View, CA 94035-0001, USA

Abstract. A Human-in-the-Loop (HITL) simulation was conducted to explore the impacts of various surface metering goals on operations and Ramp Controllers at Charlotte Douglas International Airport (CLT). Three conditions were compared: (1) Baseline, with no surface metering, (2) instructions to meet advisory times at the gate only, and (3) instructions to meet advisory times at the gate as well as the times at the scheduled taxiway spot, where aircraft are delivered to Air Traffic Control (ATC). Results showed increased compliance for taxiway spot times when compliance was first met for gate advisories. Instructing Ramp Controllers to meet advisory times at the gate improves spot time compliance and therefore surface scheduling predictability at CLT. Results also demonstrated there was increased compliance overall with gate and spot times in the second condition. This was likely due to higher Ramp Controller workload in the third condition.

Keywords: ATD-2 · Airport surface scheduling · Integrated Arrival, Departure, and Surface (IADS) traffic management · Human factors assessment in field operations · Surface metering · HITL simulation

1 Introduction

Airspace Technology Demonstration 2 (ATD-2) is an ambitious NASA project in coordination with the Federal Aviation Administration (FAA) and aviation industry partners that aims to integrate, for the first time, multiple concepts and technologies to reduce delays in the National Airspace System (NAS) [1]. The umbrella concept is Integrated Arrival, Departure, and Surface (IADS) operations. The FAA selected Charlotte Douglas International Airport (CLT) and surrounding air traffic control

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 310–320, 2020.
https://doi.org/10.1007/978-3-030-20037-4_28

(ATC) facilities as a test bed. CLT is the sixth busiest airport in the US based on the number of operations in 2018 and the busiest on the East Coast [2]. It is known as having the runway capacity of a major airport with the ramp capacity of a mid-level airport. Among the concepts and technologies being integrated into new tools for Ramp and Air Traffic Controllers at CLT is surface metering, as outlined by the FAA's Surface Collaborative Decision Making (Surface CDM) Concept of Operations [3].

A goal of metering, or time-based scheduling, on the surface is to reduce aircraft wait times in the departure runway queue, with its attendant fuel burn and emissions. Benefits of using surface metering include increased predictability in timing of aircraft movement across the airport surface, as well as economic and environmental benefits. At CLT, initial assessment of quantitative measures have shown a marked improvement in fuel burn and emissions in banks of aircraft that are metered, with an average reduction of 1,000 lb of fuel and 3,000 lb of carbon dioxide emissions per bank (CLT typically has nine banks per day) [4].

In the Surface CDM concept, surface metering works by redistributing some of the time that would otherwise be spent in the departure runway queue back to the ramp area, typically at the gate. Each aircraft subject to surface metering is assigned a time to be delivered to the "spot," or the point at which Air Traffic Control (ATC) takes control of the aircraft. During surface metering, Ramp Controllers deliver aircraft through the ramp to the spot in compliance with each aircraft's assigned spot arrival time. ATD-2's surface scheduler works to achieve the goals of the Surface CDM concept by generating target event times for aircraft to push back off gates, be delivered to the spot, and arrive at the runway. Scheduler-generated gate advisories are used by Ramp Controllers to aid them in deciding when to release aircraft from the gate and begin moving them across the ramp.

CLT has unique constraints that may impact the Ramp Controllers' ability to focus on delivering aircraft to the spot in compliance with an assigned spot time. Not only does CLT have the ramp capacity of a mid-level airport with limited ramp real estate, it also possesses extended areas within the ramp where traffic flow is reduced to a single-lane, so traffic can only flow one direction at any given time. Once an aircraft is inserted into the flow of traffic in the ramp, few, if any, options exist to adjust the aircraft's rate of travel to the spot.

A question exists as to whether Ramp Controllers can utilize alternative strategies to achieve the Surface CDM goal of spot time compliance during surface metering at airports like CLT with ramp area constraints, and the impact of those methods on surface operations and Ramp Controllers. This question was explored in a Human-in-the-Loop (HITL) simulation conducted at NASA Ames Research Center in June, 2018.

2 Method

2.1 Human-in-the-Loop Simulation

Conditions. To explore the impact of various methods of ramp operations for complying with times at the spots, the HITL compared three conditions at a simulated CLT:

1. Baseline: No metering/no advisory condition; normal operations,

2. TOBT-Only: Metering with Ramp Controller instructions to meet gate advisories for departures (Target Off-Block Times or TOBT) within a ± 2 min window, and
3. TOBT+TMAT: Metering with Ramp Controller instructions to meet gate advisories (TOBT) within a ± 2 min window for departures and, in addition, arrival times at the spots (Target Movement Area entry Times or TMAT) within a ± 5 min window.

HITL Overview. Each condition (i.e., Baseline, TOBT-Only, and TOBT+TMAT) was tested for three runs for a total of nine experimental runs. Three 70-minute traffic scenarios were used during the study; each scenario was used once in each test condition. Each of these traffic scenarios was duplicated from CLT live traffic recordings on different days during Bank 2 (the heaviest traffic bank at CLT) and thus had similar traffic loads. The three conditions were counterbalanced over time to counteract any order effects of training and experience. On each run, four Ramp Controllers rotated through each of the four CLT ramp positions: North, East, South, and West Sector. Ramp Controller training took place on the first morning and included familiarization with the ATD-2 IADS Ramp Traffic Console (RTC) interface and the three conditions.

Participants. Participants in the HITL included four experienced Ramp Controllers, one Ramp Manager, four ATC Tower Controllers, and one ATC Traffic Management Coordinator (TMC), along with eight Pseudo-Pilots and two Air Route Traffic Control Center (ARTCC) confederates.

Instruments. In addition to quantitative metrics such as target time compliance, human factor metrics were collected. Workload Assessment Keypads (WAKs) were adapted to chime every five minutes during the HITL runs, at which time Ramp Controllers rated their workload on a 1 to 5 scale, with 1 being "Very low workload," and 5 being "Very high workload." Ramp Controllers also rated their workload on post-run surveys using five of the NASA Task Load Index (TLX) rating items [5]. Situational awareness was derived from an adapted version of the 3D Situational Awareness Rating Technique[1] (SART) [6]. Ramp Controllers were also asked to rate the acceptability of various aspects of ramp operations and to describe how they handled aircraft in the TOBT+TMAT condition. A post-study survey and debrief followed.

Displays. Figure 1 shows the new ATD-2 IADS Ramp Traffic Console (RTC) which was available to Ramp Controllers in all conditions; detailed descriptions can be found elsewhere [7].

Close-ups of the flight strips are shown in Figs. 2 and 3.

[1] The 3D SART consists of a score obtained by adding an item rating "your Understanding of the traffic situation" to an item rating "the Supply of your attentional resources" and subtracting an item rating "Demand on your attention" (i.e., SA = U + (S − D)).

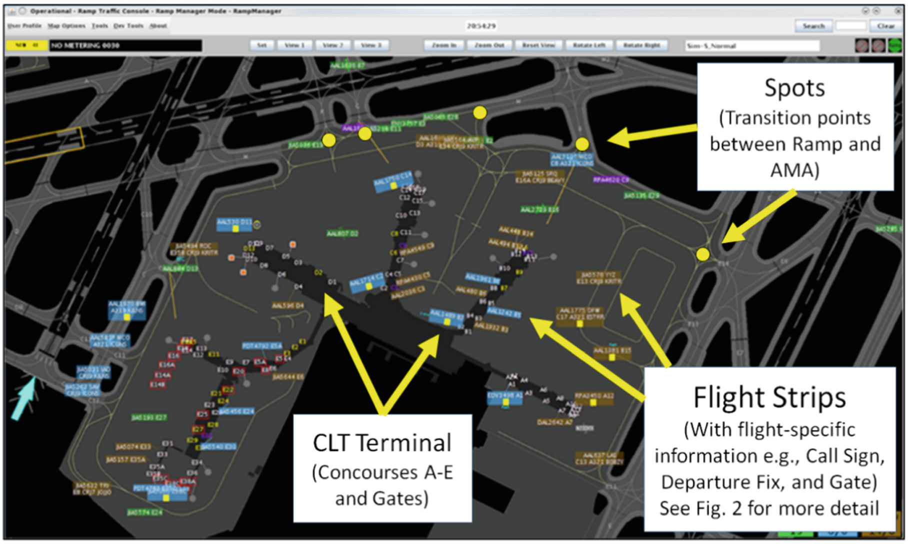


Fig. 1. Display of the Ramp Traffic Console (RTC) tool showing the CLT ramp area with five Concourses (A–E), ramp area (made up of four ramp sectors), and surrounding taxiways and runways. Example "spots," the transition points between the ramp area and the ATC-controlled Airport Movement Area (AMA), are depicted. Each flight is represented by a digital flight strip, which indicates flight-specific information.

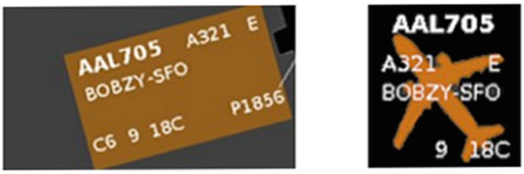


Fig. 2. Digital flight strips on the RTC display in the baseline, or no metering condition, (left) at the gate prior to pushback and (right) following pushback, while taxiing in the ramp area.

Fig. 3. Flight strips for aircraft in both metering conditions: (left) at gate showing gate hold advisory countdown ("4 min") to the TOBT (Target Off-Block Time); (middle) at which time "PUSH" is indicated; (right) after pushback with TMAT (Target Movement Area entry Time) showing "1941" for arrival at the spot.

3 Results

3.1 Quantitative Results

TMAT Compliance Increases for Aircraft Compliant with TOBT. During the runs compliance with the TOBT times at the gate and TMAT times at the spot were measured for all aircraft subject to surface metering. In both the TOBT-Only and TOBT+TMAT condition we observed that compliance with the TOBT advisory increased the chance of complying with the TMAT advisory. Similar findings appear in analysis of operational data collected at CLT during the Phase 1 field evaluation [8]. This indicates that the TOBT advisory is helpful in achieving TMAT compliance. This is an important finding as future systems will be built around controlling the flow of surface traffic via the TMAT at the spot, in contrast to ATD-2 which controls the flow of traffic via the TOBT at the gate (Table 1).

Table 1. TOBT and TMAT compliance metrics for flights subject to metering.

Compliance	TOBT-only condition	TOBT+TMAT condition	Sig. level Chi Square
TOBT (±2 min)	61.7% (21/34)	57.1% (24/42)	p = .68, 0.2 (*df* 1)
TMAT (±5 min)	85.3% (29/34)	69.0% (29/42)	p = .10, 2.7 (*df* 1)
TMAT given TOBT compliance	95.2% (20/21)	75.0% (18/24)	p = .07, 3.3 (*df* 1)

Additionally, we observed that TMAT compliance in the TOBT-Only condition was higher than TMAT compliance in the TOBT+TMAT condition. Given the relatively small sample size of HITL data the significance of this finding is unclear. There exist various factors affecting the traffic situation in each simulation run, but one possible reason for lower TMAT compliance in the TOBT+TMAT condition is the increased workload of having to keep track of aircraft scheduling times at both the gate and the taxiway spot. This increased workload could have lowered the ramp controllers' situation awareness and interfered with meeting the TMAT times. Evidence for this possibility will be explored in the next sections.

3.2 Subjective Results

Perceived Efficiency in Ramp Operations. After each run, Ramp Controllers were asked "During the busiest time in this run, how acceptable were the following in terms of operational efficiency?" As can be seen from Fig. 4, the general trend was that operations were less acceptable in the TOBT+TMAT condition (red bar). Hold times at the hardstands (holding areas for aircraft) were rated as significantly less acceptable in the TOBT+TMAT condition than Baseline.

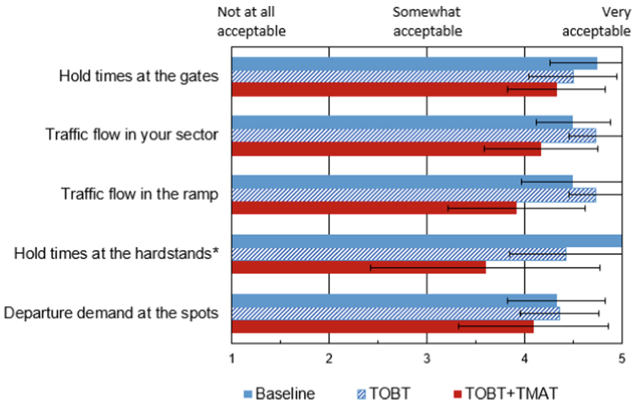


Fig. 4. Ramp controllers' perceptions of the acceptability of ramp conditions in terms of operational efficiency. $N = 36$, (12 ratings in each of 3 conditions, 4 controllers in each of 3 runs). Error bars = 95% confidence intervals. Significant ANOVAs are marked with customary asterisks (e.g. "Hold times at the hardstands" was significant at $p \leq .05$).

Working with TMATs. In the TOBT+TMAT condition, Ramp Controllers were asked, "How frequently in this run did you use TMATs to make decisions about sequencing aircraft?" The average response was "About half the time (Fig. 5)."

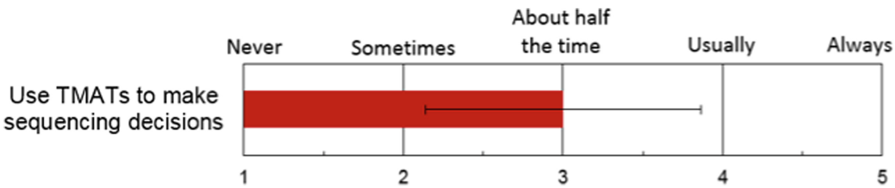


Fig. 5. Average responses of ramp controllers in the TOBT+TMAT condition on frequency of TMAT usage, $N = 12$ ratings (4 controllers in 3 TOBT+TMAT runs). Error bar is 95% confidence interval. Of the 12, 2 responses were "Never."

To explain this finding, a typical comment on the post-run survey was, "Things were flowing a bit fast. I didn't have enough time to really sequence the TMAT times." Hence high workload and time pressure appear to be contributing factors as to why more Ramp Controllers didn't use TMATs to sequence aircraft.

Ramp Controllers were asked to "Please rate how appropriate the times of the TMATs were in this run for aircraft coming from the gates in your sector and from other sectors." For aircraft originating in their own sector, TMAT times were thought to be "About right," as shown in Fig. 6.

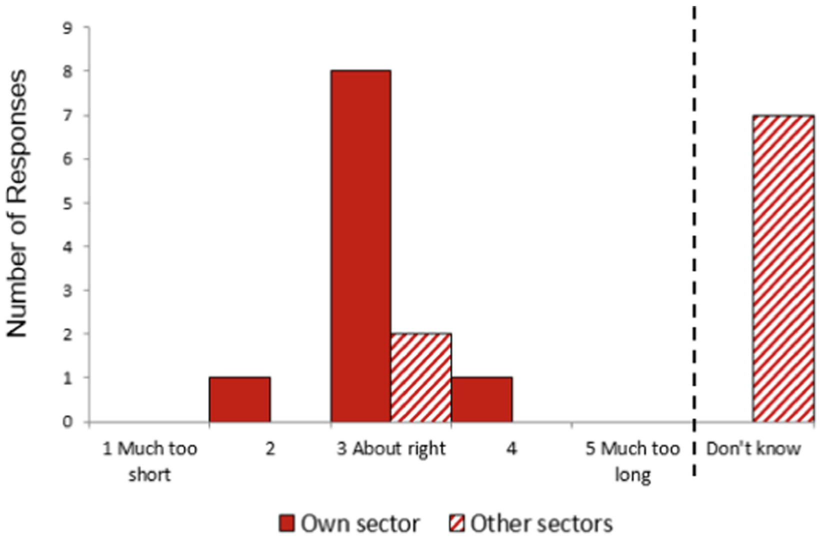


Fig. 6. TMAT times were judged "About right," unless aircraft came from another sector, in which case the ramp controllers selected "Don't know." $N = 10$ ratings for "own sector" and 9 for "other sectors" from ramp controllers who used TMATs to make sequencing decisions.

As can be seen, Ramp Controllers frequently marked "Don't know" on this question for aircraft that had originated in another sector. This shows a lack of situation awareness in this condition and indicates it is likely they were not making sequencing decisions based on TMAT times for those aircraft.

Situation Awareness. The final SART scores in the TOBT+TMAT condition were significantly lower than in the Baseline condition, as shown in Fig. 7. The scores were not significantly lower in the TOBT-Only condition.

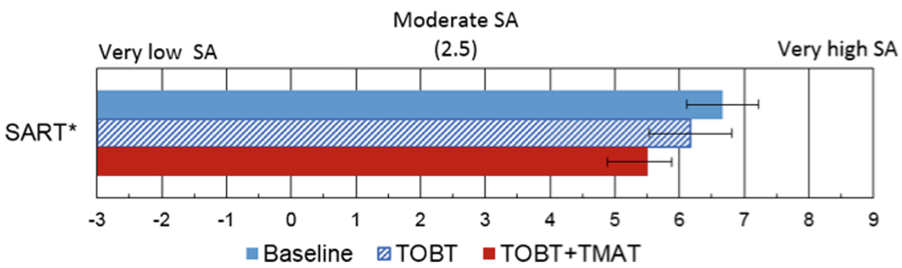


Fig. 7. Final SART scores in the three conditions. $N = 36$ (12 ratings in each condition). ANOVA is significant at $p < .05$; error bars = 95% confidence intervals.

Workload Assessment Keyboard (WAK) Results. Ramp Controller workload was significantly higher in the TOBT+TMAT condition than Baseline as indicated by the WAK (Fig. 8).

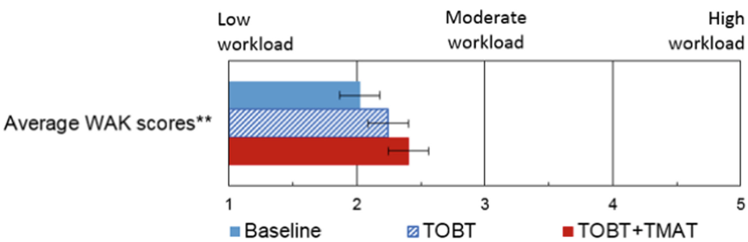


Fig. 8. WAK results in the three conditions. *N*s = 154,155,154 per condition. ANOVA is significant at $p < .01$; error bars = 95% Confidence Intervals.

Results of NASA Task Load Index (TLX) Items. Workload items "Time Pressure" and "Effort" were significantly higher in the TOBT+TMAT condition than in the other two conditions. Other comparisons were not significant (Fig. 9).

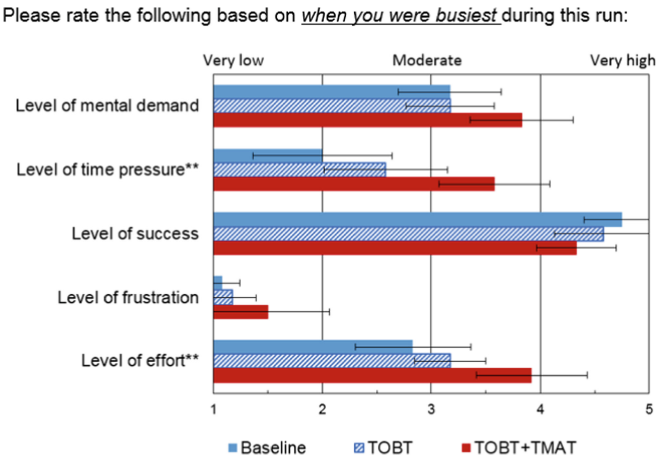


Fig. 9. Results on NASA TLX items in the three conditions. $N = 36$ ratings, (12 in each of 3 conditions). Error bars = 95% Confidence Intervals.

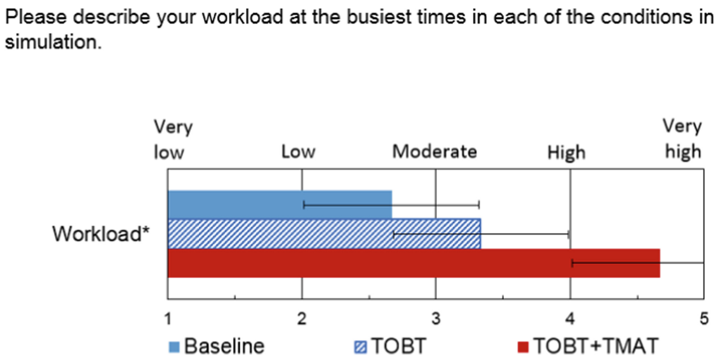


Fig. 10. Self-assessed workload in the three conditions on the post-study survey. $N = 4$. Repeated measures ANOVA, sphericity not assumed, $F(2,2) = 28$, $p = .03$. Error bars = 95% confidence intervals.

Post-Study Survey Results. Self-assessed workload on the post-study survey was significantly higher in the TOBT+TMAT condition than in the other two conditions and near “Very high”, as shown in Fig. 10 below. There was no significant difference between Baseline and the TOBT-Only condition.

4 Summary and Discussion

In anticipation of the implementation of FAA’s Surface CDM concept of surface metering, in this HITL different possibilities were explored for achieving TMAT compliance at the “spot,” the point where ATC takes control, within a ±5 min window. Three conditions were tested to understand the impacts of this goal on operations and on Ramp Controllers at a simulated CLT, where there is limited real estate for managing ramp traffic. Results from the HITL indicated that there was more compliance with the TMATs in the TOBT-Only condition, where Ramp Controllers were only asked to push aircraft back from their gates within a ±2 min window, than in the TOBT+TMAT condition, where Ramp Controllers were given explicit instructions to also try to get aircraft to the spots within a ±5 min window.

Reasons for this finding were explored in this paper. Increased workload and lowered situation awareness may have contributed to reduced TMAT compliance in the TOBT+TMAT condition. First, Ramp Controllers reported using TMATs to make decisions about sequencing only half the time, with comments explaining this in terms of workload and time pressure: “Things were flowing a bit fast. I didn’t have enough time to really sequence the TMAT times.” Second, on specific workload measures, the workload was significantly higher in the TOBT+TMAT condition than the other two conditions. This was the case for two items on the NASA TLX: level of time pressure and level of effort, and for the self-assessed workload on the post-simulation survey. Workload on this survey was rated on a 1 to 5 scale as 4.7 in the TOBT+TMAT condition compared to 3.3 in the TOBT-Only condition, and 2.7 in the Baseline condition.

Although the SART situation awareness scale used in this study indicated that situation awareness was lowest in the TOBT+TMAT condition, this reached significance only in comparison with Baseline. It should be noted, however, that the frequent “Don’t know” responses on the post-run survey also indicated a low situation awareness in the TOBT+TMAT condition. For example, Ramp Controllers frequently did not know if the TMAT times of aircraft originating from a different sector were appropriate. As one Ramp Controller put it,

“Trying to think about the TMAT times and keeping them in order can sometimes be demanding. Trying to keep order and recognize what other team members may have going on is demanding enough. Once I push and send an instruction to taxi, I usually don’t have enough time to go back and see if the TMAT time is within limits. I think the system should monitor and adjust these numbers.”

4.1 Conclusion and Next Steps

Therefore, for airports like CLT with unique limitations in the ramp area, instructing Ramp Controllers to meet scheduling times at the spots may not be feasible. Benefits of TMAT compliance can still be achieved by Ramp Controllers working toward TOBT compliance, as evidenced by improved TMAT compliance and decreased workload in the TOBT-Only condition. Additional anecdotal evidence was captured during a TOBT +TMAT run when one Ramp Controller commented, "I think once the ramp got congested, it basically made the TMAT times insufficient. I would like to have had another way to get a few of them out thru the traffic." Other ramp controllers echoed this in the post-simulation debrief stating that at CLT, once an aircraft is merged into the ramp traffic, there is very little a Ramp Controller can do to meet TMAT times. However, by following the TOBT gate advisories based on an algorithm that makes a best estimate of the time it takes for an aircraft to move from the gate to the spot, Ramp Controllers can, with little increased effort compared to non-metering operations, still deliver aircraft to the spots within the TMAT window.

Next steps are to explore scheduling alternatives in a HITL with more runs, to use an airport with a larger ramp area relative to runway capacity (Dallas-Fort Worth International Airport) and to include a TMAT-Only condition. With a larger ramp area, the Ramp Controllers may be able to meet a TMAT-Only condition.

Acknowledgments. Thanks are due to the many Ramp Controllers, ATC Controllers, and pseudo-pilots who participated in this simulation, along with the ATD-2 Team and the Future Flight Central Simulation Team that made the simulation possible.

References

1. Jung, Y., et al.: Airspace Technology Demonstration 2 (ATD-2) Phase 1 Concept of Use (ConUse), NASA/TM-2018-219770, February 2018
2. FAA Air Traffic Activity System (ATADS): Airport Rankings for 2018 by Operations (2018). https://aspm.faa.gov/opsnet/sys/opsnet-server-x.asp
3. FAA Air Traffic Organization Surface Operations Directorate: U.S. Airport Surface Collaborative Decision Making Concept of Operations (ConOps) in the Near-Term: Application of the Surface Concept at United States Airports, June 2012. https://faaco.faa.gov/index.cfm/attachment/download/33926
4. Aviation Systems Division, NASA Ames Research Center, Airspace Technology Demonstration 2 (ATD-2) Overview. https://www.aviationsystems.arc.nasa.gov/research/atd2index.shtml. Accessed 7 May 2019
5. Hart, S.G., Staveland, L.E.: Development of a NASA-TLX (Task Load Index): results of empirical and theoretical research. In: Hancock, P.S., Meshkati, N., Human Mental Workload, pp. 139–183. Elsevier Science Publishers B.V., Amsterdam (1988)
6. Taylor, R.M.: Situational awareness rating technique (SART): the development of a tool for aircrew systems design. In: Proceedings of the AGARD AMP Symposium on Situational Awareness in Aerospace Operations, CP478. Seuilly-sur Seine: NATO AGARD (1989)

7. Parke, B., Jobe, K., Chevalley, E., Verma, S., Dulchinos, V., Stevens, L.: Impact of data exchange provided by ATD-2 Tools at Charlotte-Douglas international airport. In: 9th International Conference on Applied Human Factors and Ergonomics, Orlando, FL, 21–25 July 2018
8. Coupe, W.J., Lee, H., Jung, Y., Chen, L., Robeson, I.: Scheduling improvements following the Phase 1 field evaluation of the ATD-2 integrated arrival, departure, and surface concept. In: 13th USA/Europe Air Traffic Management Research and Development Seminar, Manuscript in progress (2019)

Safety Risk Assessment of Space Launch Site Considering Human Reliability

Tun Liu, Wenjin Zhang, Xing Pan(✉), and Xiaoxia Wang

School of Reliability and Systems Engineering, Beihang University,
Beijing 100191, China
liu_tun@163.com, {zwjok,panxing}@buaa.edu.cn,
langshixuan@foxmail.com

Abstract. This paper evaluates the safety risk of space launch site from the perspective of human reliability. Firstly, the risk causes are analyzed, and it is found that the main reason affecting the safety risk of space launch site is human error. Then, a modified CREAM is proposed to analyze the human reliability of risk events in space launch site. The method uses the fuzzy function to convert discrete CPCs and control modes into continuous, and apply Bayesian reasoning to obtain more accurate HEP. Finally, on the basis of HEP data, the Bayesian network of space launch site risk is constructed, and the risk probability of the overall system is evaluated and sensitivity analysis is carried out.

Keywords: Space launch site · Human reliability · CREAM · Bayesian network · Safety risk

1 Introduction

Space launch site is a system with high reliability. Any component failure or human error may cause serious accidents [1]. Therefore, it is important to evaluate the safety risk of space launch site. The early accidents of space launch site are mostly caused by equipment failures [2]. However, with the development of technology and the emergence of various hazard analysis techniques, the safety problems caused by equipment failures have been gradually solved [3]. At present, human factors have more and more important influence on the safety risk of space launch site [4]. In order to improve the safety and reliability of space launch missions, it is urgent for space launch site to evaluate its safety risk from the perspective of human reliability, and explore the mechanism of human errors in the process of space launch, so as to reduce the safety risks of space launching site effectively.

Despite the maturity of space launch technology in various countries, and the safety risks of space launch are strictly controlled [5]. However, in recent years, there are still some major accidents in space launch missions around the world. Sala-Diakanda, S.N. et al. evaluated the average collective risk caused by rocket launch based on information fusion method, which provides a reference for risk aversion after rocket launch [6]. Veniaminov, S. et al. studied the impact of environmental factors on the safety risk of space launch through time series analysis of a large number of space launch data [7]. In a word, many researchers have assessed the safety risk of space launch from different

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 321–331, 2020.
https://doi.org/10.1007/978-3-030-20037-4_29

perspectives, but the safety risk assessment of space launch site from the perspective of human reliability is still rare.

2 Risk Causes Analysis

Space launch site is usually composed of technical zone, launch zone, and ground measure-control system. The technical zone is a special area for technical preparation, its main task is to assemble and test launch vehicles and spacecraft. The launch zone is a dedicated area for preparation before launch. The ground measure-control system is a set of facilities for tracking and measuring launch vehicles and spacecraft, and transmitting communication information [8]. Therefore, the risk of space launch site mainly consists of technology zone risk, launch zone risk and ground measure-control system risk. The risk types of launch zone include process control risk, rocket stability risk, launch delay risk and casualty risk. Fault Tree Analysis (FTA) of the space launch site is carried out as shown in Fig. 1, in which bottom events 1 to 15 represent key risk events that occur frequently in space launch site or have a large impact on the overall system. These key risk events are described in Table 1.

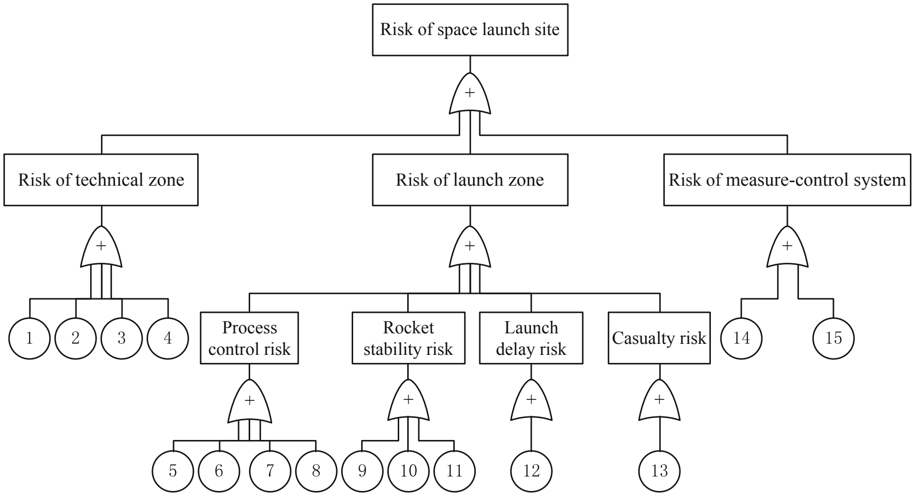


Fig. 1. FTA of space launch site risk

Statistical analysis of the failures of the space launch site in the past five years shows that the root causes of the key risk events consist of misoperation, environmental interference, management issues, operability of equipment, design defects of equipment and quality problems of equipment. The ratio of the root causes are shown in Fig. 2. Among them, environmental interference, management issues and operability of equipment mainly affect the operation of human. Design defects of equipment and quality problems of equipment can be collectively referred to as failure problems of equipment. Therefore, the root causes of key risk events can be macroscopically divided into human error risk and equipment failure risk. Through Fig. 2, the total risk of human error is 71.43%, and the total risk of equipment failure is 28.57%. It can be concluded that the safety risk of space launch site is mainly caused by human error. The human reliability analysis should be emphasized in the safety risk assessment of space launch site.

Table 1. Key risk events and description

No.	Key risk events
1	The spectrum clutter output of phaselocking frequency multiplier increased
2	Cable plug of front connecting rod is damaged by screw
3	Accelerometer +Z-channel pulse anomaly
4	The control cable socket of star-arrow separation cutter is loosen
5	The pre-computer can't start normally
6	Abnormal exit of front-end strapdown geodetic software
7	Commander's command is wrong, delayed or advanced
8	The fire-fighting truntable ladder can't be geared
9	The plug of control system is indented
10	The outlet pressure of JQ2 vacuum reducer in cryogenic power system rises sharply
11	Extension damage of servo mechanism
12	Booster compartment door unsealed
13	Plug of hard hose for filling is ejected
14	Fire in automobile power station
15	Uplink remote control command of malindy station was not issued

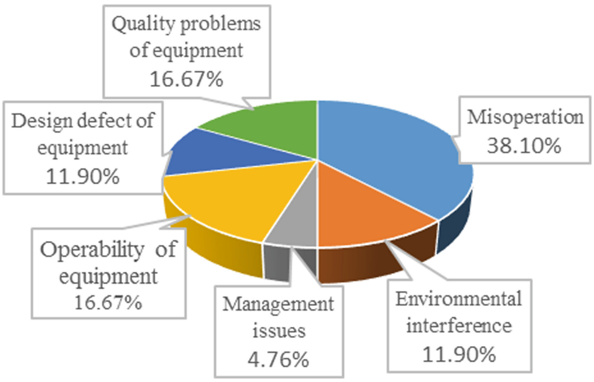


Fig. 2. Ratio of root causes

After the risk cause analysis of space launch site above, it is found that human error has the greatest impact on the safety risk of space launch site. Therefore, firstly this paper will analyze the human reliability of the key risk events of space launch site, and then construct Bayesian network to evaluate the safety risk of space launch site. The research framework of this paper is Fig. 3.

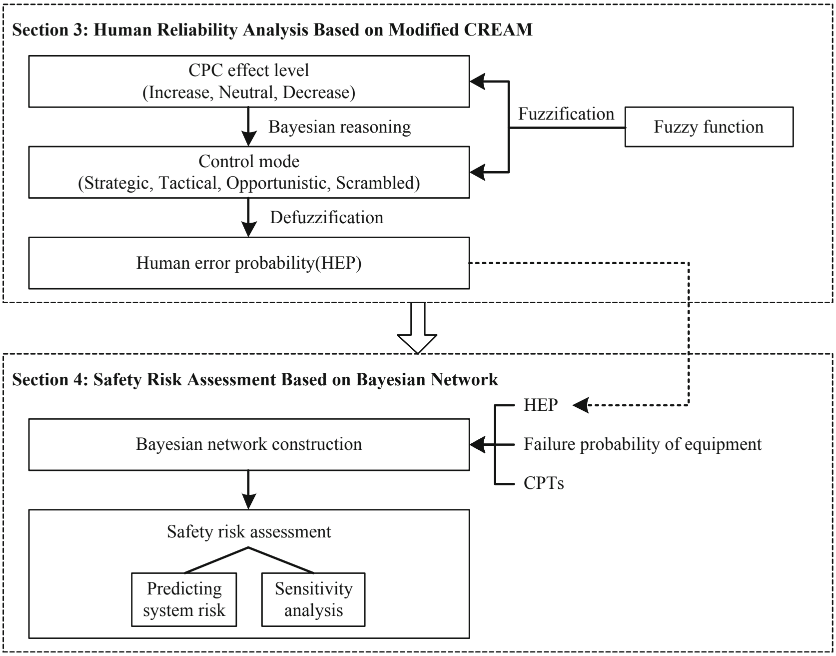


Fig. 3. Research framework of the paper

3 Human Reliability Analysis Based on Modified CREAM

3.1 Modified CREAM

CREAM (Cognitive Reliability and Error Analysis Method), as a HRA (Human Reliability Analysis) method based on Context Control Model (COCOM), its core idea is to emphasize that human error is not an isolated random behavior, but depends on the situational environment in which human complete tasks [9]. CREAM classifies human factors into nine categories, called Common Performance Conditions (CPC). Each CPC factor has three levels of impact on human reliability, which are increase (improved), neutral and decrease (reduced) [10]. According to Fig. 4, the corresponding control modes are obtained by calculating the number of CPC factors at the improved and reduced levels [11]. These control modes have their own intervals to estimate Human Error Probability (HEP).

The traditional CREAM has the following effects. Firstly, the method does not consider the influence weight of different CPC factors on human error. Secondly, the discrete control mode is used to estimate the continuous human error probability. Thirdly, the corresponding human error probability range of each control mode is wide, appropriate methods are needed to estimate the human error probability in order to ensure the accuracy of HEP. Aiming at the effects of traditional CREAM, this paper proposes a modified CREAM, which combines fuzzy function, Bayesian reasoning theory and traditional CREAM to evaluate HEP more accurately.

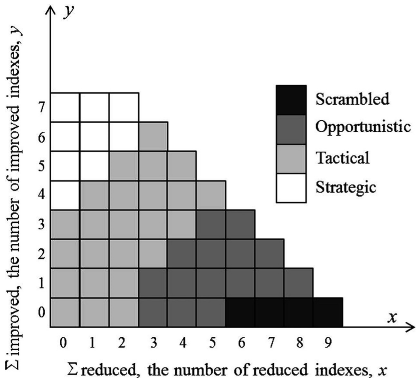


Fig. 4. Relations between CPC score and control mode [11]

3.2 Fuzzification of CPCs and Control Mode

The CPCs are determined based on the specific working environment on which operator depend to complete his tasks, so as to assess the expected impact of CPC on human reliability [12]. Konstandinidou et al. suggest that CPC can be modified according to specific cases, or new CPC can be added to expand the input parameters, so as to meet the evaluation of human reliability in different fields [13]. Based on the nine types of CPCs proposed by the traditional CREAM, this paper should determine the CPCs for human reliability analysis in the space launch site, as shown in Table 2. The effect level of each CPC on human reliability is divided into increase, neutral, and decrease. In order to improve the accuracy of the effect level assessment, the scoring domains of the three effect levels are designated as [50, 100], [10, 90], [0, 50]. The relationship between the probability of CPC's effect level on human reliability (increase, neutral, decrease) and expert scores can be mapped by a fuzzy trigonometric function, as shown in Fig. 5 [14].

For example, in the process of rocket launching, for the risk event of "Abnormal exit of front-end strapdown geodetic software", the CPC of the risk event is scored by experts. According to the mapping relationship between the expert score and the probability of CPC effect level, as shown in Fig. 5, the probability of three effect levels corresponding to each score can be obtained, as shown in Table 3.

Table 2. CPCs of space launch site

No.	CPCs
C1	Organisational management
C2	Working conditions
C3	Available time
C4	Training and experience
C5	Team collaboration quality
C6	Operator's physiological conditions
C7	Operator's psychological situation
C8	Equipment availability
C9	Information exchange

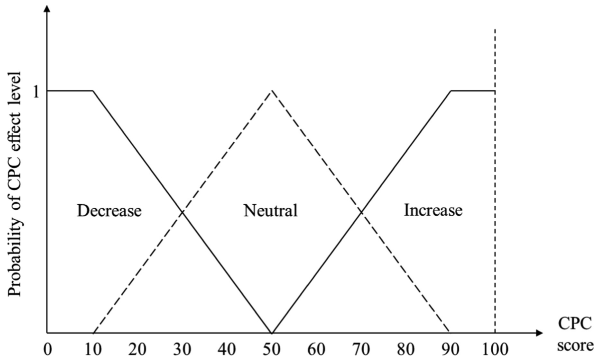


Fig. 5. Fuzzy probability function of CPCs effect levels

Table 3. CPC scores and CPC effect level probability of risk event

Risk event	Abnormal exit of front-end strapdown geodetic software								
CPCs	C1	C2	C3	C4	C5	C6	C7	C8	C9
Score	88	30	50	76	62	75	37	24	80
Decrease	0	0.5	0	0	0	0	0.325	0.65	0
Neutral	0.05	0.5	1	0.35	0.7	0.375	0.675	0.35	0.25
Increase	0.95	0	0	0.65	0.3	0.625	0	0	0.75

The traditional CREAM uses control mode to evaluate human error probability. There are four control modes, which include Strategic, Tactical, Opportunistic and Scrambled, and they correspond to different intervals of HEP. For the convenience of calculation, the interval of human error probability is converted into log10 interval, as shown in Table 4 [14]. Fuzzify the interval of human error probability corresponding to control mode, as shown in Fig. 6.

Table 4. Control modes and their probability intervals

Control mode	Strategic	Tactical	Opportunistic	Scrambled
HEP intervals	(0.000005, 0.01)	(0.001, 0.1)	(0.01, 0.5)	(0.1, 1.0)
Log10 (HEP)	(−5.3, −2)	(−3, −1)	(−2, −0.3)	(−1, 0)

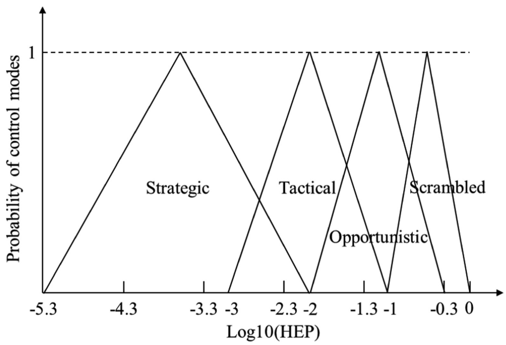


Fig. 6. Fuzzy probability function of control modes

3.3 Bayesian Reasoning

After the discrete CPCs effect levels and control modes are fuzzified, the control mode probabilities corresponding to different CPCs scores can be obtained by Bayesian probabilistic reasoning, the Bayesian reasoning model is shown in Fig. 7. The core idea of the Bayesian reasoning model is to transform the discrete correspondence between CPCs and control mode into more accurate continuous reasoning relationship.

In order to simplify the input of conditional probability between CPCs and control modes, firstly, this paper introduces node "M_i" (i = 1, 2, 3, 4) to represent the comprehensive impact of CPC_{2i-1} and CPC_{2i} on improving human reliability, and introduces node "R_i" (i = 1, 2, 3, 4) to represent the comprehensive impact of CPC_{2i-1} and CPC_{2i} on reducing human reliability. Furthermore, considers the weight of different CPC factors on the conditional probability between CPCs and "M_i" and "R_i". And then node "improved" represents the impact of all CPCs on improving human reliability, and node "reduced" represents the impact of all CPCs on reducing human reliability.

Since the correspondence between CPCs and control modes in the traditional CREAM is too discrete, this paper constructs the conditional probability of node "improved" and "reduced" to node "control mode" according to the 80–20 principle. the boundaries of the four control modes are divided by the probability of 80% and 20%. For example, when (Σimproved, Σreduced) = (5, 1), the probability of human reliability is 80% in control mode and 20% in Strategic control mode. According to this principle, the final results are as follows: the probability of Strategic is 11%, the probability of Tactical is 88%, the probability of Opportunistic is 1%, as shown in Fig. 7.

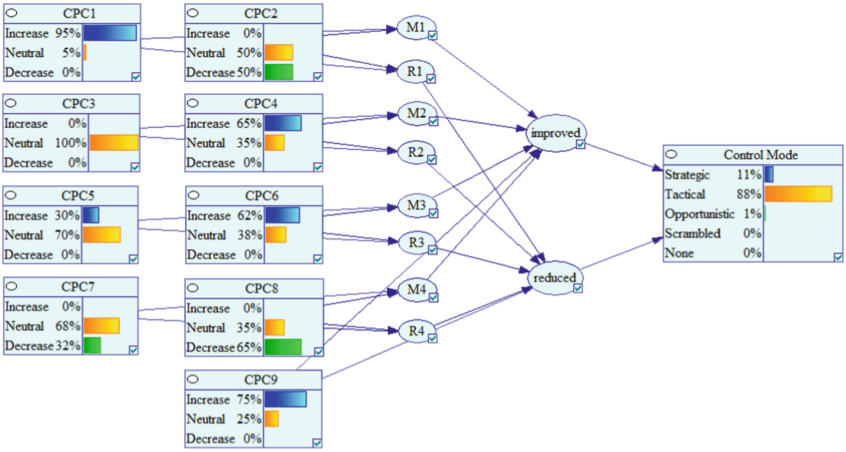


Fig. 7. Bayesian reasoning result for CPCs to control mode

After the fuzzy probability of the control mode is defuzzified [15], the HEP of the risk event "Abnormal exit of front-end strapdown geodetic software" is solved as 0.0039. According to the same method, the HEP of any risk event in the space launch site can be solved. These HEPs will be applied to the overall risk assessment of the space launch site.

4 Safety Risk Assessment Based on Bayesian Network

4.1 Bayesian Network Construction

In order to construct a Bayesian network model of space launch site risk, this paper transformed the fault tree in Fig. 1 into a Bayesian network [16]. For each bottom event in the fault tree, two parent nodes can be added to the Bayesian network, which are node "human" and "machine". Node "human" represents human error factors, node "machine" represents equipment failure factors, the Bayesian networks constructed such as Fig. 8. After constructing Bayesian network, it is necessary to input probability data to Bayesian network. The probability of node "human" is HEP, which is obtained by the modified CREAM proposed in Sect. 3. The probability of node "machine" is provided by the reliability standard of equipment or manufacturer. CPTs are obtained by expert evaluation method.

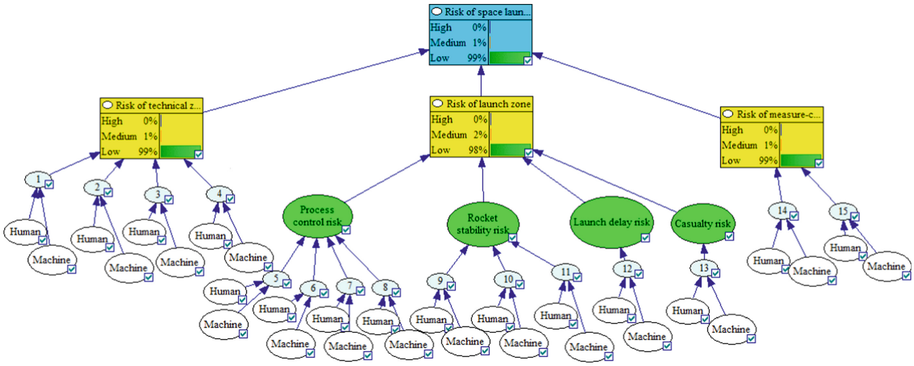


Fig. 8. Bayesian network of space launch site risk

4.2 Safety Risk Assessment

Through the probability reasoning of Bayesian network, the probability of space launch site in high risk is 5.8e-5, that of medium risk is 0.0135 and that of low risk is 0.9864. In the space launch site, the probability intervals corresponding to the three risk levels are evaluated by experts, such as Table 5. Using fuzzy function to map the relationship between risk level and risk probability, finally the system risk probability of the space launch site is solved as 0.00329%.

Table 5. Probability intervals of risk levels

Risk levels	High risk	Medium risk	Low risk
Probability intervals	(0.01, 1)	(0.001, 0.01)	(0.000001, 0.001)

Because the probability of space launch site in Low risk is the largest proportion of the other two risk levels, the sensitivity analysis of space launch site in Low risk is mainly carried out. The analysis results are shown in Fig. 9, which shows that the change of human error probability and equipment failure probability in risk events 14 and 15 has the greatest impact on total system risk, followed by the change of human error probability in risk events 12, 13, 2, 4, 3 and 1. Therefore, in order to reduce the safety risk of space launch site, the risk probability of risk events 14 and 15 should be controlled strictly, and the human error risk of risk events 1, 2, 3, 4, 12 and 13 should be controlled secondly.

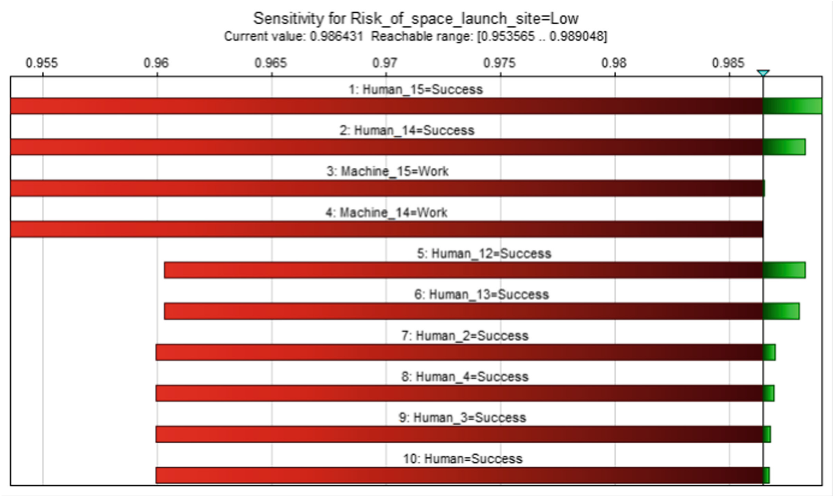


Fig. 9. Sensitivity analysis for space launch site in low risk

5 Conclusions

Most of the existing literature only considers the failure risk of equipment, but pays litter attention to the impact of human errors on safety risk. In recent years, the data of space launch show that human errors have an increasing impact on the safety risk of space launch site. Therefore, this paper evaluates the safety risk of space launch site from the perspective of human reliability.

In this paper, an modified CREAM for space launch site is proposed to solve the HEP of space launch site risk events more accurately. By introducing a fuzzy function, the discrete CPCs and control modes are continuous. Bayesian reasoning not only considers the weight of different CPCs on human reliability, but also improves the estimation accuracy of HEP.

After obtaining the HEP of each risk event in the space launch site, the Bayesian network of the space launch site can be constructed. The basic nodes in Bayesian network mainly include human error and equipment failure. The probability of human error is obtained by modified CREAM, and the probability of equipment failure is obtained by finding the reliability data of equipment. Finally, the safety risk probability of space launch site is predicted and sensitivity analysis is carried out.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grants No. 71171008/No. 71571004, the National Key Research and Development Program of China, the Special Scientific Research Project for Civil Aircraft of the Ministry of Industry and Information Technology.

References

1. Rising, J.M., Leveson, N.G.: Systems-theoretic process analysis of space launch vehicles. J. Space Saf. Eng. **5**, 153–183 (2018)
2. Marozau, I., Auchlin, M., Pejchal, V., Souchon, F., Vogel, D., Lahti, M., Saillen, N., Sereda, O.: Reliability assessment and failure mode analysis of MEMS accelerometers for space applications. Microelectron. Reliab. **88–90**, 846–854 (2018)
3. Keller, S., Collopy, P.: Value modeling for a space launch system. Proc. Comput. Sci. **16**, 1152–1160 (2013)
4. Kishi, N.: Management analysis for the space industry. Space Policy **39–40**, 1–6 (2017)
5. Haber, J.M.: Launch and reentry safety objectives. J. Space Saf. Eng. **4**, 22–28 (2017)
6. Sala-Diakanda, S.N., Sepulveda, J.A., Rabelo, L.C.: A methodology for realistic space launch risk estimation using information-fusion-based metric. Inf. Fusion **11**, 365–373 (2010)
7. Veniaminov, S., Klyushnikov, V., Kozlov, S., Nagorskiy, P.: Launch malfunctions and space weather factors. Acta Astronaut. **150**, 33–38 (2018)
8. Kiureghian, A.D.: Risk assessment of satellite launch with reusable launch vehicle. Reliab. Eng. Syst. Saf. **74**, 353–360 (2001)
9. Zhou, Q., Wong, Y.D., Xu, H., Thai, V.V., Loh, H.S., Yuen, K.F.: An enhanced CREAM with stakeholder-graded protocols for tanker shipping safety application. Saf. Sci. **95**, 140–147 (2017)
10. Hollnagel, E.: Chapter 6 - CREAM – a second generation HRA method. Cognitive Reliability and Error Analysis Method (CREAM), 151–190 (1998)
11. Ung, S.-T.: A weighted CREAM model for maritime human reliability analysis. Saf. Sci. **72**, 144–152 (2015)
12. Swain, A.D.: Human reliability analysis: need, status, trends and limitations. Reliab. Eng. Syst. Saf. **29**, 301–313 (1990)
13. Konstandinidou, M., Nivolianitou, Z., Kiranoudis, C., Markatos, N.: A fuzzy modeling application of CREAM methodology for human reliability analysis. Reliab. Eng. Syst. Saf. **91**, 706–716 (2006)
14. Zhou, Q., Wong, Y.D., Loh, H.S., Yuen, K.F.: A fuzzy and Bayesian network CREAM model for human reliability analysis – the case of tanker shipping. Saf. Sci. **105**, 149–157 (2018)
15. Sun, Z., Li, Z., Gong, E., Xie, H.: Estimating human error probability using a modified CREAM. Reliab. Eng. Syst. Saf. **100**, 28–32 (2012)
16. Ning, X., Qi, J., Wu, C.: A quantitative safety risk assessment model for construction site layout planning. Saf. Sci. **104**, 246–259 (2018)

Author Index

R. L. Boring (Ed.): AHFE 2019, AISC 956, pp. 333–334, 2020.
https://doi.org/10.1007/978-3-030-20037-4