Human Factors Information and Processes: FAA's Aircraft Evaluation Division

A Review of the Literature

Kim Cardosi, Ph.D.
Sarah Yahoodik, Ph.D.
Tracy Lennertz, Ph.D.

Final Report —February 2024

DOT-VNTSC-FAA-24-04

Prepared for:
US Department of Transportation
Federal Aviation Administration
800 Independence Avenue, SW
Washington, D.C. 20591

U.S. Department of Transportation
Volpe Center

Notice

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof.

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*
February 2024

2. REPORT TYPE
Final Report

3. DATES COVERED *(From - To)*

4. TITLE AND SUBTITLE
Human Factors Information and Processes for FAA's Aircraft Evaluation Division – A Review of the Literature

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kim Cardosi, Sarah Yahoodik, Tracy Lennertz

5d. PROJECT NUMBER
FG04C122, FG04C222

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
US Department of Transportation
John A Volpe National Transportation Systems Center
220 Binney St
Cambridge, MA 02142-1093

8. PERFORMING ORGANIZATION REPORT NUMBER
DOT-VNTSC-FAA-24-04

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
US Department of Transportation
Federal Aviation Administration
800 Independence Avenue, SW
Washington, D.C. 20591

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Aircraft Certification, Safety, and Accountability Act requires improved integration of human factors in the Federal Aviation Administration's regulatory processes and material. This paper reviews research that could contribute to the data and processes that would be useful to the FAA in Aircraft Evaluation Division (AED) tasks. The more specific the human factors information, however, the less generalizable the finding and will need to be interpreted and applied within the context of the knowledge of specific functions. The best use of the information and processes in the literature would be to develop tools and specific guidance for AED personnel. These tools could increase the standardization of AED tasks and maintain the flexibility of the AED to decide what needs to be assessed, how it should be assessed, and how the results should be interpreted and applied. Research needed to progress the development of these tools and guidance is identified as well as recommended approaches. A repository of AED decisions would be useful in helping to guide decisions and increase standardization. Finally, the feasibility of de-identifying and compiling data on pilot performance and issues identified by pilots after final design approval should be explored.

15. SUBJECT TERMS
Aircraft Evaluation Division, Flight Standardization Board, aircraft certification, decision making, operational evaluation, human factors.

16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	Unlimited	44	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

SI* (MODERN METRIC) CONVERSION FACTORS

Symbol	When You Know	Multiply By	To Find	Symbol
APPROXIMATE CONVERSIONS TO SI UNITS				
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
		NOTE: volumes greater than 1000 L shall be shown in m^3		
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
oz	ounces	28.35	grams	g
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa
APPROXIMATE CONVERSIONS FROM SI UNITS				
Symbol	**When You Know**	**Multiply By**	**To Find**	**Symbol**
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
mL	milliliters	0.034	fluid ounces	fl oz
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
g	grams	0.035	ounces	oz
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	Kilopascals	0.145	poundforce per square inch	lbf/in^2

*SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

Preface

This report was prepared by the Transportation Human Factors Division of the Safety Management and Human Factors Technical Center at the United States (U.S.) Department of Transportation (DOT), John A. Volpe National Transportation Systems Center. This work is funded by the Federal Aviation Administration (FAA) Human Factors Division (ANG-C1), in support of technical sponsors in the Aviation Safety (AVS) organization. We thank Dr. Kathy Abbott, Colleen Donovan, Christy Helgeson, Dr. Heidi Kim, Pam Munro, Clark Davenport, the Aircraft Evaluation Division, Dr. Chuck Perala, Eddie Austrian, and Kevin Siragusa for technical guidance on this task.

The views expressed herein are those of the authors and do not necessarily reflect the views of the Volpe National Transportation Systems Center or the United States Department of Transportation.

For questions or comments, please contact Kim Cardosi, kim.cardosi@dot.gov.

Contents

List of Abbreviations

Abbreviation	Term
AC	Advisory Circular
ACSAA	Aircraft Certification, Safety, and Accountability Act
ACT ARC	Air Carrier Training Aviation Rulemaking Committee
AED	Aircraft Evaluation Division
AFM	Airplane Flight Manual
AFS	FAA Office of Safety Standards of the Flight Standards Service
AFX	FAA Flight Standards Service
AIR	FAA Aircraft Certification Service
ANG-C1	Human Factors Division
ASIAS	Aviation Safety Information Analysis and Sharing
ASRS	Aviation Safety Reporting System
AVS	FAA Aviation Safety Organization
CFR	Code of Federal Regulations
EFB	Electronic Flight Bag
ESL	English as a Second Language
FAA	Federal Aviation Administration
FMS	Flight Management System
FSB	Flight Standardization Board
FSIMS	Flight Standards Information Management System
HF	Human Factors
IP	Issue Paper
NTSB	National Transportation Safety Board
OEM	Original Equipment Manufacturer
OSL	Operational Suitability Letter
OSR	Operational Suitability Report
POI	Principal Operations Inspector
RA	Resolution Advisory
SOP	Standard Operating Procedure
STC	Supplemental Type Certificate
TA	Traffic Advisory
TC	Type Certificate
TCAS	Traffic Alert and Collision Avoidance System

Executive Summary

The Aircraft Certification, Safety, and Accountability Act (ACSAA) requires improved integration of human factors in Federal Aviation Administration (FAA) regulatory processes and materials. The current research was conducted by human factors specialists at the US DOT Volpe National Transportation Systems Center to help improve the integration of human factors in Aircraft Evaluation Division (AED) assessments involving transport airplanes within context of Title 14 Code of Federal Regulations (CFR) Part 121 operations. This report consolidates findings from four sets of related analyses. Collectively, the results can provide an early foundation for improving human factors integration in the FSB process and certain AED functions.

The first section of this report details the background of this research and describes the methodological approaches used to address the operational needs. The purpose of this work was to identify where there are gaps in human factors and operational data in FAA documentation specific to AED functions: Flight Standardization Board (FSB) Operational Evaluations (including T Tests), crew complement determinations through the FSB process, and operational suitability assessments. These assessments can be made based on subject matter expertise, analysis of data, and findings presented by an applicant from part-task testing or full-mission simulations.

The results of first phase of this research are described in Section 2, *Discussions with FSB Members*. This section details the findings from semi-structured interviews conducted with 10 FAA personnel on FSB activities and processes. The results identified several ways in which the processes could be improved including: increasing the standardization of FSB processes, formal documentation of FSB methods and results, and earlier involvement of AED in the FAA's aircraft certification process. As can be seen in Section 3, *Confluence of Interview Results with Published Papers*, the need for closer integration of airworthiness and operational assessments during the aircraft certification process was also noted in several published reports, including the Joint Authorities Technical Review (JATR) in their review of work conducted during the Boeing 737 MAX certification (JATR, 2019).

Section 4, *Review of the Literature*, provides a review of the scientific literature on human factors information and processes that could be useful in AED activities. In addition to more general applied human factors topics that could be used in the AED context, topics researched included representative operational pilot behaviors, current assessment methods and testing procedures, and design and use of checklists, memory items, and procedures. While the review of the literature did not reveal any publications suitable for such guidance, there is a wealth of literature that could be used, in combination with substantial subject matter expertise, to develop guidance for example to assist AED in their tasks.). Although there is a wealth of human factors information in the literature on foundational topics, the more specific the information, such as pilot response time in seconds to a given situation, the less generalizable the finding. Such information will always need to be interpreted by subject matter experts and applied within the context of the specific functions and operational environment. However, useful information and processes could be compiled into tools and guidance for AED personnel. Use of such tools would increase the standardization of AED tasks, while maintaining the flexibility that should continue to be afforded to AED to decide what needs to be assessed, how it should be assessed, and how the results should be interpreted and applied.

Section 5 identifies the research gaps derived from the interviews and literature review and provides suggested approaches to fill these gaps. The results of the research could be used to progress the development of the tools and guidance for AED. The research areas identified include operational measurements of flightcrew workload, dynamic fault management, development of a framework to identify suitable operational credits or constraints, and assessment of pilot performance for a common type rating. A repository of AED decisions, with details as to how these decisions were reached, would be useful in helping to guide future decisions and increase standardization. The report concludes with a recommendation that the feasibility of deidentifying and compiling operational data on pilot performance and issues identified by pilots after final design approval should be explored.

1. Introduction

The Aircraft Certification, Safety, and Accountability Act (ACSAA), passed by the United States (U.S.) Congress in December 2020, includes several requirements to further integrate human factors in Federal Aviation Administration (FAA) regulatory material and processes. One purpose of this work was to identify gaps in human factors and operational data in FAA documentation specific to three AED functions: Flight Standardization Board (FSB) Operational Evaluations (including T Tests), crew complement determinations through the FSB process, and operational suitability assessments. Another purpose was to obtain data from existing literary sources to fill identified gaps and provide recommendations as to how to address gaps that could not be filled.

FAA Advisory Circular (AC) 120-53B, *Guidance for Conducting and Use of Flight Standardization Board Evaluations – with Change 1*, identifies the functions and processes of the Flight Standardization Board (FSB):

> When requested by industry, the Flight Standards Service (AFS), through the Aircraft Evaluation Groups (AEG)[1], undertakes analyses of new and related aircraft and their associated systems for the purpose of providing recommendations for pilot training and qualification. These recommendations are documented in FSB reports for each specific or related aircraft and may be used by a certificate holder to develop its training and qualification programs. The FSB is responsible for specification of minimum training, checking, currency, and type rating requirements, if necessary, for U.S.-certificated civil aircraft.

As described in AC 120-53B, evaluations are conducted to determine the aircraft type rating and the pilot training and qualification requirements. In support of these decisions, T Tests, proposed by the applicant, are used to determine the level of differences between a candidate and legacy aircraft.

In the conduct of this task, the U.S. DOT Volpe Center conducted ten semi-structured interviews with FAA personnel on FSB processes. Due to the scope of this research, discussions were limited to transport category operations conducted under Part 121.The purpose of these interviews was to provide an initial exploration of the areas in which the processes and materials available to FSB could be improved and the ways in which human factors could be better integrated into FSB activities. Interview questions were provided in advance and focused on information, materials, and processes used by the FSB. Interviewees were informed that no comments would be attributed to individuals by name. The discussions ranged from 30-60 minutes and were guided by the respondent's specific area of expertise and answers to the questions below:

- In your opinion, what are the limitations and advantages of the current FSB process?
- How do you think the FSB process could be improved?
- What information do you use (e.g., data, checklists, processes, guidance, methods, tests) to support and document your evaluation?

[1] The Aircraft Evaluation Group (AEG) was the predecessor to the AED.

- What pilot performance information, if any, would you like available to you (to aid in your assessments) that has not been available in the past?
- What information, if any, would you like available to you (to aid in your assessments) that has not been available in the past?

Six interviewees had served as a Chairperson of a FSB. The remaining four interviewees were Human Factors Specialists. Human Factors Specialists were a relatively recent addition to the FSBs; they averaged five years at FAA and two years participation on FSBs. In addition, a representative from AIR with 17 years of experience in certification was consulted. While the opinions offered by the interviewees were in general agreement, it is important to note that these views are a sample and may not be representative of others with FSB experience.

The general themes from the interviews are presented here. Note that the interviews were unable to explore all topics in detail.

2. Discussions with FSB Members

Interviews with FSB members identified several areas in which human factors information would be useful as well as ways to progress the goal of better integration of human factors in the FAA's regulatory processes. These discussions identified the following key points:

- The main advantage of the current FSB process is that it is flexible; flexibility must be balanced with standardization.
- The process would benefit from better integration of FSB responsibilities into the aircraft certification process.
- The determination of operational suitability and minimum training requirements by AED should be a compulsory component of the aircraft certification processes led by AIR.
- More specific guidance is needed on the testing process.
- FSB Chairpersons would benefit from more information on human factors.
- New hires that serve on FSBs would benefit from more specific information on the FSB processes.
- Ways to provide Human Factors Specialists and FSB Chairpersons the opportunity to supplement their operational experience should be explored.
- The roles and responsibilities of Human Factors Specialists in AED need to be clearly defined.
- A repository of previous FSB decisions, with details as to what was done and why, would be helpful.
- Continued data collection on pilot performance, as well as problematic issues identified by pilots, after final approval would be useful.

One of the identified benefits of the current FSB processes is the flexibility afforded to the Chairperson. The Chairperson can make decisions, for example, regarding the level of testing needed and how the testing will be conducted. This allows the FSB to quickly scope the process to the complexity of the project and modify it as issues arise. This flexibility was identified as enabling the FSB to keep up with the technology and not be unduly burdensome to the applicant by requiring testing where none is needed. However, this flexibility was also identified as needing to be tempered with more standardized processes. While FSB procedures and results under the same Chairperson would be expected to be consistent, discussions with interviewees revealed variability across Chairpersons. Based on these discussions, it would be difficult to expect decisions to be consistent across FSBs with different Chairpersons without more standardized procedures and decision criteria. The resulting lack of standardization can also lead to erroneous assumptions made on the part of applicants based on previous experiences with different Chairpersons and/or OEMs.

Many of those interviewed stated that new hires to AED would benefit from more specific information on the FSB processes. Additionally, several respondents stated that formal training for all Aviation Safety new hires who participate in FSB activities should be improved to supplement individual operational experiences, as needed. For example, more than one FSB member suggested that new hires without specific operational experience might benefit from shadowing a Principal Operations Inspector (POI) to learn more about the relevant operational environment. This would allow new hires to observe operations, for example, in a Certificate Management Office, and/or observe airline operations or training to gain a more in-depth understanding of air carrier operations. Given the breadth and complexities of equipment, procedures, and environments that AED may encounter, such supplemental training would be a valuable means to augment their subject matter expertise. It would also be helpful if on-the-job training for new hires included the opportunity to experience the FSB processes from the earliest phases through to the final report. Such a process would step through what actions are required, including what needs to be assessed and how, identification of end products, and a list of documentation that is to be delivered to other FAA personnel, such as to POIs. Once developed, this information would increase the standardization of the process and would help to put the importance of AED component tasks in perspective.

More than one Chairperson interviewed indicated that Human Factors Specialists need to understand the operational environment in order to be maximally useful to the FSB processes. This perception can lead to an underutilization of the human factors expertise provided to their group. The application of human factors information and existing data can contribute to more objective, reliable (i.e., repeatable), and standardized decisions. To realize these benefits, all FSB Chairpersons will need to have a fundamental understanding of the ways in which human factors could be applied to their tasks and the ways in which the Human Factors Specialists available to them can assist. While all Chairpersons interviewed had at least a minimal understanding of human factors, none had a solid understanding of what human factors “brings to the table” and how to best use these Human Factors Specialists. Chairpersons and other FSB members who are not Human Factors Specialists would benefit from a tutorial that provided useful information on pilot performance and human factors methods. This would help inform the FSB members when and how the Human Factors Specialists could contribute to FSB tasks.

The Aircraft Evaluation Group (AEG)[2] *Job Task Analysis Worksheet* identifies the areas of tasks under FSB jurisdiction; Subtask 8 under “Conduct FSB Evaluation” states “Prepare a list of lessons learned from the evaluation and distribute it to interested parties such as the manufacturer/applicant, FSB members and other FSB chairman in the office, other AEGs and foreign authorities if they participated.” All of those interviewed agreed that such a repository of these lessons learned would be useful and should be constructed.

Several of those interviewed added that, in addition to documenting lessons learned, documents that describe the issues addressed in FSB activities, how they were addressed, and how the results were used in the final determination would be useful to retain in a repository. One interviewee stated that they use Issue Papers to reach resolution on process opinion differences, but these Issue Papers are not required to be published as part of the FSB report. Documentation of past decisions and how they were made could be useful references for future FSBs, particularly as AED inspectors and FSB Chairpersons with previous experience with similar projects retire. Chairpersons noted that without such documentation, the knowledge of past procedures and observations used to make assessments as well as any

[2] The AEG was the predecessor to the AED.

knowledge of the challenges and successes that resulted from the FSB decisions will likely retire with the Chairpersons involved.

Finally, several interviewees stated that it would be helpful to have access to in-service performance data and pilot reports of issues identified in actual operations. Such data could be used to verify assumptions made in previous analyses and shape future decisions.

Some FSB Chairpersons expressed a desire for a tutorial on human factors and how to integrate human factors in their assessments. In addition, most of the FSB members interviewed indicated that specific information on pilot performance in the following areas was highly desirable: capabilities and errors (e.g., perception, information processing, etc.), data on pilot accuracy and response times, and specific processes used in human factors testing and evaluation (e.g., measurement of mental workload).

Human Factors Specialists have been hired by FAA, in part, to participate in FSB activities. However, interviews revealed that the specific roles of these Specialists were not clearly defined. Information from these interviews suggest that to be the most useful, Human Factors Specialists should be involved in FSB activities and deliberations from the beginning, not brought in late for a ‘final check’. In addition, the timing of the involvement of FSB members, including the Human Factors Specialists, dictates the timing of the opportunity for them to learn the system(s) being assessed and to provide input. In any evaluation, the earlier issues are identified, the more valuable the input is to affect design improvements, procedures, and training.

Timing of AED involvement in the aircraft certification process

The predecessor to the AED was the Aircraft Evaluation Group (AEG). One responsibility of the AEG that carried over to AED was the determination of the operational suitability of the aircraft and its systems. This includes a requirement to “address broad-based safety concerns that provide a ‘bottom line’ for operational safety including all disciplines, such as airworthiness, crew procedures, flight operations, maintenance, human factors, air traffic, airports, facilities, or other aspects of a safe operation” (Imrich, 1981, p. 6). The importance of this responsibility cannot be overstated. As FAA Order 8110.4C (2007), *Type Certification with Change 6,* states, “It is possible an aircraft could be type certificated and not be determined to be operationally acceptable under the applicable regulations” (p.54). The Order further states that this situation in which a type-certificated aircraft is determined to be unsuitable for specific operations can be avoided with “proper and timely AEG involvement” (p. 54).

We have seen that there is strong consensus in the literature that AIR and AED activities need to be more closely integrated. This was noted more specifically in the interviews with FSB members as the recommendation that the determination of operational suitability and identification of the minimum training and checking should be required as part of the overall aircraft certification process. This would require that AED be involved earlier in the process than is currently the norm. While AED activities are usually conducted before certification activities are complete, full AED participation may not occur until the final phases of a certification project. As a result, The AED is not typically afforded the opportunity to fully participate in AIR-led discussions between the applicant and FAA. While it might not be practical for the FSB to be involved before AIR determines that systems are safe and perform their intended function, there are several advantages to earlier AED involvement. If the AED were involved earlier in aircraft certification processes, potential operational issues could be identified earlier; the earlier issues are identified, the easier they are to resolve. Earlier involvement would also relieve some of the time pressure often experienced to complete testing so that the aircraft can be put into service. Another benefit of bringing the AED in earlier is that it would give these specialists more time to familiarize themselves with the candidate aircraft, systems, equipment, and their intended use by a target population in a specific operating environment; they would be afforded opportunities as those provided by the manufacturer to AIR flight test pilots. The additional exposure to the relevant intricacies of the new aircraft/system will only become more valuable as the complexity of these systems increase. Ideally, these and other relevant interactions between AIR and the applicant would be participatory, not observational, and lead to discussion about potential issues that could affect operational suitability or training (such as assumptions made in the design about pilot behavior) earlier than the current process affords. Complex systems will benefit from not only early, but also continuous, involvement over the lifecycle of the equipment as the uses of the equipment evolve.

It was also discussed that, in the current process, the assessment of training needed for a candidate aircraft may be restricted to comparison to the base aircraft chosen by the applicant. Since the FSB is able, but not required, to consider the cumulative effects of training on all variations of the same aircraft type, the assessment could underestimate the complexity of the effects of a pilot's experiences on multiple variations of the same aircraft type; this could result in hidden risks. Continuous involvement throughout the lifecycle of a product would support a greater understanding of these complexities as well as the issues involved in the use of equipment or displays for other than their originally intended function or use in different operational environments. From a human factors standpoint, guidance on how to identify and assess issues associated with such complexities would be helpful toward increasing standardization.

Interviews with FSB members revealed that new applicants may not understand the purpose and sequence of aircraft certification processes performed by AIR and AED. Additionally, they said that it is often assumed by those outside and inside the FAA that AIR handles all of the human factors issues in the aircraft certification process, even though human factors issues are critical in the assessment of operational suitability and in determination of requirements for training, checking, and currency. Perhaps the most common example of this concerns the handling qualities of an aircraft. While AIR requirements address the aerodynamic performance of an aircraft, which in turn affects flight characteristics, the assessment of 'handling qualities' is in the purview of a T2 test conducted by AED.

One of the interviewees noted that, even within FAA, Order 8110.4C, *Type Certification with Change 6* is often referred to as the "Certification Order" when in fact it also addresses responsibilities of Flight Standards. Better integration of AED into Aircraft Certification processes would help clarify these issues so that FAA and applicants have a common understanding of roles and responsibilities.

3. Confluence of Interview Results with Published Papers

The current research confirmed what was noted long ago–that FAA aircraft certification processes would benefit from better integration of AED and AIR. For example, Abbott, Slotte, and Stimson (1996) identified this need in their examination of pilot interactions with automated systems. They document the results of a human factors study that focused on flightcrew difficulties in interacting with the flight deck automation. The scope of the investigation included flight deck design, flightcrew training, flightcrew qualifications, air carrier operations, and the appropriate use of automated systems. The multi-disciplinary team concluded, in part, "New tools and methods need to be developed, and existing ones improved" (FAA, 1996, p. 25) to support regulatory authorities in evaluating flight deck designs for pilot performance problems. Another conclusion was that "The FAA should improve and increase interaction between the Flight Standards and Aircraft Certification Services" (FAA, 1996, p. 78). These findings mirror the input from interviewees who expressed a desire for operational data, tools, and processes to facilitate the alignment of AED and AIR functions throughout Aircraft Certification processes.

In 2002, FAA and industry reiterated the need for more formal coordination between AIR and Flight Standards in *The Report of the FAA Associate Administrator for Regulation and Certification's Study on the Commercial Airplane Certification Process*. The report noted that:

> The AEG was created to resolve this communication deficiency and serves as the liaison organization between Aircraft Certification and Flight Standards. However, there remains a lack of adequate formal guidelines that define the interface between Aircraft Certification and Flight Standards. Without such protocols, there is an ongoing risk that important safety data and precursor information may not be communicated in a way that assures proper analyses and appropriate action. While efforts have been initiated to address this issue, it remains unresolved. (p.66).

The report also noted that:

> There are informal processes that have evolved between Flight Standards and Aircraft Certification, but they are neither consistent nor complete. The lack of documented formal business processes between these offices compromises effective communication and coordination that may affect the FAA's ability to address industry safety issues effectively and industry's ability to fully comply with FAA regulations and requirements. (p. 68).

The need for closer integration of airworthiness and operational assessments during the aircraft certification process was also noted by the Joint Authorities Technical Review (JATR) chartered by FAA. The mission of the JATR was to review:

> the work conducted during the B737 MAX certification program, to assess whether compliance was shown with the required applicable airworthiness standards related to the flight control

> system and its interfaces, and to recommend improvements to the certification process if warranted. Of particular concern to the FAA in chartering the JATR was the function, evaluation, and certification of the MCAS function on the B737 MAX3. The JATR team's review also focused on flight crew training and operational suitability of the design. The JATR team considered whether the appropriate regulations and policy were applied, as well as how applicable regulations and policy material could be improved to enhance safety. The FAA did not charter the JATR to review the entire certification process for all aspects of the aircraft, nor did it task the team to review details related to returning the B737 MAX to service. (2019, p. II)

One of the resulting recommendations of the JATR was:

> "Based on the JATR team's findings and observations related to the operational design assumptions of crew response applied during the certification process for the flight control system of the B737 MAX, JATR team members recommend that the FAA *require* [emphasis added] the integration of certification and operational functions during the certification process" (2019, p. 43).

Another key recommendation was that the FAA "should be provided all system differences between related aircraft in order to adequately evaluate operational impact, systems integration, and human performance" (2019, p. 43).

Academic research performed in support of AIR certification activities found that manufacturers would welcome better integration of aircraft certification and operational functions in the approval process. Mumaw et al. (2021) interviewed representatives from four original equipment manufacturers (OEMs) who routinely conduct 14 CFR Part 25-related certification activities. They also interviewed FAA AIR personnel from Headquarters, Atlanta, Los Angeles, Seattle, and Wichita. Mumaw and colleagues noted that, "The value of early involvement with the FAA was a recurring topic. However, this process was not seen as systematic or well-organized, even though it is encouraged on both sides" (p. 48). These views on AIR activities mirror the need identified for early AED involvement. Through the interviews with OEM applicants, Mumaw et al. (2021) also learned that applicants would like a defined path that would allow applicants to use the results from early human factors testing in the aircraft certification process. The same would apply to AED activities.

Better integration of aircraft certification and operational activities could also afford AED the opportunity to review information and test results acquired in AIR-led activities. Even if not directly relevant to the operationally-focused AED evaluations, the information could provide valuable insights to inform AED decisions. It could also facilitate the relatively new requirement for FAA to review any human factors testing conducted by the applicant, for which the applicant is seeking to use to justify training or qualification credits. Specifically, the requirement is to ensure that "the applicant has accounted for realistic assumptions regarding the time for pilot responses to non-normal conditions in designing the systems and instrumentation of such airplane." (ACSAA, 2020, Section 119, Domestic and International Pilot Training, p. 1157).

Early and continuous involvement of AED could be beneficial for the manufacturers as well as FAA. AED

Operational Suitability feedback to OEMs could result in improved designs that enhance usability before the equipment is finalized and/or beneficial changes to proposed training and procedures. For example, systems designed to display information to flightcrews have become increasingly complex. While AIR approves the design and installation of flight deck systems and equipment, it does not require an assessment of all the information that a display can convey. The addition of information or functions to a display after it is certified could increase the number of steps needed to access the information for which the display was originally intended. As one interviewee noted, an assessment of the number of actions needed to return to a normal operational configuration to manage a system failure or malfunction that requires flightcrew action would be included in an operational suitability assessment; the results could provide feedback for refinements of the design, procedures or training requirements.

4. Review of the Literature

A review of the literature was conducted for human factors information and processes that would be useful in AED processes. Specifically, a search was conducted for useful information that could be used to address the needs of AED on the following topics:

- Representative operational pilot behaviors,
  - Common types of errors,
  - Pilot response times,
  - Factors that affect pilot accuracy and response time,
- Assessment methods, appropriate for use in conduct of FAA responsibilities,
  - Methods for determining what should be tested and how:
    - Subject matter expertise,
    - Table-top assessment,
    - Part-task simulation,
    - Full-mission simulation,
  - Testing procedures with aircraft or flight simulations,
    - Pilot selection,
      - Number of pilots,
      - Background experience,
      - Other relevant characteristics,
    - Testing conditions to be included,
    - Data collection,
      - Workload assessment (both physical and cognitive),
    - Data analysis and interpretation,
- Design and use of checklists, memory items, and procedures.

Applied human factors

There are several documents that contain fundamental human factors information and procedures that could be of use to AED personnel without a human factors background. *Human Factors for Flight Deck Certification Personnel* (Cardosi & Huntley, 1993) presents a compilation of material that was designed specifically for FAA AIR flight test engineers and flight test pilots involved in assessments of flight deck systems. This document contains information on fundamental characteristics of human operators that are applicable to flight deck operations, with chapters on:

- Auditory perception,
- Color vision and other visual processes,
- Perception of form and depth,
- Information processing,
- Attention and display compatibility,
- Decision making,
- Divided attention, time sharing, workload and human error,

- Automation,
- Display design,
- Workload assessment, and
- Human factors testing and evaluation.

This 416-page document also contains useful details of pilot performance found in studies by Boeing and others. Despite the age of the document, much of the information, for example, on visual and auditory information processing, remains useful. The information on flight deck automation may be of historical interest– as well as the accurate prediction that "[cockpit] automation will continue to increase" (p. 239). Much research has been done on the topic of flight deck automation that could be used to update such a document.

The most useful reference document for human factors airworthiness regulatory and guidance material for flight deck displays and controls is *Human Factors Considerations in the Design and Evaluation of Flight Deck Displays and Controls – Version 2* (Yeh, Swider, Jo, & Donovan, 2016). This extensive, 367-page document identifies specific guidance on human factors issues to consider in the design and evaluation of avionics displays and controls for all types of aircraft (14 CFR parts 23, 25, 27, and 29). This document contains detailed sections on numerous aspects of visual displays (including Head-Up Displays), annunciations and alerts, workload, and automation, along with providing several examples of usability evaluations. This document was developed to facilitate the identification and resolution of human factors issues that are frequently reported by FAA Aircraft Certification Specialists in determining airworthiness. While it can serve as a useful reference document for AED, it does not speak to operating requirements or issues that need to be considered in an operational assessment.

Human Factors in the Design and Evaluation of Air Traffic Control Systems (Cardosi & Murphy, 1995) is a textbook and checklist booklet combination that was designed for line controllers who participate in the design and testing of new systems for air traffic control. In addition to containing some of the same basic information on human performance that would be useful in developing a compilation of human factors material for AED personnel, some of the information and a checklist format included in the document could be useful in developing a tool for assessments of operational suitability. The chapter on testing and evaluation is particularly relevant as it details the importance of including representative users performing representative tasks in a realistic operational environment as the best predictor of expected performance in simulation studies of operational suitability. In addition, it discusses the use of statistical tests with an explanation of statistical versus operational significance. Expanding on this distinction between statistical and operational significance could be useful to AED personnel without a background in experimental design or statistics in assessing the quality of tests that they review, the generalizability of results, and the validity of assumptions drawn from presented data.

AED Operational Evaluations

As stated in FAA AC 120-53B, *Guidance for Conducting and Use of Flight Standardization Board Evaluations*, all operational evaluations conducted by AED are done at the request of an applicant:

> When requested by industry, the Flight Standards Service (AFS) of the FAA, through the Aircraft Evaluation Groups (AEG), undertakes analyses of new and related aircraft and their associated systems for the purpose of providing recommendations for pilot training and qualification. These recommendations are documented in FSB reports for each specific or related aircraft, and may be used by a certificate holder to develop its training and qualification programs. (p. 2).

AED assessments of operational suitability, training requirements, testing requirements, or crew complement can employ a variety of methods. These decisions may be made based on subject matter expertise, analysis of data, and/or findings presented by an applicant, part-task testing, or full-mission simulations. Several interviewees noted that for increased standardization in FSB processes, specific guidance is needed in terms of how to determine when an evaluation is required, the type of assessment that is needed, and criteria for pass/fail/acceptable/unacceptable performance. Operational suitability assessments of newly certified aircraft are one of the primary responsibilities of the AED. The definition of operational suitability is the determination that an aircraft or system meets the applicable operational regulations (e.g., Title 14 of the Code of Federal Regulations (14 CFR) parts 91, 121, 135) and so may be used in the National Airspace System (NAS). The Air Carrier Training Aviation Rulemaking Committee (ACT ARC) recommended clarification of definitions, AED responsibilities, and parameters used in FSB operational suitability assessments (ACT ARC Recommendation 21-3, *Additions and Clarifications to the "Definitions" Section of AC 120–53B, Change 1*). FSB members interviewed also expressed the need for more details to be provided for the terms: "operational suitability" and "flight characteristics vs. handling qualities." They also expressed a need for criteria for determining what type of test is appropriate in various assessments.

However, one recommendation from the ACT ARC would not be supported either by the FSB Chairpersons interviewed or from a human factors standpoint. The ARC recommended that "the FAA should be required to establish parameters, criteria, and number of flights used for the FSB operational suitability assessment" (ACT ARC Recommendation 21-3, p. 3). While it would be helpful to identify potentially relevant issues to be examined with examples of operational definitions and parameters for measurement, the specification of methodological details such as the number of flights to be used would be counterproductive. The AED processes are designed to be agile, enabling the FAA to tailor evaluations to the complexity of the issues and assess what is needed without undue burdens of excessive testing and data collection. The type of FSB evaluation that is required is determined by the Chairperson based on the issues identified to be assessed.

Evaluations called "T Tests" (not the same as a statistical t-test) are used to determine the aircraft type ratings and pilot training and qualification requirements for new, derivative, or modified aircraft. T Tests are also used to establish the training and qualification credits that can be earned based on similarities between related aircraft. These T Test processes are described in FAA Advisory Circular (AC) 120–53B, Change 1, *Guidance for Conducting and Use of Flight Standardization Board Evaluations*. While nothing was found in the scientific literature that addressed T Tests, there are published recommendations from stakeholders. The Air Carrier Training Aviation Rulemaking Committee (ACT ARC) had specific recommendations regarding T Tests. ACT ARC Recommendation 21-5, *Flight Standardization Board T*

Tests recommended that FAA "conduct a review of current T Test processes with the goal of improving and expanding the T Test methodologies described in AC 120–53 based on new technologies in aircraft design and training and lessons learned from previous evaluations" (p. 1).

The results of interviews with FSB members also echoed the need for more specific guidance regarding the T Test process. For example, Appendix 3 of AC 120-53B states that:

> Not all changes or modifications to an aircraft or on occasion, the certification of a related aircraft may require flight-testing to assess their impact upon type rating. Type rating determination through analysis may be considered if the changes do not influence aircraft handling, introduce no significant change to systems operation or pilot procedures, and can be addressed at level A or B training (p. 2).

While it is certainly true that not all assessments are required to be conducted in an aircraft, it is unclear how this determination is made. Similarly, it is unclear how to determine whether 'aircraft handling' in the applicant aircraft is the same as in the base aircraft without either conducting an assessment in an aircraft or the requirement of specific data from the applicant.

More specifically, feedback from the interviewees included the need for operational definitions for all criteria and measures to be used in T Tests. For example, since the pass/failure based of a T-2 Test is an "assessment" of the "degree of difficulty" in performing maneuvers in the candidate aircraft compared to the base aircraft, "degree of difficulty" needs to be operationally defined as well as how this is to be assessed. Similarly, guidance needs to be added to address how decisions are made in the assessment of T-3 test results. For example, the failure of one crew could be due to a single test subject, specific aspects of the crew complement or dynamic, or improper training. When should 'repeated attempts' be allowed? When should an individual pilot's performance be considered an 'outlier' and justifiably discounted? In the past, such decisions were made solely on the basis of expert judgement and not documented. From a human factors and standardization standpoint, such decision-making processes should be codified.

Consistent documentation is a prerequisite for a growth in standardization. Documentation of the details of the procedures used, and results obtained, in evaluations would help to ensure that the processes are repeatable and comparable. Consistent documentation allows decisions to be tracked over time so that future AED activities can benefit from past assessments. For each AED evaluation, documentation should specify the process in terms of expertise involved, purpose of the evaluation, issues identified, procedures used, analyses conducted, and operational decisions reached. This documentation would aid future efforts in similar evaluations and could provide a draft roadmap to facilitate upfront planning, document progress, and facilitate better coordination between AED and AIR. Such documentation would also help manufacturers (applicants) and AIR have a clear understanding of the information needs of AED and vice versa. AED decisions based on pre-defined processes and criteria would also help to avoid apparent inconsistencies in results across FSBs. Such standardization would strengthen the FSB process by making the methods and resulting decisions less subjective and more defensible. This in turn would better serve the objectives of AC 120-53B.

4.1.1 T Test: Pilot selection

Experts in any field behave differently than novices; they process information differently and use different strategies to diagnose and solve problems. For this reason, professional flight test pilots cannot provide the best predictions of the representative pilots' behavior. In fact, despite many attempts to model pilot performance, there is currently no valid method or source of data that can be used to reliably predict pilot behavior in actual operations. This is one of the reasons why FSB testing is critical. AED is uniquely positioned to identify and assess the intricate interactions of the specific product, user group, procedures, training, and operational conditions.

When T Tests require pilots to interact with equipment, the tests are usually conducted with a relatively small number (2-3) of crews in addition to FAA pilots. The practical limitation in the number of pilots able to be used in a test means that expertise is required to interpret the results. When pilots are drawn from different air carriers, any differences in performance observed could be due to due to procedures or cultures that are air carrier specific. It is also the case that selection of FSB test subjects could skew the results (in either direction). Knowing this, pilot selection and interpretation of the results in operational evaluation requires the expertise of the AED.

Notice N 8900.649, *Use of Air Carrier Pilots During Flight Standardization Board Evaluations for Transport Airplanes*, states that "FSB test subjects are the identified FAA pilots and/or air carrier pilots that participate in the FSB activities as appropriate for the size, scope, and specific requirements" (p. 2). The FSB members interviewed identified that more specific guidance in pilot selection for FSB activities was desired. Similarly, the Air Carrier Training Aviation Rulemaking Committee (ACT ARC) Recommendation 21-6, *Flight Standardization Board Test Subjects and Membership* recommended the inclusion of line pilots in the participant pilot population.

ACSAA (2020, Section 128, Pilot Operational Evaluations, p. 1168) requires that assessments include air carrier pilots of varying levels of experience as well as foreign carrier pilots. The FSB members interviewed agreed that the inclusion of air carrier pilots in FSB tests involving air carrier operations was a beneficial change. Air carrier pilots have useful insights due to their knowledge of the airlines Standard Operating Procedures (SOPs) and safety culture and complement the expertise of FAA pilots. Air carrier pilots from a given airline within a given range of years of airline experience cannot be expected to show the same behavior or share the same opinions, for example, as to the difficulty of a given task. The pilots used in an operational assessment should be representative of the target users. While it is important to include varying levels of experience within the target population, varying the levels of experience beyond the target population would be counterproductive.

In contrast, several members interviewed noted (supported by the scientific literature) that including pilots from foreign air carriers as participants in an assessment would pose additional challenges in interpreting the results. Cultural norms, either societal or airline, could affect the test results as certain cultures are more averse to openly identifying potential faults or to admitting to difficulties in a task. The challenges of including pilots of foreign air carriers are magnified by the fact that the FAA has no control over the training or operational experience those pilots receive. Because of this, FSB Chairpersons cannot be sure of what level of performance to expect from them or how to interpret the results if they should fail. For example, pilots with English as a second language (ESL) have been found to perform differently with Quick Reference Handbooks than native English speakers, particularly if they are accustomed to using their native language on the flight deck (Sevillian, Silveria, & Graeber, 2015). Most notably, the "time to respond to system malfunctions/failures was longer due to the inability for ESL flight crew members to understand the translation of the technical words" (p. 16). In short, the inclusion of line air carrier pilots in addition to FAA pilots in FSB assessments is widely regarded as positive, but the inclusion of pilots from a variety of backgrounds and experience levels would increase the variability in observed performance.

Variability in performance increases the number of pilots needed to be included in testing to see reliable trends. For these and other reasons, it was the consensus of the FSB members interviewed, and this independent human factors assessment, that the FSB Chairperson, with input from HF Specialists and other SMEs, should lead the determination of both pilot selection and the interpretation of the results. From a human factors perspective, this is a sound approach since this team is in the best position to determine what is needed when, given the fluid and dynamic nature of FSB testing.

4.1.2 Pilot response time

Although a repository of data on pilot performance would be useful, not much could be found in the literature. This, in part, is due to the limited applicability of data obtained in simulation studies. While manufacturers sometimes collect data on pilot response times and error rates in the development and testing of new systems, these data are considered proprietary and are rarely published. Still, there are foundational data of interest. Results of a series of flight simulation studies on one aircraft indicated that with an executive system (that is, one that requires immediate action), it will take approximately two to three seconds to detect that the message is there, five to six seconds to decide what to do about it, and one to two seconds to initiate a response (Boucek, Erickson, Berson, Hanson, Leffler, & Po-Chedley, 1980; Boucek, Po-Chedley, Berson, Hanson, Leffler, & White, 1981; Boucek, White, Smith, & Kraus, 1982; see also Berson, Po-Chedley, Boucek, Hanson, Leffler, & Wasson, 1981). This led to the conclusion that a total of eight to eleven seconds should be allotted for a pilot to respond to an alert that requires immediate action.

This finding can be contrasted to the pilot response times in early simulation studies of Traffic Alert and Collision Avoidance System TCAS (Chappell, Billings, Scott, Tuttell, Olsen, & Kozon, 1989). In one study, the range of pilot response times for making an input to the controls from the onset of the Resolution Advisory (RA) was measured at one to seven seconds (*standard deviation* = 1.3 seconds). The response times measured to a later prototype in which the Traffic Advisory (TA) alerted pilots to the conflict for 40 seconds before the RA, and pilots received multiple RAs in the simulation, was less than 2 seconds (*standard deviation* = 1.4 sec; Chappell, 1990). Pilots had a longer than normal response time to reversal RAs (e.g., a CLIMB RA after a DESCEND RA). The FAA (2011) TCAS II manual states that, "In modeling aircraft response to RAs, the expectation is the pilot will begin the initial 0.25 g acceleration maneuver within five seconds" (p. 29).

While not based on referenced data, the National Transportation Safety Board (NTSB) used 35 seconds as the representative time required for an A320 flightcrew to recognize a bird strike and the extent of the engine thrust loss and decide on a course of action (NTSB/AAR-10/03)[3]. This was used as the delay in the simulations conducted to determine whether the airplane that ditched in the Hudson River could have glided to and landed at LGA or TEB after the catastrophic bird strike.

Situations in which pilots do not have the knowledge to understand what is happening and why can result in pilots responding inappropriately. Dehais, Peysakhovich, Scannella, Fongue, and Gateau (2015) examined pilot responses to a change in altitude commanded by an automated system. While the purpose of the study was to demonstrate a proof of concept for a tool that would indicate an inconsistency between the pilot's settings and the autopilot behavior, some performance data were described. This simulation study used pilots from the Institut Supérieur de l'Aéronautique et de l'Espace (a French aerospace engineering school) with a wide range of experience with automated functions (range of 15-700 hours with a mean of 200 hours). In the "impossible descent automation surprise", pilots in a simulator were instructed by ATC to increase speed and descend 5,000' with a minimum of 1000'/min. To prevent an unacceptable speed, the automated function leveled the aircraft. This was unexpected for 9 of the 16 pilots who completed the simulation. Only 7 pilots were able to correctly assess the situation and reduce the speed with a mean reaction time of 18 seconds (the standard deviation was not computed). Unfortunately, there was no attempt to examine pilot performance as a result of experience. Cues that identified a pilot input that would conflict with an automated function reduced the response time and resulted in all pilots responding appropriately. This could be useful as an example that situations that result in an 'automation surprise' take longer to resolve than cued situations.

Even with a uniform cohort of airline pilots, performance will vary with the typicality and complexity of the event. The more unusual or complex the event, the more time will be needed for the pilot to diagnose the problem and respond appropriately. Specifically, research has shown that atypical events result in more pilot errors and greater response times than a random selection of safety events in simulated flight conditions (Clewley & Nixon, 2023).

It would be worth mining, compiling, and cataloging available data and noting any limitations. In terms of Part 121 operations, there is a substantial amount of raw data and pilot safety reports contained in the FAA's Aviation Safety Information Analysis and Sharing (ASIAS) database. The ASIAS website was developed to promote the "open exchange of safety information in order to continuously improve aviation safety". However, to fully satisfy the objective of enabling users, including FAA, to perform meaningful queries, ASIAS data and safety reports would need to be de-identified and analyzed, so that only portions of them would be made publicly available, much like reports submitted to the Aviation Safety Reporting System (ASRS). The cost of such an effort might be prohibitive and the agreements in place regarding ASIAS data collection and use could not be violated. Nonetheless, strategies for enabling

[3] Loss of Thrust in Both Engines after Encountering Flock of Birds and Subsequent Ditching on the Hudson River US Airways Flight 1549 Airbus A320-214, N106US Weehawken, New Jersey, January 15, 2009

the FAA to mine relevant data to inform operational decisions should be explored. For example, depending on the quality of the data available, it may be suitable for analysis to assess the amount of operational time available on the flight deck, for a specific flightcrew procedure (Kourdali & Sherry, 2017).

The FAA is currently compiling and reviewing additional information on pilot performance data. However, any study of pilot response time must be interpreted in the context of the following caveats:

- We cannot expect performance data from one set of pilots to be directly applicable to the full pilot population.
- Performance in a simulation study must be considered 'best case', given the wide range of variables that can affect performance in the 'real world'.
- Performance is task, context, and population specific. Even when error rates are established for basic flightcrew tasks for pilots of a given experience level, the context within which the tasks are being performed will influence performance.
    - Additionally, performance will vary with training, experience, corporate culture, workload, fatigue, and a host of other factors.

4.1.3 Assessment of pilot performance for a Common Type Rating

One of the functions of AED is to determine whether qualification on one aircraft is sufficient for qualification on another, and if not, what additional training or operational limitations are needed. Common type ratings, which allow pilots to fly more than one aircraft type with the same qualifications, are a cornerstone to aviation. However, the amount of research done on this issue is not commensurate with the importance to the aviation industry or the critical nature of the issues involved. Only one useful article on this issue was found in this review of the literature. Braune (1989) offers a sound discussion of some of the pilot performance issues involved with common type ratings. Noting that there are important nuances in positive and negative transfer that have operational implications, he identifies the following four areas of commonality which would provide a basis for considering a common pilot type rating:

1. **Aircraft Handling Qualities.** Pitch and roll characteristics, speed limitations (over-speed, stall, flap/gear extension/retraction speeds, speeds, and pitch angles on approach for various conditions) controls inputs/responses.
2. **Crew Procedures**. Emergency procedures, particularly memory items requiring immediate crew response, including a comparison of emergency procedures checklists and memory items.
   Note: For a common type rating, abnormal procedures should be identical or similar but not confusable. Differences exist where dictated by aircraft design and are acceptable since they are neither committed to memory nor are they routine actions for normal operation. Normal procedures (routine actions learned through training/operation) are identical/similar to a higher degree than abnormal procedures to ensure ease of management/control for rote operation.

3. **Flight Deck Geometry—Arrangement.** Arrangement of primary and secondary flight and power plant controls and major panel areas with respect to the crew in both aircraft.
4. **Controls/Displays.** Design and operation of subsystem controls and displays under normal/abnormal/emergency conditions. Identical location of controls and displays within each major panel.

At the time, Braune (1989) considered this a comprehensive list of areas for which commonality would suffice for a common type rating. However, since the publication of this paper, the importance of considering more than displays and controls has come to light. The critical importance of automated functions and other factors will also need to be considered in any assessment.

4.1.4 Need for a repository of FAA decisions

The Air Carrier Training Aviation Rulemaking Committee (ACT ARC) recommended compiling FAA certification decisions. Recommendation 20-5, *Standard Issue Papers*, stated that an increase in standardization in format, content, and application of FAA Issue Papers would benefit both applicants and the AED.

> The FAA uses Issue Papers (IP) to document the negotiation and resolution of certification and other issues. IPs provide a structured means for describing and tracking the resolution of significant technical, regulatory, and administrative issues that occur during a project. The IP process establishes a formal communication vehicle for addressing significant issues between the applicant and the FAA, and they are also useful in addressing novel or controversial technical issues. Aircraft Evaluation Groups (AEG) representing the interests of the FAA's Flight Standards Service (AFX) have used IPs to address Flight Standards considerations during type certification. These IPs have encompassed the evaluation of operational and maintenance aspects of certification and the continuing airworthiness requirements of newly certificated or modified products and parts*. Although the use of IPs by the AEGs is described in FAA Order 8110–112A, no guidance or policy within AFX describes the use and application of these IPs by the AEGs, whether as part of a certification project or as a separate AEG evaluation* [emphasis added] (p. 1).

The ACT ARC made several notable recommendations including:

- Reviewing and standardizing the format and content of all Operational Issue Papers (IPs) across. AEGs,
- Developing a process for development, recording, tracking, and dissemination issue papers to OEMs (both specific and generic), and
- Reviewing past AEG activities and IPs for the purpose of developing guidance as to both the possible topics of "specific" IPs (e.g., special use, unique systems and their use, and specific evaluation process) and the required rationale necessary for specific assessments.

This confluence of recommendations from the ACT ARC and the FSB members interviewed who recommended compiling FSB detailed decisions gives strong support for the operational research needed to develop a repository of FAA decisions on issues relevant to aircraft certification, operational suitability, and the like. Since there is no requirement to document why or how AED decisions were made, nor to document decisions (e.g., that led to an assessment of operational suitability) in the FSB report, research would be needed to recover as much information as possible and put it in a format that could be useful in future AED tasks.

A repository of AED decisions could also help reduce variability in how these decisions are implemented. Order 8900.1 states that the FSB's primary responsibilities are to determine the pilot requirement for type ratings, to develop minimum training recommendations, and to ensure initial flightcrew member competency in accordance with AC 120-53. However, FSB recommendations to the Principal Operation's Inspectors (POI's) role in approving training for operations are not based on a regulatory requirement and the AC states that FSB recommendations provide a means, "but not the only means" of compliance. As one respondent stated that, while there is nothing inherently wrong with this approach, it creates inconsistencies in results. It also means that the AED cannot ensure that any alternative methodology chosen by the applicant complies with the original FSB intent. Additional sources of variability are introduced by the fact that criteria for what is acceptable will vary from POI to POI, as will the resources available to POI (e.g., typically large carriers will have more resources than small carriers), and the individual safety culture and compliance attitude of the carriers. Having decisions documented that could serve as reference points could increase standardization.

4.1.5 Is a more holistic approach to operational suitability needed?

In ACT ARC Recommendation 20-10, *Operational Evaluation of Non-Installed Equipment*, the ACT ARC recommended that the FAA consider creating and documenting as policy an operational suitability evaluation process applicable to non-installed equipment, i.e., equipment that is not part of the aircraft type design but that is carried on-board by the aircraft operator, used in operations, and which could impact the operational safety of the aircraft. They stated that:

> The operational evaluation and authorization for use of EFBs [Electronic Flight Bags] and other non-installed equipment typically requires a wide range of expertise (e.g., safety, human factors, security, pilot procedures, and training) which is not available to all operators or Federal Aviation Administration (FAA) local offices. For this reason, original equipment manufacturers (OEMs) have been applying for operational suitability evaluation of non-installed equipment, in particular EFB applications, to their FAA Aircraft Evaluation Groups (AEG) in order to obtain a general statement of operational suitability, which eases and streamlines operational authorization and entry into service at the operator level. However, operational suitability evaluation of noninstalled equipment by the AEGs is not a documented process and there is no assurance to industry that AEGs can perform the evaluation on a continuing basis. (p. 1).
>
> The FAA has, in the past, published the results and findings of these evaluations on the Flight Standards Information Management System (FSIMS) as Operational Suitability Reports (OSRs). For recent evaluations, the FAA has not published OSRs on FSIMS, but has produced Operational Suitability Letters (OSLs), which it has sent directly to applicants, who distribute them to their

> end-users. The process and associated procedures to conduct the operational evaluation of non-installed equipment and the publication of results in OSRs or OSLs is not documented, and there is no assurance to industry that the FAA will continue to fulfill requests by applicants and operators. Non-installed equipment, in particular EFBs, are widely used by operators and failure to properly evaluate and authorize some functions can have an impact on operational safety. Therefore, as a short-term objective, the ACT ARC recommends the FAA consider developing and documenting as policy an operational suitability evaluation process applicable to such noninstalled equipment. (p. 2).

There is no requirement for AIR to certify displays or tools that are not installed on the flight deck. Nor is there a requirement for FAA to evaluate all information that may be displayed on a certified display device. This means that the only assessment of the interplay of the certified and supplemental information/displays would be made by the AED in their operational suitability assessment. Specifically, FAA 14 CFR Part 25 was intended to address the design and approval of *installed* equipment intended for flightcrew members to use from their normally seated positions on the flightdeck. AC 25.1302-1, however, states that it:

> applies to flightcrew interfaces and system behavior for all installed systems *and equipment used by the flightcre*w [emphasis added] on the flightdeck while operating the airplane in both normal and non-normal conditions. It applies to those airplane and equipment design considerations within the scope of part 25 for type certificate (TC) and supplemental type certificate (STC) projects (p 5).

Since the relevant CFRs and ACs were written however, there has been a proliferation of possible equipment used by pilots that is not installed. Because AIR does not certify equipment that is not physically installed on an aircraft, it is our interpretation of the existing documentation that the AED must be prepared to assess the use of *all equipment* expected to be used by the flightcrew (and their potential interactions) in their assessment of operational suitability and approval for use. This adds a level of complexity (as well as scope) to the FSB assessments. FAA should determine whether a more holistic approach to operational suitability is needed to include *all* equipment and information used by the flightcrew.

5. Research Gaps and Suggested Approaches

Several research gaps were identified from the interviews with FSB members. The subsequent review of the literature identified some relevant research, but none to effectively fill these gaps. The results of all the research areas identified could be used to inform the development of standardized guidance and tools for AED assessments.

Standardized Guidance and Tools for T tests

The results of the interviews with FSB members and literature review revealed research gaps in several components of the T Test process. ACT ARC Recommendation 21-5, *Flight Standardization Board T tests* recommended the FAA provide clear guidance on the use of evaluation devices and assessment that may be used for conducting T Tests as a part of the operational evaluation process. The FSB members interviewed identified a need for increased standardization in FSB processes and specific guidance for: pilot selection, determining when an evaluation is required, the type of assessment that is needed, criteria for acceptable performance, and documentation of the results.

There is no single "optimum" safety assessment method or feature for aircraft certification (Mumaw, Billman, & Feary, 2021). The appropriate method will always depend on the specifics of what needs to be tested. Flexibility in applying appropriate methods to show compliance with FAA regulations will continue to be necessary and beneficial to all involved parties. This means that the FSB will need to continue to be agile as well in their ability to decide what needs to be assessed and how best to assess the critical issues.

Suggested approach. Research to fill this gap would begin by documenting tasks and subtasks of FSB processes – these would be identified by FSB subject matter experts. These tasks and subtasks could be mapped to identified operational human factors requirements and measurable tasks to evaluate said human factors requirements.

Similar research has been conducted in the area of flight deck airworthiness certification that maps FAA design requirements to tasks that can be observed and measured in certification tests. Mumaw, Haworth, Billman, and Feary (2020) examined the pilot performance issues that should be evaluated for a Flight Deck Interface. This document contains sections on physical ergonomics, design for usability, data integration and display content, attention and task management, flightcrew problem solving, and flightcrew teaming.

For each performance issue discussed, Mumaw et al. (2020) identify the relevant 14 CFR Part 25 rules and guidance, thus providing a detailed mapping of FAA requirements to design features that could be used to satisfy the requirements and issues that need to be considered in an assessment. An example from the text states:

> To comply with the § 25.1302(d) requirement that a design enables the flightcrew to "manage errors," to the extent practicable, the installed equipment design should meet the following criteria:
>
> (a) Enable the flightcrew to detect, and/or recover from errors;
> (b) Ensure effects of flightcrew errors on the airplane functions or capabilities are evident to the flightcrew and continued safe flight and landing is possible;
> (c) Discourage flightcrew errors by using switch guards, interlocks, confirmation actions, or similar means, and
> (d) Preclude the effects of errors through system logic and/or redundant, robust, or fault tolerant system design. (p.9).

While the focus of this research was aircraft certification, it could be used as a model for identifying specific issues to be considered, and tasks to be included, in AED testing. For example, while Mumaw et al. (2020) operationalize the requirement for flightcrew intervention capabilities per § 25.1302 as "The interface design should aid pilots in locating and accessing guidance for pilot actions when tasks are triggered by conditions in the system or in the world" (p. 42), the issues of assessing "that the pilot can: know there is a task to be performed, identify the set of actions to take in response, and access those actions" (p. 42) is equally applicable to an operational suitability assessment.

Guidance could be compiled from the scientific literature along with input from stakeholders to help determine when subject matter expertise is sufficient to make decisions as to what is operationally suitable and when an assessment is required. If an assessment is needed, guidelines could identify issues that should be considered in determining the type of assessment that is needed, criteria for pilot selection, and criteria for acceptable performance. Several FSB members interviewed stated that it would be useful to have a tool or guidance that maps identified flight characteristics and/or handling qualities to appropriate T testing methods and then maps the results to minimum training requirements. Of course, it would not be possible to map all possible differences in flight characteristics and or handling qualities to required testing, nor is it possible to predict all the variations that the AED might be presented with in the future. However, the recommendation for a general framework of mapping assessment results to minimum training requirements is consistent with the recommendation for more detailed guidance in testing and evaluation. While such guidance would still require expertise to apply the framework and interpret the results, it could help increase standardization.

One of the FSB Chairpersons interviewed noted that while more specification of testing procedures and pass/fail criteria are needed, overspecification (such as a "redline" performance measure) would not be appropriate. From a human factors standpoint, flexibility for expert judgment on complex issues will always be needed, in part, because of the limited objective data available on complex pilot performance issues and the limited applicability of simulation data.

The goal of the development of this specific guidance and tools would be to progress the stated intent of AC 120-53B, *Guidance for Conducting and Use of Flight Standardization Board Evaluations* in their tasks of determining the pilot requirements for type ratings, developing minimum training recommendations, and ensuring operational suitability of aircraft systems. These tools would help to accomplish this by providing concrete translation of requirements into testable conditions, identifying useful performance measures and metrics, and providing recommendations for data analysis and interpretation. These tools should be developed as 'living documents' that are updated as the results of applicable research on pilot performance and its measurement progress. Use of these tools would increase the standardization of methods used to assess applicant programs across FSBs.

Assessment of applicants' assumptions of pilot response times and error rates

The scientific literature was reviewed for any data on pilot response times and errors that could potentially be useful to AED in their assessment of applicants' assumptions. The information found has been described and is insufficient to be directly applicable to AED activities. It is also important to consider the potential for *misapplication* of such data. In a review of the misapplication of response times for human perception, Francis, Tyrrell, and Owens, (2020) conclude that (even) "driving behavior is both complex and variable, and no single PRT (perception response time) value, adjusted or not, will ever be universally applicable" (p. 1). More to the point, in their interviews with OEMs, Mumaw, Billman, and Feary (2021) found that "None of the applicants required or found valuable the use of human error probabilities in safety assessments during aircraft certification" (p. 49). Throughout their review of FAA and applicant AIR testing processes, Mumaw et al. (2021) found no required or recommended approval method to determine a human error probability or likelihood for any pilot actions. In fact, they found "no one who desires to assign human error probabilities during the certification process; those interviewed feel it should remain a process driven by performance data and subjective judgment. Indeed, 14 CFR 25.1309 states that it is inappropriate to assign Human Error Probabilities" (p. 9). It would be equally inappropriate to seek to assign a human error probability in an operational assessment; such assessments will always require subject matter expertise to conduct and apply.

Aircraft systems are assigned an estimated failure likelihood based on known characteristics of components and subsequent expert analysis. Any attempt to similarly assign estimated error rates for flightcrew performance would be misguided for several reasons. While mechanical and electrical components would be expected to perform the same under the same conditions, humans would not. Components can only perform in ways in which they are programed, pilot performance is much more variable and complex. There is no prototypical pilot that can be used as the definitive test pilot, nor can the behavior from a subset of pilots be generalized to others with potentially different skills and training.

Suggested approach. To support the assessment of applicants' assumptions of pilot response times and error rates, the development of repository of pilot performance should be considered. One contributor to a repository could be de-identified ASIAS data and safety reports, so that portions of them could be made publicly available, much like reports submitted to the Aviation Safety Reporting System (ASRS). The feasibility of the strategic mining of such data should be investigated.

Operational measures of flightcrew workload for AED assessment

As previously mentioned, one of the information needs identified in the interviews with FSB members was measures of flightcrew workload, particularly mental workload, that could be used in operational suitability evaluations. While AIR personnel are responsible for measuring workload and determining minimum flightcrew through aircraft certification processes, Flight Standards personnel are responsible for assessing workload and determining crew complement in specific operational conditions through the FSB process. In all cases, the crew complement determined by FSB must be equal to, or more conservative than, the AIR determination.

AIR uses airworthiness requirements and guidance, such as 14 CFR 25.1523 and AC 25.1523-1 to examine human factors topics such as flight deck design and pilot workload. The stated purpose of AC 25.1523-1, Minimum Flightcrew is to set forth "a method of compliance with the requirements of § 25.1523 of the CFR, which contains the certification requirements for minimum flightcrew on transport category airplanes" (p. 1). Minimum flightcrew is determined, in part, by 'workload on individual crewmembers' (p. 1). The AC further states that "Appendix D of Part 25 contains the criteria for determining the minimum flightcrew under § 25.1523" (p. 4). While this AC contains useful information for airworthiness considerations, it does not provide sufficient level of detail in the description of the methods, data collection, and data analysis – nor does it provide the decision criteria – needed to ensure that FAA decisions are reliable (i.e., repeatable) or that the FAA assessments include valid measures (i.e., measure what they purport to measure).

AED considers workload in their determination of crew complement based on operational requirements (91, 121, 135, etc.) and planned operational capabilities (e.g., extended operations). Example human factors topics include pilot training and qualification to operate a specific aircraft in the NAS, operational pilot workload, and conditions related to operational approvals (e.g., international flights). AED also considers workload in their decisions on training, checking and currency requirements, and review of operational procedures. Because the assessment of workload conducted by AED is in the operational context, and what is operationally suitable will vary with the operational environment, the AED assessment of crew workload is more complex than that of AIR.

The science of measuring pilot workload has a long and rich history (e.g., Corwin, Sandry-Garza, Biferno, Boucek & Logan, 1989; Roscoe, 1987; Roscoe & Ellis, 1990; Roscoe, 1993). Until recently, the most reliable and valid assessments were subjective. The emerging technologies in wearable devices for human performance monitoring have demonstrated to be useful in assessing the cognitive workload on the flight deck and could be used to assess differences between aircraft (Izzetoglu & Richards, 2019). New measures of assessing flightcrew mental workload combined with tools to apply human factors information and processes in AED assessments would be beneficial to FAA and applicants.

Suggested approach. A survey of physiological measures useful in assessing mental workload can be found in Ayers, Lee, Paas and Merrienboer (2021). The available research on existing physiological workload metrics could be used to further test the validity of wearable devices in measuring workload during flight, potentially opening the door to a new and non-intrusive method of real-time, objective workload monitoring. However, additional research will need to be conducted to validate and demonstrate the reliability of wearable devices in measuring physiological workload that was found in Izzetoglu and Richards (2019). Contingent on the results of this research, a protocol could be developed to assess physiological and cognitive workload in operational contexts on the flight deck, which in turn could inform AED decisions.

Repository of FAA decisions

FSB members interviewed recommended compiling FSB detailed decisions. This gives strong support for the operational research needed to develop a repository of FAA decisions on issues relevant to aircraft certification and operational suitability. Research would be needed to identify any recoverable decision criteria from past assessments. The feasibility of this research has not yet been determined, but without a requirement to document how AED decisions were made, this information may not be recoverable. Guidance could be developed for the documentation of AED decisions so that a repository could be available to help guide future decisions. This could reduce variability in how these decisions are implemented in the future.

Suggested approach. The feasibility of documenting past procedures, decisions, and lessons learned could be determined through discussions with AED personnel. Guidance should be developed for the documentation of future decisions that specifies the methods and decision criteria used.

Dynamic fault management

As early as 1982, it was identified that most aviation accidents were caused by multiple failures as opposed to a single failure. In a study that was conducted jointly by the three major U.S. manufacturers of commercial transport aircraft at that time (Boeing, Lockheed, & McDonnel Douglas), Hanson, Howison, Chikos. and Berson (1982) examined the reports of accidents and incidents in NTSB database and conducted simulation studies. They concluded that most accidents “were not the results of a single failure, deviation, mistake, or misjudgment; they were the result of accumulating or compound problems. They were shown to be the result of combinations of causes, primary and contributing... Many of the crew errors were the result of improper responses to situations precipitated by other causal factors. In such cases, instead of the crew providing a recovery from the situation, they often aggravated it” (p. 12).

In a 2013 report, interviews with operators revealed an overall concern with the management of malfunctions. Specifically, they were concerned about the: decision-making, judgment, knowledge, and skills surrounding the management of equipment failures, whether failures of the automated systems or failures of other systems” (FAA, 2013, p. 35). Further, operators noted that “some training programs and non-normal procedures did not adequately address partial failures and uncertain situations” (FAA, 2013, p. 35).

Additionally, it was noted that failures for which there are no crew procedures or checklists may be becoming more prevalent (FAA, 2013). This makes training to assess the situation, (including suspecting sensor failures) and to intervene appropriately all the more critical. The same report noted, however, that current training was insufficient to provide the flightcrews with the knowledge, skills, and judgment to successfully manage systems used for flight path management.

Recently, after a review of safety assessment practices used in certification for transport airplanes, Mumaw, Billman, and Feary (2021) noted that, “there seems to be an assumption of independence (specific to pilot actions); that is, the occurrence of a failure does not change the likelihood of another failure. This is especially relevant to pilot performance because it implies that system failures do not influence pilot performance”. This is unlikely since stress can adversely affect performance. Similarly, the authors noted that multiple failures, or failures that occur at a fast pace can be more difficult for flightcrews to manage than a single more serious failure. There can be more subtle forms of failure that may be more likely to lead to undesirable flightcrew performance, e.g., competing operational objectives. Are there subtle failures that the crew might not notice that could cascade into a more serious failure if unnoticed? The effects of system failures on pilot performance, and their effects on pilot interpretation of subsequent events merits further investigation – particularly in the context of complex automated functions.

There are design requirements that specify the need for indicators as to what pilot actions need to be taken to recover from the failure of a critical system (AC 25.1309). Singer and Dekker (2000) examined the interaction between the ways in which warning systems function and the flightcrew’s management of recovery from the failure. They found that a system that guides the pilot in what to do next or that shows the pilot what is still operational has clear performance benefits. Au (2005) showed that, in the absence of indicator lights that annunciated the failure condition and pointed the pilot to the correct response, pilots tended to fixate on the most prominent cue and perform the checklist appropriate to that cue (whether or not it was the correct checklist). Au noted that the single most prominent cue does not always lead the pilot to the correct checklist. Even after identifying the situation correctly, pilots erred in their recall of the memory items.

It is usually sufficient for simple systems to rely on individual warnings for single functions in a system with status shown on the display (such as an illuminated icon of a gas pump as an alert for low fuel in a car in addition to the gas gauge). However, with complex aircraft, dynamic fault management is a challenging task: warning systems and associated flight deck procedures designed to address critical failures need to evolve along with the growing complexity of flight deck systems and operations.

Suggested Approach. An experimental approach to study dynamic fault management in specific situations could be developed from the methods used by Sarter and Woods (1994). By using a simulated flight scenario with a variety of probes, they explored the pilots’ mental model of the Flight Management System (FMS). The study demonstrates how a simulation can be used to explore pilots’ mental models of the FMS. This study revealed “a variety of latent problems in pilot-FMS interaction that can affect pilot performance in nonnormal time-critical situations” (p.2). For example, 14 pilots (70%) showed deficiencies in performing tasks such as aborting a takeoff at 40 kts with autothrottles on, anticipating certain mode indications, describing when go-around mode becomes armed in landing, and describing the consequences of loss of certain functionality. The same methodology used in this study could be used to further our understanding of dynamic fault management. In turn, these data could provide information to AED to better inform decisions.

Development of a framework to identify suitable operational credits or constraints and assessment of pilot performance for a Common Type Rating

One of the stated intents of AC 120-53B is to provide “a recommended framework for application of suitable credits or constraints to better address new technology and future safety enhancements.” An implementable framework, however, has yet to be developed. One of the recommendations from the interviews with FSB members describes a prerequisite to developing such a framework —operational definitions for “exceptional pilot skill” and “normal airmanship” need to be established. Currently, these terms are defined by the applicant, but not operationally defined with specific measures. Throughout the 14 CFR design and certification requirements for aircraft, the statement is made that the aircraft should be designed so that it does not require exceptional piloting skill. However, no definition or quantitative measures are provided for exceptional skill, hence, it is subjectively evaluated, both in airworthiness certification and operational evaluations.

Similarly, Item 11a in the FAA AC 25.1309-1A, *System Design and Analysis*, states that:

> When assessing the ability of the flightcrew to cope with a failure condition, the warning information and the complexity of the required action should be considered (reference Paragraph 8g(5)). If the evaluation indicates that a potential failure condition can be alleviated or overcome during the time available without jeopardizing other safety-related flightcrew tasks and *without requiring exceptional pilot skill* [emphasis added] or strength, credit may be taken for correct and appropriate corrective action, for both qualitative and quantitative assessments. Unless flightcrew actions are accepted as normal airmanship, they should be described in the FAA-approved AFM [Aircraft Flight Manual] or AFM revision or supplement (p. 15).

While there is a requirement for FAA to determine that “exceptional pilot skill” is not required to resolve a failure condition, no operational definitions for “exceptional pilot skill” or “normal airmanship” could be found in the literature review. An operational definition of *normal airmanship* was offered by Mumaw, Billman, and Feary (2021) but this parameter was not based on data, nor could it be directly applied in AED assessments.

The issues associated with operational credits would be applicable to the issues that need to be considered in the development of criteria for a common type rating. A landmark paper (Braune, 1989) identified important nuances associated with a common type rating with operational implications that need to be further explored. Noting that research is needed to fully understand the effects of mixed fleet operations on crew performance, Braune identified the need to investigate: 1) perceived vs actual differences in positive and negative transfer, 2) the incremental effects of differences, and 3) the impact of mixed fleet operations on crew performance). At the time, this was considered a comprehensive list of areas for which commonality would suffice for a common type rating. However, since the publication of this paper, the importance of considering much more than displays and controls has come to light. The critical importance of automated functions and other factors will need to be considered in any assessment.

Suggested approach. Extensive research would be needed to explore all these factors that contribute to identifying criteria for a common type rating. This should be considered in a future FAA research plan. The foundation of this research would be the results of research in the areas of pilot response times and errors, operational measures of flightcrew workload assessment, and the framework for identifying suitable operational credits.

A plan for data collection on pilot performance/issues after final design approval

Continued data collection on pilot performance and issues after final approval would be useful for several reasons. Such data could be used to determine if the assumptions made by the manufacturer and FSB were accurate, to determine if training was effective and sufficient, to identify latent problems early before they result in an adverse event, and to further research on the time and experience required for pilots to feel proficient with different systems. The design, as submitted by the applicant, can contain assumptions used by the manufacturer of expected pilot behavior (e.g., time required for a pilot to recognize a malfunction or determine the proper procedure for a given failure mode). These assumptions, which may also be used by AIR and FSB, are not empirically validated. At the time of this writing, while the applicant is not required to provide data to support these assumptions, AED could request that such data be reported.

Suggested approach. Investigate the feasibility for deidentifying and compiling data on pilot performance and issues identified by pilots after final design approval.

6. Conclusion and Next Steps

There are several competing factors in any AED assessment. The assessment must, with limited FAA resources, be thorough in its examination of safety critical issues while at the same time not placing an unfair economic burden on the applicant by requiring excessive testing. This complexity speaks to the need to support the current flexibility in AED processes. However, the need to balance flexibility with standardization is clear. FAA decisions must be defensible and repeatable. This means that the assessments must be as objective as possible, and the results must be as detailed as possible, well-documented, and preserved for future FAA activities – all while allowing the applicant to preserve the economic benefit of proprietary technology.

In this review, we have provided useful quantitative findings where they exist. We have also seen missteps in trying to quantify pilot skill. In terms of next steps, we have identified topics that merit further investigation; an example of this is in the area of dynamic fault management. In addition, tools could be developed using existing knowledge to increase the standardization with which said knowledge is applied. Because of the complicated nature of the assessment, it would be useful to provide both background information and a template for a test checklist to the subject matter experts on the evaluation team. However, the template must have enough flexibility to allow subject matter experts to use their judgement where there are no absolutes. Developing and workshopping a template that balances guidance with flexibility would be an important next step to increasing the standardization of assessments. Further, it would be useful to compile the best of this information into a set of tools that could be used to support AED evaluations.

Finally, we also identified research that should be conducted in the continuous improvement of information available to support AED decisions. Expanding this line of research will ensure that AED has the most comprehensive and relevant information possible to make their decisions.

7. References

Aircraft Certification, Safety, and Accountability Act, Division V of P.L. No. 116-260 (2020). https://www.congress.gov/116/plaws/publ260/PLAW-116publ260.pdf.

Air Carrier Training Aviation Rulemaking Committee (ACT ARC) (2021). *ACT ARC Recommendations*. Recommendation 21-3 Additions and Clarifications to the "Definitions" section of AC 120–53B, Change 1. https://www.faa.gov/sites/faa.gov/files/about/office_org/headquarters_offices/avs/ACT_ARC_Rec_21-3.pdf

Air Carrier Training Aviation Rulemaking Committee (ACT ARC) (2021). *ACT ARC Recommendations*. Recommendation 21-5 Flight Standardization Board T-Tests. https://www.faa.gov/sites/faa.gov/files/about/office_org/headquarters_offices/avs/ACT_ARC_Rec_21-5.pdf

Air Carrier Training Aviation Rulemaking Committee (ACT ARC) (2021). *ACT ARC Recommendations*. Recommendation 21-6 Flight Standardization Board Test Subjects and Membership. https://www.faa.gov/sites/faa.gov/files/about/office_org/headquarters_offices/avs/ACT_ARC_Rec_21-6.pdf

Air Carrier Training Aviation Rulemaking Committee (ACT ARC) (2020). *ACT ARC Recommendations*. Recommendation 20-5 Standard Issue Papers. https://www.faa.gov/sites/faa.gov/files/about/office_org/headquarters_offices/avs/ACT_ARC_Rec_20-5.pdf

Au, H. (2005). Line Pilot Performance of Memory Items. *2005 International Symposium on Aviation Psychology*, 29-35. https://corescholar.libraries.wright.edu/isap_2005/140

Berson, B.L, Po-Chedley, D.A., Boucek, G.P., Hanson, D.C., Leffler, M.F., & Wasson, R.L. (1981). *Aircraft Alerting Systems Standardization Study Volume II Aircraft Alerting Systems Design Guidelines* (Report No. FAA/RD-81/38/1). Boeing Commercial Airplane Co., Seattle, WA.; Federal Aviation Administration, Washington, DC. Aircraft Alerting Systems Standardization Study. Volume II. Aircraft Alerting System Design Guidelines. | National Technical Reports Library - NTIS

Boucek, G. P., Erickson, J. B., Berson, B. L., Hanson, D. C, Leffler, M. F., & Po-Chedley, D. A, (1980) *Aircraft Alerting Systems Standardization Study Phase I. Final Report.* (Report No. FAA/RD-80-68).

Boucek, G. P., Po-Chedley, D. A., Berson, B. L., Hanson, D. C, Leffler, M. F., & White, R. W. (1981) *Aircraft Alerting Systems Standardization Study, Volume I: System Validation and Time-Critical Display Evaluation* (Report No. FAA-RD-81-381).

Boucek, G., White, R., Smith, W., & Kraus, J. (1982) *Traffic alert and collision avoidance system*. Report No. DOT/FAA/RD-82/49.

Braune, R. (1989). The Common/Same Type Rating: Human Factors and Other Issues. *SAE Transactions*, *98*, 1455-1459. http://www.jstor.org/stable/44471740

Cardosi, K. and Huntley, M.S. (1993). *Human Factors for Flight Deck Certification Personnel.* (Report no. DOT/FAA/RD-93/5). https://rosap.ntl.bts.gov/view/dot/8679

Cardosi, K. M., & Murphy, E. (1995). *Human factors in the design and evaluation of air traffic control systems* (Report no. DOT-VNTSC-FAA-95-3). United States. https://rosap.ntl.bts.gov/view/dot/8708

Clewley, R. & Nixon, J. (2023). A new facet of category theory: cognitive disadvantage and its implications for safety in the cockpit. *Ergonomics,* 66 (6): 762-771.

Chappell, S.L. (1990) Human factors research for TCAS. *Proceedings of the Royal Aeronautical Society Conference on the Traffic Alert and Collision Avoidance System (TCAS) -Its Development, Testing, Installation and Operational Use.*

Chappell, S. L., Billings, C. E., Scott, B. C., Tuttell, R. J., Olsen, M. C., & Kozon, T. E. (1989). *Pilots' use of a traffic alert and collision-avoidance system (TCAS 2) in simulated air carrier operations. Volume 1: Methodology, summary and conclusions* (Report no. NASA-TM-100094-VOL-1). NASA Ames Research Center. Wayback Machine (archive.org)

Corwin, W. H., Sandry-Garza, D. L., Biferno, M. H., Boucek Jr, G. P., Logan, A. L., Jonsson, J. E., & Metalis, S. A. (1989). *Assessment of crew workload measurement methods, techniques and procedures. Volume 1 – Process, methods and results* (Report No. WRDC-TR-89-7006).

Federal Aviation Administration (2011) *Introduction to TCAS II version 7.1.* https://www.faa.gov/documentlibrary/media/advisory_circular/tcas%20ii%20v7.1%20intro%20booklet.pdf

Federal Aviation Administration (2013). *Advisory Circular (AC) 120-53B - Guidance for Conducting and Use of Flight Standardization Board Evaluations - With Change 1*

Federal Aviation Administration (2022). *Notice N 8900.649 Use of Air Carrier Pilots During Flight Standardization Board Evaluations for Transport Airplanes.*

Federal Aviation Administration (2023). *Flight Standardization Board (FSB). https://www.faa.gov/aircraft/air_cert/airworthiness_certification/fsb#:~:text=The%20FAA%20typically%20establishes%20a,for%20the%20aircraft%20in%20question.*

Federal Aviation Administration (2007). *Order 8900.1 Vol 8, Ch 2, Sec 5, 8-108 - Flight Standards Information Management System (FSIMS).*

Federal Aviation Administration (2013). *Final Report of the Performance-based Operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group*.

Federal Aviation Administration (1996). *Human Factors Team: The Interfaces Between Flightcrews and Modern Flight Deck Systems.* FINALVER.DOC (faa.gov)

Federal Aviation Administration (1988). *Advisory Circular (AC) 25.1309-1A - System Design and Analysis*.

Federal Aviation Administration (1993). *Advisory Circular (AC) 25.1523-1 - Minimum Flightcrew*.

Francis, E. L., Tyrrell, R. A., & Owens, D. A. (2020). Perception response time and its misapplication: An historical and forensic perspective. *Theoretical Issues in Ergonomics Science, 21*(3), 327–346. https://doi.org/10.1080/1463922X.2020.1736204

Hanson, D. C., Howison, W. W., Chikos, S. F., & Berson, B. L. (1982). Aircraft Alerting Systems Standardization Study. Phase IV. Accident Implications on Systems Design. Boeing Commercial Airplane Co Seattle, WA.

Izzetoglu, K., & Richards, D. (2019). Human performance assessment: evaluation of wearable sensors for monitoring brain activity. In M.A. Vidulich & P.S. Tsang (Eds.), *Improving Aviation Performance through Applying Engineering Psychology* (pp. 163-180). CRC Press.

Joint Authorities Technical Review (JATR) (2019). Boeing 737 Max Flight Control Systems Observations, Findings, and Recommendations. https://www.faa.gov/sites/faa.gov/files/2022-08/Final_JATR_Submittal_to_FAA_Oct_2019.pdf

Kourdali, H.K., & Sherry, L. (2017). Available Operational Time Window: A Method for Evaluating and Monitoring Airline Procedures. *Journal of Cognitive Engineering and Decision Making, 11*, 371-381. DOI: 10.1177/1555343417727190.

Mumaw, R., Billman, D., & Feary, M. (2021). *Safety Assessment Methods for Flightcrew Performance Phase 1 Final Report* (Submitted to FAA May 3, 2021).

Mumaw, R., Billman, D., & Feary, M. (2021b). *Safety Assessment Methods for Flightcrew Performance Task 2 Report*. (Submitted to FAA May 31, 2021).

Mumaw, R. J., Haworth, L. A., Billman, D., & Feary, M. S. (2020). *Evaluation Issues for a Flight Deck Interface; CAST SE-210 Output 2: Report 4 of 6.* (Report no. NASA/TM-20205010986).

National Transportation Safety Board (2010). Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River US Airways Flight 1549 Airbus A320-214, N106US Weehawken, New Jersey January 15, 2009. NTSB/AAR-10/03

Roscoe, A. H., & Ellis, G. A. (1990). *A subjective rating scale for assessing pilot workload in flight: A decade of practical use*. Royal Aerospace Establishment Farnborough (United Kingdom).

Roscoe, A. H. (1987). *The Practical Assessment of Pilot Workload: Flight Mechanics Panel of AGARD*. Advisory Group for Aerospace Research and Development Neuilly-sur-Seine (France).

Roscoe, A. H. (1993). Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics*, *36*(9), 1055-1062.

Sarter, N. B., & Woods, D. D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology*, *4*(1), 1-28.

Sevillian, D., S., Silveria, M., & Graeber, D. (2015) Flight deck engineering: Impact of flight deck crew alerting and information systems on English as a second language flight crew performance in airline flight operations. Downloaded from https://spacemedicineassociation.org/download/members_archives/yia_awardYIA.paper_2015.doc

Singer, G., & Dekker, S.W.A. (2000). Pilot performance during multiple failures: An empirical study of different warning systems. *Transportation Human Factors*, *2*(1), 63-76.

Yeh, M., Swider, C., Jin Jo, Y., & Donovan, C. (2016). Human Factors Considerations in the Design and Evaluation of Flight Deck Displays and Controls – Version 2. DOT-VNTSC-FAA-17-0.

U.S. Department of Transportation
John A. Volpe National Transportation Systems Center
220 Binney St.
Cambridge, MA 02142-1093

617-494-2000
www.volpe.dot.gov

DOT-VNTSC-FAA-24-04