Categorization of Select Cockpit Performance Evaluation Techniques

Eric M. Brighton; David M. Klaus

INTRODUCTION: The modern aircraft cockpit has evolved into a complex system of systems. Numerous performance evaluation metrics and techniques exist that can measure the effectiveness of cockpit components in terms of how they influence the human operator's ability to perform tasks relevant to mission success. As no prior review of these metrics has been found in the literature, this effort attempts to do so, albeit without applying the metrics to a novel cockpit evaluation.

METHODS: These metrics and techniques are discussed and presented in five defined categories as they relate to evaluating cockpit subsystems: ergonomics and anthropometrics; human-computer interaction; data management and presentation; crew resource management and operations; and ingress and egress.

DISCUSSION: While this effort is significant and novel, it is not necessarily comprehensive. In conclusion, it is noted that no single holistic quantitative metric to evaluate cockpit design and performance yet exists. Utilizing some of the preexisting metrics presented to develop such a metric would be beneficial in efforts to evaluate aircraft cockpit designs and performance, as well as aiding future cockpit designs.

KEYWORDS: human factors, aircraft cockpit, aviation, aircrew performance, human engineering, anthropometrics, human computer interaction, crew resource management.

Brighton EM, Klaus DM. *Categorization of select cockpit performance evaluation techniques*. Aerosp Med Hum Perform. 2023; 94(9):696–704.

Aerospace cockpits and control systems are evolving in ever-more complex workspaces. Presently, the act of flying an aircraft is being supplemented with vast amounts of data management as well as the rise of automation. More often than not, the challenge of aerospace engineering lies in the design of a "system of systems," rather than a focus on traditional disciplinary topics such as aerodynamics and flying qualities calculations. Space exploration and commercial space tourism present other unique rubrics, where cockpits, depending on design and purpose, will be required to function in both horizontal and vertical layouts, ranging from 1-G level flight to high-G ascent and entry accelerations to weightlessness. NASA has identified inadequate crewmember human–computer interaction as a potential issue for future long-duration space missions.[26] Similarly, the U.S. Navy (USN) has sought additional support for human systems engineering and human performance assessment and modeling.[27]

When optimizing the design of a modern aircraft cockpit, whether for crewed or remotely piloted vehicles, for general aviation, commercial flight, or military applications, many factors can influence the effectiveness of the human operator. A USN study in the late 1980s identified 13 distinct technologies that are incorporated in a crewed cockpit, and this number has likely increased since then.[47] While many of these devices are normally running transparently in the background, others require ongoing direct interface or manual intervention in the event of an anomaly. In particular, for military operations on the modern battlefield, poorly designed and implemented cockpit ergonomics and interfaces can significantly, and adversely, impact the effectiveness of combat aircraft as an effective weapons system.[2] And in any flight environment, overreliance on automated systems has been shown to be a concern that can cause piloting skills to atrophy and impact performance in critical situations.[44] For these and other reasons, an

From the University of Colorado-Boulder, Boulder, CO, United States.

This manuscript was received for review in October 2022. It was accepted for publication in June 2023.

Address correspondence to: Mr. Eric Brighton, University of Colorado-Boulder, Aerospace Engineering Sciences Department, 3775 Discovery Dr., Boulder, CO 80303, United States; brighton@colorado.edu.

DOI: https://doi.org/10.3357/AMHP.6185.2023

effective means of assessing how well these integrated cockpit systems perform has become increasingly important for modern cockpit design.

Several individual evaluation metrics exist to assess how well the cockpit system accommodates the needs of the pilot. Various approaches are widely accepted as industry standards (e.g., NASA TLX, Cooper-Harper, etc.) and, although there are recognized shortcomings in some of these methods, they represent the current state-of-the-art for evaluating general, commercial, and military aircraft. Because of the demanding mission goals for military combat aircraft, optimizing the human–machine interface is especially critical. No standardized process, however, could be identified for conducting a systematic evaluation of cockpit system performance in military aircraft, nor could a published means of compiling a comprehensive outcome from the individual metric inputs be found. This gap extends to designing pilot interfaces for remotely piloted drones as well.

The design trade study process employs identification and weighing of individual parameters, as well as codependent factors, that collectively affect the optimal solution differently than their individual performance would suggest (e.g., an ergonomically designed, comfortable seat that is difficult to ingress/egress). Combinations of cockpit systems are not only codependent, but also mission dependent. Therefore, establishing a holistic metric that represents the overall integrated design performance would provide useful insight to cockpit design. This work provides a categorical summary of key existing cockpit performance evaluation techniques as a foundation for defining a comprehensive, integrated approach.

BACKGROUND

The USN desired to standardize a systems engineering approach to cockpit design with the Advanced Technology Crew Station (ATCS) program, with McDonnell Douglas and Boeing electing to participate.[23,30] This new codified approach to aircraft cockpit design sought to overcome the discrepancies where pilot–operator performance was the limiting factor to overall system effectiveness.[23] Ultimately, the results of the ATCS program were never formalized by the Department of Defense, but they influenced future carrier-borne fighter aircraft cockpit design and motivated smaller efforts to formalize approaches to cockpit and aircrew system design.[6] While ATCS produced multiple techniques and tools for development and validation of cockpit subsystem design, none of these deliverables yielded an overall holistic evaluation metric or tool for the entire cockpit.[14,23,30]

ATCS also attempted to firm up connections between research and application in the late 1980s by identifying 13 technologies utilized in cockpit system design, including seating and escape, controls and displays, man–machine functional requirements and interface, and computer/software.[47] From this starting point, one can assume that the number of technologies has only increased in modern cockpits.

The Department of Defense is presently looking toward sixth-generation fighters being fielded in the 2030s.[25] The USN first identified the F/A-XX in June 2008, with Analysis of Alternatives completed in June 2019.[18] A hypothetical next-generation cockpit design would include advanced technologies such as automation, artificial intelligence, augmented reality, embedded or conformal cockpit structures and aerodynamics, and possibly an opaque canopy. Ultimately, this work intends to contribute insight towards achieving that goal.

METHODS

Groupings of quantitative evaluation metrics have been proposed for crew utilization of space habitat systems.[11] Extending this approach to the modern cockpit environment, in both aircraft and piloted spacecraft, would provide an insightful tool for designers to evaluate cockpit and control layouts in the preliminary design phase of the acquisition lifecycle. Research is moving forward in numerous subareas that could ultimately be compiled into a holistic cockpit evaluation metric, which is the motivation of this current study. Approaching these needs from a categorical perspective, a suite of metrics can be grouped into those affecting the following: ergonomics and anthropometrics; human–computer interaction; data management and presentation; operations/crew resource management (CRM); and ingress and egress. The following sections provide a summary of existing cockpit and crew performance evaluation techniques grouped under these functional headings. Defining a standardized multi-metric-based approach merging select combinations of these current techniques would provide cockpit designers with insights from a systems engineering perspective while still in the preliminary design phase, where modifications and optimizations can readily be incorporated before designs are finalized. Additional references are presented in **Table I**, **Table II**, **Table III**, **Table IV**, and **Table V** for completeness but not necessarily thoroughly discussed in the text. While some metrics may offer beneficial evaluation techniques across multiple categories, for the purpose of cataloging, each metric is assigned to its “best fit” category as assessed by the authors.

RESULTS

Ergonomics and Anthropometrics

“Ergonomics” and “anthropometrics” are two terms with unique definitions used when discussing cockpit layouts, the former addressing fit of a workspace and the latter characterizing dimensional reach envelope of the human operator. “Ergonomics” implies designing equipment to maximize how efficiently people are working in an environment. “Anthropometrics” refers to measuring size and proportions of the human body. “Poor ergonomics” is cited as a significant enabling factor of perception errors leading to RAF aviation accidents, including visual illusion, disorientation, and misinterpreted displays.[4]

Table I. Table of Anthropometrics & Ergonomics Evaluation Metrics.

METRIC	DESCRIPTION
Card-Sorting[38]	Contrasts with MCDM as it factors in end user/ pilot's desires in location placement for cockpit items. Pilots place indicators (cards) around the cockpit display panel. Comments and reasoning behind placement are also noted. Preference means and frequency after all (eight) pilots complete the card-sorting are then analyzed to reach a preferred design location.
Hess Force/Feel Metrics[17]	Set of five metrics (that must be all used in conjunction with one another) based on pilot Cooper-Harper ratings and measurement of "stick and rudder" control force inputs. Designed for use in aircraft/cockpit certification.
Different Types of Uncertain Linguistic Multiple Attribute Combination Decision-Making (DTULDM)[5]	Cockpit ergonomics evaluation. Approaches cockpit layout as multiple attribute decision-making (MADM) problem. Very rigorous mathematical background to deal with "fuzziness" of ergonomics evaluation.
Anthropometric Measurements[37]	Helicopter aircrew focus.

In practical application, the U.S. Air Force (USAF) and USN use very similar processes for anthropometric accommodation evaluations of aircraft cockpits. These evaluations are conducted when the introduction of new aircraft or aircraft modifications impact cockpit accommodation. These evaluations identify anthropometric restrictions for a particular aircraft and determine maximum and minimum values for areas such as sitting height, sitting eye height, buttock-knee length, and thumb tip reach. The primary focus is avoiding any impingement on controls or control panels and allowing for egress (i.e., ejection seat) clearance. It is noteworthy that crash-worthiness is not necessarily considered in these anthropometric and ergonomic studies.

Senol presents the largest recent body of work regarding cockpit layout anthropometry. The bulk of this research focused on rotary-wing cockpits, although fixed-wing cockpits were included. It addresses the "dialogue between the operator and the device" with both quantitative and qualitative metrics: multi-criteria decision-making (MCDM) (quantitative) and card-sorting (qualitative metric).[37,38] MCDM is referenced in subsequent research evaluating positioning of analog indicators on display panels.[37] The focus shifts to pilot visual field and ability to reach necessary cockpit controls and displays. The research recommendations include that pilot selection may be necessary to limit cockpit design accommodations.[36] While both MCDM and card-sorting may be useful in evaluating ergonomics and anthropometrics, MCDM may yield the greatest benefit as a data management and presentation evaluation tool and is categorized as such.

Multiple Chinese research groups are active in cockpit ergonomics research, some to an exhaustive level of mathematical detail. Zhang and Sun go as far as discussing coordinate transformation between pilot torso and cockpit reference frames when discussing cockpit layouts and body movements required of aircrew. A thorough examination of different flight profiles and regimes is presented, decomposing each in-flight task into items requiring either pilot attention or inputs.[50] Chen et al. discuss differences between Russian, American, and Chinese cockpit designs, but focuses on Chinese aircraft characteristics and quantitative evaluations with multiple attribute decision-making.[5] Similarities may be drawn to Senol's MCDM metric categorized in data management and presentation.[37] Wang et al. evaluate cockpit ergonomics in a virtual CAD model, which could be very useful in the preliminary design review stages of a cockpit design lifecycle.[46] However, they do not present their evaluation criteria in detail. Their scope has been limited to commercial aircraft cockpits.

Table II. Table of HCI Evaluation Metrics.

METRIC	DESCRIPTION
Communication-Human Information Processing (CHIP)[39]	Risk communication and warning/ACAWS evaluation tool. Based on model for traffic signs, medicine and food labels, etc. Includes framework based on source, channel, attention, memory, attitudes, motivation, and behavior. Uses various methods to measure attention, including eye movement, Detection or Reaction Time (D-RT), and Self Reports questionnaires.
Detection or Reaction Time (D-RT)[39,48]	Direct method of measuring attention. Susceptible to false positives when used to judge if participants have detected a warning. Cheaper and easier when compared to eye movement measurements, but drawback in that the researcher does not gain information on the visual path taken to locate the warning.
Crew Station Design Tool (CSDT)[45]	Proprietary software tool to help designers optimize displays and controls layout in workstations, with a focus on fixed-wing aircraft. Prioritizes criteria such as frequency of use, sequence of use, and both.
Stanton Input Device Evaluation[41]	Evaluation of cognitive and physical performance of menu navigation devices.
Modulation Transfer Function (MTF)[19 (p. 380)]	Developed in CRT world, unclear if LED/Flat Panel Display would require modifications.
Modulation Transfer Function Area (MTFA)/TQF[19 (p. 381)]	Developed in CRT world, unclear if LED/Flat Panel Display would require modifications.
Snyder's Threshold Sensitivity Curve[19 (p. 381)]	Further adaption of MTF requiring psychophysical experiments; developed in CRT world, unclear if LED/Flat Panel Display would require modifications.
Haworth-Newman Avionics Display Readability Scale[28 (p. 97)]	Naval Postgrad School proposal; developed in CRT world, unclear if LED/Flat Panel Display would require modifications.
Perceivable Just Noticeable Differences (PJND)[19 (p. 160)]	Measure color and luminance difference, used in (relatively) recent Eurofighter development; developed in CRT world, unclear if LED/Flat Panel Display would require modifications.

Table III. Table of Relevant Data Management Evaluation Metrics.

METRIC	DESCRIPTION
Multi Criteria Decision-Making (MCDM)[37]	On a 0–100 scale, pilots are evaluated on misapprehending information on flight safety and the frequency of that information's use during flight. This is graded on 24 locations in the cockpit. Disadvantage is that the pilot's desires for location placement are not factored in, only the design engineer's.
Eye-Movement Patterns[16 (p. 346)]	Utilized eye-movement records to determine in which order pieces of text were read.
Cloze Tests[16 (pp. 346-347)]	Text-reading comprehension test concurrent while user is reading.
Comprehension Tests[16 (p. 351)]	Text-reading comprehension test after user has completed reading.
Johnson Criteria/Detect Identify Recognize[19 (p. 118)]	Standard practice for evaluating sensor-assisted vision.
USAF Tri Bar Test Pattern[19 (p. 379)]	Standard practice for evaluating display image quality.
Display Readability Rating[28 (p. 111)]	Modified Cooper-Harper.
Display Flyability Rating[28 (p. 111)]	Modified Cooper-Harper.
HUD Optical Measurements/HUD Photometric Measurements[28 (p. 114)]	Often restricted to laboratory environment.

Hess focuses indirectly on ergonomics by developing metrics for evaluating pedal force and feel systems in transport aircraft. Using these metrics in conjunction with Cooper-Harper ratings allows for an evaluation of "stick and rudder" cockpit control layouts.[17]

Human–Computer Interaction (HCI)

The ability of aircrew members to interact with the increasingly digital systems of modern aircraft is critical to optimal cockpit design and performance, which falls under the field of HCI. Singer and Dekker provide an insightful body of work for evaluating cockpit design focusing on HCI.[40] Singer's background as a military test pilot leverages a high degree of operational knowledge to the cockpit design evaluation problem. Although dated, they walk through an excellent synopsis of 2001 European commercial aircraft cockpit certification, identifying a gap between technical focus (which, they suggest, the European Union and Federal Aviation Administration can become fixated on) at the expense of operability, and rely entirely on subjective evaluations for human–machine interactions. These subjective evaluations are prone to pilot-to-pilot biases.[40] Singer's dissertation deviates from quantitative metrics on cockpit design validation, proposing a modified design process for commercial air transport aircraft based on flight test

Table IV. Table of Operations/CRM Evaluation Metrics.

METRIC	DESCRIPTION
NASA-TLX (Task Load Index)[15 (p. 541),39]	Mental workload measurement for multidimensional characteristics. Considered the most widely used metric because of ease in administering. Participants rate 6 different scales in 20 intervals, ratings are then converted to values of 0–100. Laboratory research-based. Dimensions are also evaluated on relevance, and then weighted. Sometimes used in simulations but limited application in real-world flight operations.
SWAT (Subjective Workload Assessment Technique)[22 (p. 533)]	More time-consuming than NASA-TLX, requiring an hour to fully implement. Also multidimensional. Unique feature is that it's based on psychological model of how judgments of mental workload are formed by participants. Rate on three dimensions, each scale having 3 points. (1-1-1 = lowest workload, 3-3-3 = highest workload).
Human Factors Analysis and Classification Tool (HFACS)[32,39]	Uses data gathered from accident investigations to develop human error classifications into operational error. Less useful as a design evaluation tool and more of a human error event analysis tool. Also lacks references to novel cockpit technologies.
Bedford Workload Scale[28 (p. 72)]	Modified Cooper Harper. Captures workload, does nothing for capturing performance.
Defense Research Agency Workload Scale (DRAWS)[19 (p. 89)]	Similar to TLX, scale 0–100 (and beyond for overload).
Jarret's Three Classes of Objective Assessment Techniques[19 (p. 89)]	Measure performance directly—difficult to quantify all tasks this way. Loading operator to maximum sustainable effort. Assessment of physiological variables (blink rate, heart rate, blood pressure, heart rate variation/arrythmia, sweat rate, muscle tension, and the concentration of adrenal hormone secretions in the blood and urine).
W/INDEX[29]	Workload analysis tool based on Wickens' Multiple Resource Theory.
China Lake Situation Awareness (CLSA)[13,28 (pp. 59-66)]	Subjective questionnaire.
Crew Situation Awareness (CSA)[13,28 (pp. 59-66)]	Observers rate crew coordination.
Situation Awareness Global Assessment Technique (SAGAT)[13,28 (pp. 59-66)]	Intrusive questionnaire in simulator scenarios.
Situation Awareness Probe (SAP)[13,28 (pp. 59-66)]	Questionnaire similar to instructor pilot questions during pilot training.
Situation Awareness Rating Technique (SART)[10,13,28 (pp. 59-66)]	Subjective analysis tool where operators rate a system design based on the demand for attentional resources, supply of resources, and understanding of overall situation provided in a given scenario.
Situation Awareness Subjective Workload Dominance (SA-SWORD)[13,28 (pp. 59-66)]	Paired judgment rating matrix that produces numerical output.
Situation Awareness Supervisory Rating Form (SASRF)[13,28 (pp. 59-66)]	Peer evaluation of other pilots/crew members.
Physiological Measurements[22 (p. 531)]	Cardiac activity, heart rate, blood pressure, brain activity, etc.

Table V. Table of Ingress and Egress Evaluation Metrics.

METRIC	DESCRIPTION
Time Measurement[12]	Egress time.
Available Safe Egress Time[50]	Commercial aircraft passenger safety and behavior modeling in emergency situations.

experience and empirical results.[39] Singer discusses multiple facets of human factors involved in crewed aircraft flight, including task saturation,[15] aircrew training,[32] situational awareness, reaction time, and human information processing.[48]

Stanton et al. thoroughly investigated and evaluated different methods of aircraft display control inputs: trackballs, rotary controllers, touch pads, and touch screens.[41] Their conclusions are varied for different tasks but identified touch screens as the most commonly "best rated" input devices for multiple input tasks.

The published work of Marstall et al. focuses on marketing their proprietary cockpit design. However, their previous research may serve to expand the knowledge base of cockpit display and instrument evolution, particularly with regards to improving legacy aircraft flight instruments.[21]

The research scope of Walters et al. was very limited, serving more as a white paper for their proprietary design tool rather than presenting research conclusions. It still is noteworthy for putting forward a framework for evaluating cockpit designs based on flight regime and function.[45]

Data Management and Presentation

A clear example of data management and presentation is depicted in **Fig. 1**. On the left side is the traditional "steam gauge" or "blind flying" panel universally accepted in aviation since its development in the 1930s by William Ocker and Jimmy Doolittle, until modern electronic displays supplanted analog gauges in commercial and military aircraft in the last 30 yr. On the right side is an example of a modern "glass cockpit" display of an Attitude Display Indicator (ADI) and Horizontal Situation Indicator (HSI). Noteworthy is the fact that both left and right presentations display exactly the same data relevant to aircraft operation (i.e., indicated airspeed, altitude, heading, navigational aid data referenced to a VOR navaid). It's simply two different methods of presenting the same data. Analogous to a digital vs. analog wristwatch, one or the other display presentation may be more comfortable to different population sets of aircrew.

Common aircraft data displays include the now ubiquitous (at least as far as commercial airline and military aircraft are concerned) Heads Up Display (HUD). The next evolution of the HUD, specifically in the realm of tactical combat aircraft, is Helmet Mounted Displays. Heads Down Displays include Multi-Function Display (MFD), and the aforementioned analog "steam gauges." Advisory Cautions and Warning System (ACAWS) audible aircraft alerts provide data presentation in an auditory capacity. ACAWS in an aircraft system should not be confused with the NASA Orion Advanced Caution and Warning System, which is a spacecraft application for use in NASA's next-generation crewed spacecraft beyond low Earth orbit to provide autonomous alert monitoring and relaying.

Thomas and Rantanen focus on human factors involving a pilot's ability to process air traffic information. Their approach segments into computer display, such as alerting algorithms and false alarms, and other human factors issues such as aircrew workload and display dimensionality. They ultimately identify a gap requiring quantitative analysis of newer technology and cockpit displays.[43]

Dehais et al. broach the topic of anesthesia induced by sterile cockpit operations, where the sterile auditory environment in a cockpit during certain flight regimes can lull an aircrew into inattention.[8] This is also discussed by Broom et al. but with a greater focus on Cockpit/Crew Resource Management.[3] Dehais focuses on higher workload flight regimes (i.e., approach and landing) where 40% of a population of general aviation pilots failed to detect an auditory, critical alarm. They propose case-based learning as a solution to inattentiveness or auditory alarm misperception. Other research is investigating the implementation of cockpit display of traffic information (CDTI) and how associated alerts can be better tuned based on pilot preference and how cockpit alerting system test methods can be refined.[31,43] Kolbeinsson et al. published a high

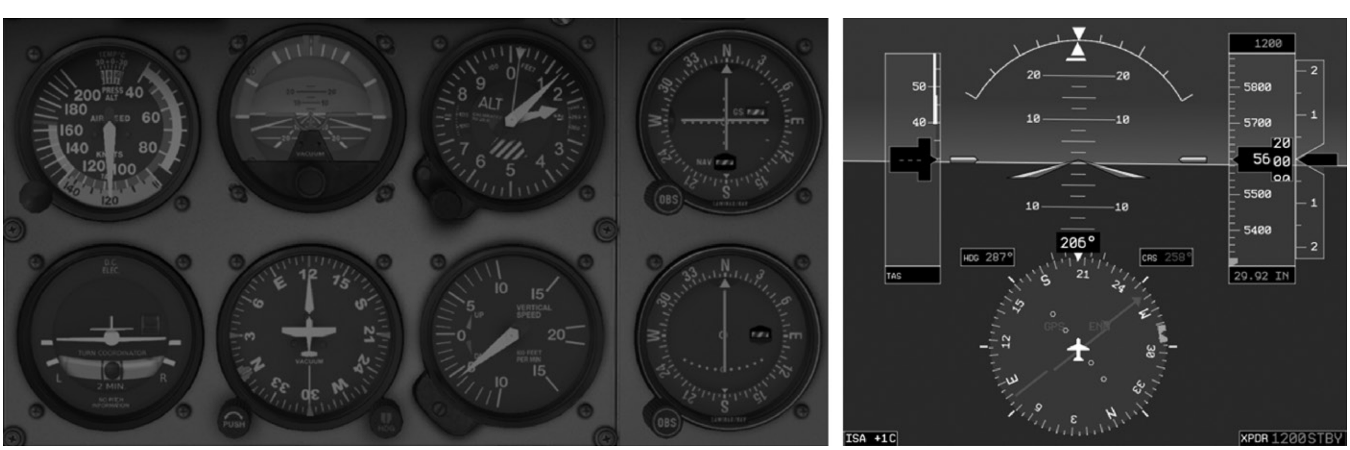

Fig. 1. Comparison of legacy steam gauge blind flying panel to attitude display indicator and horizontal situation indicator.

level investigation of different cockpit display icons with designs varying in shape, size, and color to denote additional information with a case study of aircraft "friend or foe" identification and threat level.[20]

One of Senol's previously discussed quantitative metrics, MCDM takes a quantitative approach to evaluating both anthropometrics and interaction between operators and cockpit devices in rotary wing aircraft. It appears to offer the greatest benefit as an evaluation technique categorized under Data Management & Presentation, although as with some metrics it may be useful applied across more than one category.[38]

Operations/Crew Resource Management (CRM)

Crew Resource Management can be defined as "cognitive, social and personal resource skills that complement technical skills and contribute to safe and efficient task performance".[7] CRM has a quantifiable impact on crew performance, but how does cockpit influence it? Recent evolutions of the Lockheed C-130 aircraft have removed the Flight Engineer crew position from the US Air Force aircrew, now relying on computer generated diagnostic codes to prompt the aircrew members for in-flight maintenance advisory items. While this yields benefits for streamlining aircrew training and staffing, it may not necessarily prove to be beneficial for CRM in workload intensive phases of flight. Similar examples are presented in the burgeoning arena of Remotely Piloted Aircraft and their associated crew stations.

A popular method of capturing workload is by subjective/operator opinion techniques, or subjective measures of mental workload.[22] The Bedford Workload Scale is essentially a modified Cooper Harper evaluation scale applied to workload, with the noteworthy constraint that it does nothing to capture performance.[33] NASA Task Load Index (NASA-TLX) is another widely technique that is applied in simulations but limited in real world flight operations due to required interruption to complete the questionnaire.[22] Farmer's Defense Research Agency Workload Scale (DRAWS) is similar to NASA-TLX on a scale of 0 to 100 and beyond for task overload.[19]

Jarrett categorizes three classes of objective assessment techniques:[19]

1. Measuring performance directly, which is difficult to quantify all tasks this way.
2. Loading operator to maximum sustainable effort.
3. Assessment of physiological variables (blink rate, heart rate, blood pressure, heart rate variation/arrythmia, sweat rate, muscle tension and the concentration of adrenal hormone secretions in the blood and urine).

Similarly, Megaw suggests empirical, analytical, psychophysiological, and the previously mentioned subjective/operator opinion techniques.[22] Indeed the advent of eye tracking technology, as well as psychophysiological data acquisition systems, have both become increasingly commonly used for capturing aircrew workload and performance in recent years.

It's noteworthy when Eggemeier and Wilson conclude that a battery of measurements is required to capture workload in multitask environments,[9] and R. Newman and Greeley propose using a combination of test and evaluation techniques and success criteria as no single metric defines acceptability.[28] W/INDEX stands out as a predictive workload analysis tool based on Wickens' Multiple Resource Theory.[29]

Numerous other metrics and techniques to assess operator and crew situational awareness are captured by Gawron et al. and R. Newman and Greeley.[13,28] These primarily take the form of assessment questionnaires and include the following:

1. China Lake Situation Awareness (CLSA), a subjective questionnaire;
2. Crew Situation Awareness (CSA), in which observers rate crew coordination;
3. Situation Awareness Global Assessment Technique (SAGAT), a well-known, objective, knowledge-based approach developed, but can be an intrusive questionnaire in simulator scenarios;
4. Situation Awareness Probe (SAP), a questionnaire similar to IP questions during pilot training;
5. Situation Awareness Rating Technique (SART);
6. Situation Awareness Subjective Workload Dominance (SA-SWORD), a paired judgement rating matrix that produces numerical output; and
7. Situation Awareness Supervisory Rating Form (SASRF), a peer evaluation of other pilots/crew members.

Crichton examined five principles for improving simulator-based training for teams involved in complex, technical work environments.[7] While not focused on CRM, it discussed similar nontechnical skills being applied to improve overall team performance, as well as identifying behavioral markers and evaluation metrics. Broom et al. discuss inattentiveness induced by a sterile auditory environment in a cockpit, with results measured in different sound environments.[3] Contrasting their approach from the Data Management effort of Dehais et al. was the focus on CRM issues in a sterile cockpit environment.[8]

Salas et al. investigated the lack of empirical studies supporting findings that CRM implementation is beneficial. This is noteworthy as the US Navy and US Air Force often cite their own internal statistics for CRM improving aircraft mishap rates significantly. However, Salas et al. may be the first published academic investigation into the impact of CRM.[35] While the focus of Miller and Hannen's research is on implementing a new rotorcraft user interface for cockpit information management not previously utilized, in doing so they broke down pilot-perceptible behaviors in several categories for analysis. Initial trials of subjective evaluations of cockpit layouts were supplemented with objective performance data.[24] In a separate study Russi-Vigoya and Patterson took the unique approach of investigating eye fixation of private pilots utilizing glass cockpits in flight simulations, showing common trends during failures and poor weather conditions.[34]

Ingress and Egress

The ultimate measurement of how well a cockpit is designed for ingress and egress, particularly emergency egress, is time, followed by required assistance. These are the two metrics set forth for by the Federal Aviation Administration but specifically pertaining to passengers on transport category aircraft.[12] Ejection seat technology features heavily in tactical military aircraft cockpit designs and remains a primary consideration due to the specific design requirements for ejection seat use (i.e., ensuring aircrew member's limbs are not impinged during ejection).

Stedmon et al. compare human behavior modeling between passenger airline and railroad environments, particularly noting the lack of research on passenger rail egress models and the abundance of passenger airline research.[42] While research focuses on aircrew egressing a cockpit specific environment would be of greater significance than passenger cabin egress, it still yields potential interest in commercial spaceflight (i.e., space tourism) applications. Zhang et al. goes into further detail on commercial aircraft passenger safety and behavior modeling in emergency situations. They discuss two specific models—the Fire Dynamics Simulator and Pathfinder—to measure safe egress from an Airbus A380 case study.[49]

Bienefeld and Grote analyzed commercial aviation cockpit and cabin crews and their behavior in simulated in-flight emergencies and developed a structured observation scheme to objectively evaluate crew performance. This could potentially be factored in to determining useful in how cockpit design facilitates or hinder crew performance during emergencies.[1]

DISCUSSION

No previous wide-ranging review focusing on aircraft cockpit design and performance evaluation metrics and techniques was found in the literature. As such, the information presented here attempts to summarize numerous evaluation methods across multiple aspects of cockpit design and performance. The metrics and techniques summarized here cover many of the commonly used preexisting subjective and objective approaches; however, this should not be considered a definitively comprehensive list.

Multiple research efforts, combined with research gaps identified by NASA and the US Navy, emphasize the importance of human-computer interaction during crewed flight. Quantitative evaluation metrics such as these allow for initial evaluation of emerging aircraft and spacecraft systems early in the design process. Table I, Table II, Table III, Table IV, and Table V present a brief overview of commonly used quantitative metrics focused on aircraft cockpit subsystems and human factors considerations. Each table presents the metrics and techniques cataloged into five categories covering aspects of cockpit design. These five categories are graphically depicted in **Fig. 2** in sequential order as an aircrew member may encounter them in aircraft operations.

As demonstrated in the literature, numerous metrics and methodologies exist to evaluate design effectiveness and performance of aircraft cockpit components and subsystems intended to aid human operators in task and mission

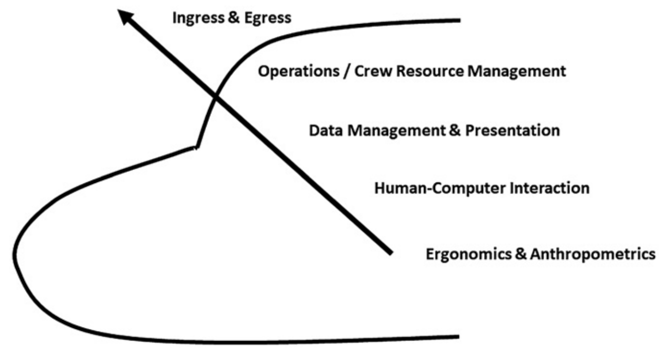

Fig. 2. Cockpit evaluation categories.

accomplishment. It would be useful to designers, systems engineers and the flight test community in evaluating and optimizing modern aircraft cockpit designs to combine these categories into a comprehensive quantitative evaluation metric to evaluate the cockpit system as a whole. By leveraging preexisting, traditionally used metrics as described here, one can quantify the effectiveness of various aircraft cockpit subsystems design and performance based on the human operator's ability to perform the intended tasks. However, to the best of our knowledge, no quantitative holistic method or metric for evaluating overall cockpit design and performance yet exists. In conclusion, it may be possible with further research to develop and validate such a holistic methodology to evaluate the integrated aircraft cockpit system performance as a means for improving the design and operations of the vehicle.

ACKNOWLEDGEMENTS

Financial Disclosure Statement: The authors have no competing interests to declare.

Authors and Affiliations: Eric M. Brighton, M.S., B.S., Aerospace Engineering Sciences (Ph.D. candidate), and David M. Klaus, Ph.D., M.S., Aerospace Engineering Sciences (Professor), University of Colorado, Boulder, Colorado, United States.

REFERENCES

1. Bienefeld N, Grote G. Shared leadership in multiteam systems: how cockpit and cabin crews lead each other to safety. Hum Factors. 2014; 56(2):270–286.
2. Bronk J. The mysterious case of the missing Russian Air Force. RUSI. 2022. [Accessed July 15, 2022]. Available from: https://rusi.org/explore-our-research/publications/commentary/mysterious-case-missing-russian-air-force.
3. Broom MA, Capek AL, Carachi P, Akeroyd MA, Hilditch G. Critical phase distractions in anaesthesia and the sterile cockpit concept. Anaesthesia. 2011; 66(3):175–179.
4. Chappelow JW. Error and accidents. In: Rainford D, Gradwell DP, Ernsting J, editors. Ernsting's aviation medicine. 4th ed. London (UK): CRC Press; 2006:349–358.
5. Chen J, Yu S, Wang S, Lin Z, Liu G, Deng L. Aircraft cockpit ergonomic layout evaluation based on uncertain linguistic multi-attribute decision-making. Adv Mech Eng. 2014; 6:698159.

6. Cohen D. Identification of advanced technology crew station decision points and information requirements. Defense Technical Information Center; 1993:356. Report No.: ADA326336. [Accessed 10 July 2023]. Available from: https://apps.dtic.mil/sti/citations/ADA326336.
7. Crichton MT. From cockpit to operating theatre to drilling rig floor: five principles for improving safety using simulator-based exercises to enhance team cognition. Cogn Technol Work. 2017; 19(1):73–84.
8. Dehais F, Causse M, Vachon F, Régis N, Menant E, Tremblay S. Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. Hum Factors. 2014; 56(4):631–644.
9. Eggemeier FT, Wilson GF. Performance-based and subjective assessment of workload in multi-task environments. In: Damos DL, editor. Multiple-task performance. London (UK): Taylor & Francis; 1991:217–276. [Accessed 10 July 2023]. Available from: https://www.taylorfrancis.com/chapters/edit/10.1201/9781003069447-13/performance-based-subjective-assessment-workload-multi-task-environments-thomas-eggemeier-glenn-wilson.
10. Endsley MR. Situation awareness measurement in test and evaluation. In: O'Brien TG, Charlton SG, editors. Handbook of human factors testing and evaluation. Mahwah (NJ): Lawrence Erlbaum Associates; 1996:159–80.
11. Fanchiang C. A quantitative human spacecraft design evaluation model for assessing crew accommodation and utilization [Ph.D. Dissertation]. Boulder (CO): University of Colorado Boulder; 2017.
12. Federal Aviation Administration. Federal aviation regulations. Washington (DC): Federal Aviation Administration; 2023. Sec. 25.803. Emergency evacuation. [Accessed 10 July 2023]. Available from https://www.ecfr.gov/current/title-14/chapter-I/subchapter-C/part-25/subpart-D/subject-group-ECFR88992669bab3b52/section-25.803.
13. Gawron V, Weingarten N, Hughes T, Adams S. Verifying situational awareness associated with flight symbology. Paper presented at the AIAA 37th Aerospace Sciences Meeting and Exhibit; January 11–14, 1999; Reno, NV. Reston (VA): American Institute of Aeronautics and Astronautics; 1999.
14. Glenn F, Boardway J, Wherry R Jr, Cohen D, Carmody M. The advanced technology crew station: development and validation of a workload assessment technique for cockpit function allocation. Defense Technical Information Center; 1993:57. Report No.: ADA279544. [Accessed 10 July 2023]. Available from https://apps.dtic.mil/sti/citations/ADA279544.
15. Hart SG, Staveland LE. Development of NASA-TLX (Task Load Index): results of empirical and theoretical research. Adv Psychol. 1988; 52:139–183.
16. Hartley J. Is this chapter any use? Methods for evaluating text. In: Wilson JR, Corlett EN, editors. Evaluation of human work. 3rd ed. Boca Raton (FL): Taylor & Francis; 2005:335–351.
17. Hess RA. Metrics for the evaluation of pedal force/feel systems in transport aircraft. J Aircr. 2008; 45(2):651–662.
18. Insinna V. At a budgetary crossroads, the US Navy's aviation wing must choose between old and new. Defense News. 2020. [Accessed October 20, 2022]. Available from https://www.defensenews.com/air/2020/06/01/at-a-budgetary-crossroads-the-us-navys-aviation-wing-must-choose-between-old-and-new/.
19. Jarrett DN. Cockpit engineering. Abingdon (UK): Routledge; 2005:89.
20. Kolbeinsson A, Falkman G, Lindblom J. Showing uncertainty in aircraft cockpits using icons. Procedia Manuf. 2015; 3:2905–2912.
21. Marstall J, Miller ME, Poisson RJ. Collaboration in the cockpit: human–system interaction beyond the autopilot. Ergon Des. 2016; 24(1):4–8.
22. Megaw T. The definition and measurement of mental workload. In: Wilson JR, Corlett EN, editors. Evaluation of human work. 3rd ed. Boca Raton (FL): Taylor & Francis; 2005:525–547.
23. Mejzak RS, Sparta ML, Warner NW. Crew system engineering methodology: process and display requirements. Paper presented at the IEEE/AIAA 10th Digital Avionics Systems Conference; October 14–17, 1991; Los Angeles (CA). New York (NY): IEEE; 1991:387–392.
24. Miller CA, Hannen MD. The rotorcraft pilot's associate: design and evaluation of an intelligent user interface for cockpit information management. Knowl Based Syst. 1999; 12(8):443–456.
25. Mizokami K. "Sixth generation" fighters jets are already taking shape. Popular Mechanics. 2017. [Accessed August 27, 2018]. Available from: https://www.popularmechanics.com/military/aviation/a25832/sixth-generation-fighter-jets-already-taking-shape/.
26. NASA. Human research roadmap, risk of inadequate human-computer interaction. [Accessed January 10, 2017]. Available from: https://humanresearchroadmap.nasa.gov/risks/risk.aspx?i=164.
27. Naval Air Systems Command. NAVAIR science and technology basic and applied research solicitation number: NAWCAD-BAA-17-03. [Accessed January 10, 2017]. Available from: https://www.fbo.gov/?s=opportunity&mode=form&id=854b996fbc57aab22547702b50d10132&tab=core&_cview=0.
28. Newman RL, Greeley KW. Cockpit displays: test and evaluation. Abingdon (UK): Routledge; 2016:59–76.
29. North RA, Riley VA. W/INDEX: a predictive model of operator workload. In: McMillan GR, Beevis D, Salas E, Strub MH, Sutton R, Van Breda L, editors. Applications of human performance models to system design. Boston (MA): Springer US; 1989:81–89. [Accessed July 10, 2023]. Available from: http://link.springer.com/10.1007/978-1-4757-9244-7_6.
30. Notaro J. Crew centered armament system for high technology cockpit. 1999:9–10. Report No.: ADA368287. [Accessed 10 July 2023]. Available from: https://apps.dtic.mil/sti/citations/ADA368287.
31. Pritchett AR, Vándor B, Edwards K. Testing and implementing cockpit alerting systems. Reliab Eng Syst Saf. 2002; 75(2):193–206.
32. Rigner J, Dekker S. Modern flight training – managing automation or learning to fly? In: Dekker S, Hollnagel E, editors. Coping with computers in the cockpit. 1st ed. Hants (UK): Ashgate; 1999:145–151. [Accessed 2022 Sept. 12]. Available from https://www.taylorfrancis.com/books/9780429864216.
33. Roscoe AH, Ellis GA. A subjective rating scale for assessing pilot workload in flight: a decade of practical use. Farnborough (UK): Royal Aerospace Establishment Farnborough; 1990. Report No.: TR 90019. [Accessed 10 July 2023]. Available from https://apps.dtic.mil/sti/citations/ADA227864.
34. Russi-Vigoya MN, Patterson P. Analysis of pilot eye behavior during glass cockpit simulations. Procedia Manuf. 2015; 3:5028–5035.
35. Salas E, Fowlkes JE, Stout RJ, Milanovich DM, Prince C. Does CRM training improve teamwork skills in the cockpit?: two evaluation studies. Hum Factors. 1999; 41(2):326–343.
36. Şenol MB. Anthropometric evaluation of cockpit designs. Int J Occup Saf Ergon. 2016; 22(2):246–256.
37. Şenol MB, Dağdeviren M, Çilingir C, Kurt M. Display panel design of a general utility helicopter by applying quantitative and qualitative approaches. Hum Factors Man. 2010; 20(1):73–86.
38. Şenol MB, Dagdeviren M, Kurt M, Cilingir C. Evaluation of cockpit design by using quantitative and qualitative tools. Paper presented at 2009 IEEE International Conference on Industrial Engineering and Engineering Management. December 8–11, 2009; Hong Kong, China. New York (NY): IEEE; 2009:847–851.
39. Singer G. Methods for validating cockpit design: the best tool for the task. Royal Institute of Technology, Dept. of Aeronautics; 2002. [Accessed July 10, 2023]. Available from https://www.diva-portal.org/smash/get/diva2:9102/FULLTEXT01.pdf.
40. Singer G, Dekker S. The ergonomics of flight management systems: fixing holes in the cockpit certification net. Appl Ergon. 2001; 32(3): 247–254.
41. Stanton NA, Harvey C, Plant KL, Bolton L. To twist, roll, stroke or poke? a study of input devices for menu navigation in the cockpit. Ergonomics. 2013; 56(4):590–611.
42. Stedmon A, Lawson G, Lewis L, Richards D, Grant R. Human behaviour in emergency situations: comparisons between aviation and rail domains. Secur J. 2017; 30(3):963–978.
43. Thomas LC, Rantanen EM. Human factors issues in implementation of advanced aviation technologies: a case of false alerts and cockpit displays of traffic information. Theor Issues Ergon Sci. 2006; 7(5): 501–523.
44. Volz KM, Dorneich MC. Evaluation of cognitive skill degradation in flight planning. J Cogn Eng Decis Mak. 2020; 14(4):263–287.
45. Walters B, Bzostek J, Small R, Stachowiak C. Comparing methods for placing controls and displays in a cockpit. Proc Hum Factors Ergon Soc Annu Meet. 2004; 48(20):2431–2435.

46. Wang L, Wei X, He X, Sun X, Yu J, et al. The virtual evaluation of the ergonomics layout in aircraft cockpit. Paper presented in 2009 IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design. November 26–29, 2009; Wenzhou, China. New York: IEEE; 2009:1438–1442.
47. Winslow L, Warner N, Thomasson T. Advanced Technology Crew Station (ATCS) design methodology. Washington (DC): U.S. Navy; 1989.
48. Wogalter MS, Hancock PA, Dempsey PG. On the description and definition of human factors/ergonomics. Proc Hum Factors Ergon Soc Annu Meet. 1998; 42(10):671–674.
49. Zhang Q, Qi H, Zhao G, Yang W. Performance simulation of evacuation procedures in post-crash aircraft fires. J Aircr. 2014; 51(3):945–955.
50. Zhang Y, Sun Y. Reuse of pilot motions for improving layout design of aircraft cockpit. J Comput. 2013; 8(9):2269–2276.