Advances in Human Factors of Transportation, Vol. 148, 2024, 121–127
https://doi.org/10.54941/ahfe1005202

Crew Resource Management Evaluation for Pilots: A Modern Perspective

Kiranraj Pushparaj, Hui Wang, David Esser, and Barbara Holder

School of Graduate Studies, College of Aviation, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA

ABSTRACT

Crew Resource Management (CRM) has been an integral part of aviation for decades. However, evaluation of CRM for pilots has always been a challenge. As such, an exploration of the literature was commenced to approach CRM evaluation from a modern perspective. Some teams promoted the integration of technology into the assessment process to better understand the behavioral markers currently being utilized. Others encouraged a larger revamp, by shifting towards a more holistic assessment based on advances in cognitive science. Both these approaches were discussed, and it is likely that a blend of both will be required to address this research problem. However, to ensure robustness and long-term sustainability of CRM evaluation, more research must be conducted from a cognitive science perspective, to inform which assessment frameworks are more effective and reliable.

Keywords: Crew resource management, Nontechnical skills, Aviation, Cognition

INTRODUCTION

Since the 1970s, nontechnical skill factors of pilots were found to be the leading cause of aircraft accidents; prompting the aviation industry to mandate Crew Resource Management (CRM), in the flight deck, as a pivotal segment of pilot training and continual assessment throughout the career of a pilot. (Gontar, Fischer & Bengler, 2017a). According to the Federal Aviation Administration's CRM Advisory Circular 120-51B, CRM refers to the effective use of all available resources: human resources, hardware, and information to attain safe and efficient flight operations. CRM training and assessment is based on the understanding that high technical proficiency of the pilots is the foundation for safe and efficient operations. CRM alone cannot compensate for poor technical mastery or a lack of proficiency. Further, high technical proficiency does not ensure safe operations without effective crew coordination.

CRM incorporates critical nontechnical skills, such as leadership, communication, judgement, and decision making. (Mearns, Flin & O'Connor, 2001). However, the assessment of CRM knowledge and skill elements are often subjective, and are assessed in realistic scenarios of flight events. To date, assessment metrics and methods have not been empirically validated, which highlights a potential for research to modernize and enhance these

methods (Helmreich, Foushee, 2019). Consistent and objective CRM assessment is needed to ensure the flight crews are equipped with the required skills to meet operational demands. Any change however, will require the buy-in and engagement of aviation industry stakeholders, airlines, pilots, and regulatory authorities. This approach was reiterated during the development of the NOTECHS framework to introduce CRM for pilots in Europe, where the ease of use by evaluators, who are not psychologists, was prioritized heavily (O'Connor et al., 2002).

Most CRM and other nontechnical skills assessment tools use behavioral marker frameworks that are utilized to guide assessment, whereas technical skills assessment tools use rating scales to describe performance criteria associated with dimensions of technical performance or steps of a procedure. Various behavioral marker systems have been deployed in aviation, such as Line/LOS Checklist (LLC), Line Operations Safety Audit (LOSA), and Targeted Acceptable Responses to Generated Events or Tasks (TARGETs), to evaluate pilots' CRM performance. However, even experienced evaluators often yield inconsistent assessments when using these evaluation methods (Gontar, Patrick, Hoermann, 2015).

Competency-based frameworks, which involve the demonstration of proficiency in the underlying requisite knowledge, skills, and abilities have also been utilized (MacLeod, 2021). The International Air Transport Association (IATA) published a framework for identifying varying levels of pilot proficiency through behavioral markers for nine distinct competencies (Sun et al., 2023). Despite this, there are several limitations, such as a lack of a globally accepted competency-based framework, as well as the time consuming and costly processes involved in outlining the necessary competencies, and their associated knowledge, skills, and abilities (Lin, Shahhosseini & Janke, 2018).

Some have also proposed the use of technologies such as eye-tracking, physiological indicators, and recorded simulator flight data outputs, as an approach to provide greater objectivity to behavioral markers (Knabl-Schmitz et al., 2023). However, the practicality of using and implementing these techniques in practice is uncertain due to labor restrictions and operational constraints. While these techniques may be appropriate for CRM evaluation in other industries, the very nature of flying is vastly different from fields like healthcare. As such, an exploration of the literature was conducted, to identify ways to supplement insights gained by behavioral markers, with cognitive indicators for which pilots are evaluated.

STRATEGIES TO MODERNIZE CRM ASSESSMENT

Some mechanisms, such as peer evaluations and self-evaluations are in place in existing assessment frameworks to reduce subjectivity by improving transparency (Bates et al., 1997). However, studies have shown that these procedures can be counter-productive, with peer evaluation scores significantly surpassing that of self-evaluations due to personality traits and lack of formal training to assess pilots (Gontar et al., 2014). There seems to be two different approaches to address the subjectivity in CRM assessment for pilots.

The use of supplementary techniques has been touted as a viable option to provide more data to the evaluators (Uenking, 2000). This data-driven approach to CRM assessment aims to enable evaluators to understand the intent behind the actions of pilots during scenario-based evaluations, and result in a more objective evaluation of pilots' performances. However, this approach requires additional data management tasks that could add extensive labor hours in the evaluation process. The possible benefits of this improved objectivity in CRM assessment need to be weighed against the impact on time and cost.

Another proposed strategy is to shift the focus of assessments away from purely behavioral markers, and towards a more holistic performance evaluation framework. This kind of approach could also help to identify social and cognitive aspects of CRM (Gontar, Hoermann, 2014). Furthermore, behavior is but one representation of cognition. A deeper, more comprehensive approach, backed by research and experimental validation, could yield more objective assessments that are also operationally acceptable.

It should be noted that both these strategies are not mutually exclusive. There can be ways to blend to a data driven approach with a more holistic cognition-based approach. Perhaps it should even be encouraged. However, for the purpose of this review, they will be considered separately because they are philosophically different. The former focuses on breadth, by incorporating additional data sources, while the latter focuses on depth, by applying recent developments in cognitive science.

DATA DRIVEN APPROACH

Advocates of data driven approaches to assessment assert there is a need for empirical substantiation, especially when recommending changes to current practices (Seah et al., 2021). There have been several technological solutions posited to make CRM assessment more objective for pilots. For example, eyetracking can be used to examine pilots' gaze patterns and head movements during CRM evaluation (Knabl-Schmitz et al., 2023). Eyetracking was demonstrated to be a noninvasive method for gleaning valuable insight into underlying causes of suboptimal pilot performance by furnishing a higher resolution of data (Lounis, Peysakhovich & Causse, 2021). The gaze and monitoring patterns of experimental subjects enhanced post assessment debriefs, where instructors were able to provide personalized feedback, even to pilots who passed, on areas they could improve upon.

Despite this potential, there are several concerns with introducing eyetracking into the assessment toolkit. Compatibility with simulators continues to be an issue, such as the need for additional equipment and software to integrate eyetracking into simulators. Furthermore, additional training may be required for instructors, to code and use the equipment reliably for assessment. The cost of training evaluators and additional technology required to assimilate eyetracking into simulators may constrain its implementation.

Proponents of technology that leverage physiological input from pilots to augment assessments make a similar argument that constructive information can be gained with the use of equipment such as electroencephalography

(EEG), electrocardiogram (ECG), and electrooculogram (EOG) (Seah et al., 2021, Uenking, 2000). While the potential of these technologies in various applications is undeniable, even an apparently minimally invasive device, such as an EEG skullcap, could be problematic in the cramped Flight Deck, which is typically a constricted environment. Even in a simulator environment, the use of such equipment may not be practical.

A potential alternative is the use of electrodermal activity (EDA), which can be measured using equipment that are no bigger than watches (Motogna, Lupu-Florian & Lupu, 2021). These devices have similar functionality to the ECG, in addition to measuring physiological arousal. While they may not have the ability to track monitoring patterns or neural activity, they do provide an additional dimension of data that can provide greater context to behavioral markers.

However, the same cost and training considerations that apply to eyetracking equipment also apply to the use of physiological equipment. In addition, the collection of physiological and neural data from pilots is against current labor laws in the United States. While these techniques may be applicable in a laboratory or experimental environment, they will not be feasible in a training environment.

One technique that may not require additional equipment is speech analysis. Verbal callouts are a prominent component of standard operating procedures for pilots (Sassen, 2005). Past studies have even shown that crews evaluated to have better performance have an increased frequency and willingness to communication with their crew (Mosier, Fischer, 2017). However, nuance is required in speech analysis, as studies also showed conflicting results, where increased frequency of communication led to a degradation in performance (Gontar, Fischer & Bengler, 2017b). A deeper analysis revealed that subpar performances were a result of crew members talking simultaneously and interrupting each other. Sequential speech patterns were then found to be a more effective way to conduct speech analysis. Nonetheless, no apparent consensus has been reached in this area of research, and may not yet be suitable for use.

HOLISTIC COGNITION BASED APPROACH

While technological advances may have yielded new technologies that can be integrated into CRM measurement, cognitive science is also a field that has certainly evolved greatly since CRM was first introduced. The fundamental understanding of cognition has progressed from passive, static processes that only involves the brain processing stimuli centrally, to one that entails "dynamical and reciprocal real-time interaction with the environment" (Newen, De Bruin & Gallagher, 2018). Accordingly, flying, which is inherently a cognitive activity, involves all aspects of 4E cognition. Flying is enacted and embodied, and embedded in an environment for action that is extended across the pilots, their interactions with the flight deck, and other actors in the system, such as air traffic controllers and ground crews.

As such, CRM assessment may require a more fundamental shift, away from evaluating specific behavioral markers, and towards a more dynamic

interaction-based evaluation. New frameworks on how pilots may predict and recognize situations and resources, and how they leverage those resources to organize their activity, exist in the literature (Wilson, Golonka, 2013). However, these recent developments in the understanding of cognition have not made their way into the field of aviation, or for advancing pilot training. Given the cognitive nature of flying, as well as the structured environment that makes up flight decks, this approach has the potential to make CRM assessment more robust.

It will take concerted research efforts to navigate this path, as many research questions are yet to be answered. The cognitive skills and knowledge required for effective performance on CRM tasks need to be established. Reliable training and evaluation of these skills and knowledge can only be conducted after this important first step. By leveraging modern cognitive science, it is possible to formulate new systems of training and assessment that are more accurate, effective, and capable of obtaining acceptance among aviation stakeholders.

CONCLUSION

Technological advances have highlighted various equipment and fields of research, as potential solutions to reducing the subjectivity in CRM assessment. One, or a combination of these devices may emerge as a viable addition to current practices. A multimodal evaluation with questionnaires, physiological output, and behavioral markers have been proposed in the past (Uenking, 2000). However, with the current state of the art, as well as regulatory restrictions, some of these methods remain an impractical solution.

Flying is by nature a cognitive task, so it makes sense that research should be guided by modern cognitive science. While research is certainly becoming increasingly interdisciplinary, the fundamental guiding principles of a task like flying may be better guided centrally by cognitive science, with secondary inputs from other fields to augment as needed. The distinctive constraints that exist in aviation should not be ignored in the process of finding similarities and applicable research in other fields.

ACKNOWLEDGMENT

This research is funded by the Federal Aviation Administration (FAA) NextGen Human Factors Division (ANG-C1) in support of the FAA Aviation Safety (AVS) organization. Embry-Riddle Aeronautical University Principal Investigator, Dr. Barbara Holder, would like to thank the FAA program manager, Dr. Chuck Perala, and AVS technical sponsor Dr. Kathy Abbott for their support of this work. The views expressed in this paper are those of the authors and do not reflect views of Embry-Riddle Aeronautical University (ERAU) or the Federal Aviation Administration (FAA).

REFERENCES

Bates, M. J., Colwell, C. D., King, R. E., Siem, F. M. & Zelenski, W. E. 1997, *Pilot Performance Variables*, Citeseer, Technical Report AL/CF-TR-1997-0059, US Air Force Armstrong Laboratory, Brooks AFB, TX.

Gontar, P. & Hoermann, H. 2014, "Flight Crew Performance and CRM Ratings Based on Three Different Perceptions", *Aviation Psychology: Facilitating Change (s)*.

Gontar, P., Fischer, U. & Bengler, K. 2017a, "Methods to Evaluate Pilots' Cockpit Communication: Cross-Recurrence Analyses vs. Speech Act–Based Analyses", *Journal of Cognitive Engineering and Decision Making,* vol. 11, no. 4, pp. 337–352.

Gontar, P., Fischer, U. & Bengler, K. 2017b, "Methods to Evaluate Pilots' Cockpit Communication: Cross-Recurrence Analyses vs. Speech Act–Based Analyses", *Journal of Cognitive Engineering and Decision Making,* vol. 11, no. 4, pp. 337–352.

Gontar, P. & Hoermann, H. 2015, "Interrater Reliability at the Top End: Measures of Pilots' Nontechnical Performance", *The International Journal of Aviation Psychology,* vol. 25, no. 3–4, pp. 171–190.

Gontar, P., Hoermann, H., Deischl, J. & Haslbeck, A. 2014, "How Pilots Assess Their Non-Technical Performance–A Flight Simulator Study", *Advances in human aspects of transportation.part i,* pp. 119–128.

Helmreich, R. L. & Foushee, H. C. 2019, "Chapter 1 - Why CRM? Empirical and Theoretical Bases of Human Factors Training" in *Crew Resource Management (Third Edition)*, eds. B. G. Kanki, J. Anca & T. R. Chidester, Academic Press, pp. 3–52.

Knabl-Schmitz, P., Cameron, M., Wilson, K., Mulhall, M., Da Cruz, J., Robinson, A. & Dahlstrom, N. 2023, "Eye-tracking: From concept to operational training tool", *Aviation Psychology and Applied Human Factors,* vol. 13, no. 1, pp. 47–57.

Lin, Y., Shahhosseini, A. M. & Janke, C. 2018, *Board 58: Concept Map-based Aviation Competency Mapping and Training*, ASEE Conferences, Salt Lake City, Utah.

Lounis, C., Peysakhovich, V. & Causse, M. 2021, "Visual scanning strategies in the cockpit are modulated by pilots' expertise: A flight simulator study", *PLOS ONE,* vol. 16, no. 2, pp. e0247061.

MacLeod, N. 2021, *Crew resource management training: A competence-based approach for airline pilots,* CRC Press/Routledge/Taylor & Francis Group, Boca Raton, FL, US.

Mearns, K., Flin, R. & O'Connor, P. 2001, "Sharing 'worlds of risk'; improving communication with crew resource management", *Journal of Risk Research,* vol. 4, no. 4, pp. 377–392.

Mosier, K. L. & Fischer, U. M. 2017, "Judgment and decision making by individuals and teams: Issues, models, and applications" in *Decision Making in Aviation* Routledge, pp. 139–198.

Newen, A., De Bruin, L. & Gallagher, S. 2018, *The Oxford handbook of 4E cognition,* Oxford University Press.

O'Connor, P., Hörmann, H., Flin, R., Lodge, M., Goeters, K. & JARTEL Group, T. 2002, "Developing a Method for Evaluating Crew Resource Management Skills: A European Perspective", *The International Journal of Aviation Psychology,* vol. 12, no. 3, pp. 263–285.

Sassen, C. 2005, *Linguistic Dimensions of Crisis Talk: Formalising Structures in a Controlled Language,* John Benjamins Publishing.

Seah, B. Z. Q., Gan, W. H., Wong, S. H., Lim, M. A., Goh, P. H., Singh, J. & Koh, D. S. Q. 2021, "Proposed Data-Driven Approach for Occupational Risk Management of Aircrew Fatigue", *Safety and Health at Work,* vol. 12, no. 4, pp. 462–470.

Sun, H., Zhou, X., Zhang, P., Liu, X., Lu, Y., Huang, H. & Song, W. 2023, "Competency-based assessment of pilots' manual flight performance during instrument flight training", *Cognition, Technology & Work,* vol. 25, no. 4, pp. 345–356.

Uenking, M. 2000, "Pilot biofeedback training in the cognitive awareness training study (CATS)", *Modeling and Simulation Technologies Conference*, p. 4074.

V. Motogna, G. Lupu-Florian & E. Lupu 2021, "Strategy For Affective Computing Based on HRV and EDA", *2021 International Conference on e-Health and Bioengineering (EHB)*, p. 1.

Wilson, A. D. & Golonka, S. 2013, "Embodied Cognition is Not What you Think it is", *Frontiers in Psychology,* vol. 4.