FROM THA TO REAL-TIME MONITORING: INTEGRATING NEW VIEW OF SAFETY INTO FLIGHT TEST PROGRAMS

Thiago Fontes Macedo[1] & Moacyr Machado Cardoso Júnior[1]

[1]Instituto Tecnológico da Aeronáutica

Abstract

Test Hazard Analysis (THA) is a safety management process widely used in flight testing. Its basis is supported by the bowtie technique where the risk management focuses on interrupting the cause-effect line of action using mitigations and/or barriers. Technology and complexity have increased, requiring flight test organizations to attach subprocesses over the standard THA to absorb and account for uncertainties in all levels of the activity. In-flight testing is an activity that constantly requires aircraft systems experimentation in conditions never tested before. For this reason, an adapted THA was used to connect risk management to all organizational levels and to proactively control the performance variability of flight test systems aiming at the prevention of undesirable outcomes. Therefore, this work intended to verify to what extent the adapted THA process controls System Performance Variability (SPV) in flight test systems when compared to the standard one. The method walked through the characterization of mechanisms and tools used on an envelope expansion flight test project that were attached to the standard process, followed by the construction of a questionnaire fitted to the flight test community and intended to collect subject matter expert data regarding how the THA process accounts for SPV in complex scenarios. Results indicated that the adapted THA was able to connect communication feedback loops to other organizational levels, to proactively control safety margins and thus variabilities of system performance, and to actively learn from indicators measurements that triggered system modifications. According to the subject matter experts, the adapted THA is 44% more capable of accounting for SPV and constraints in flight test systems.

Keywords: Test Hazard Analysis, Aircraft Flight Test, Sociotechnical atm system complexity, risk management, and System Performance Variability

1. Introduction

Aviation is an activity with hazards that are actively managed from the aircraft's conception to its operation. Civil or military aircraft must undergo a development and certification process to ensure safe airworthiness. Flight test engineering is defined as the engineering activity associated with the test in flight or on the ground of an aircraft or item of an aircraft. It enables the development of an aircraft, where compliance is verified against performance specifications, flight handling requirements, investigation of new concepts, substantiation of design assumptions, etc. [1]. Several actors, such as test crews, ground support, aircraft instrumentation, and certification authorities, among others, are involved, which characterizes the involvement of several parties and different systems in the process. As a result, its complexity requires comprehensive and well-established risk management before, during, and after the flights. This fundamental process is a core characteristic of any flight test program and is exercised in all flight test communities.

The regulatory agencies around the globe along with the industry have several ways to establish risk assessment systems, but most of them converge on the process known as Test Hazard Analysis (THA) Figure 1. This process derived from the bowtie method and Fault Tree Analysis [2] [3] [4], that tries to look at hazard causes and effects listed according to human judgment or human value. Management occurs when mitigation measures and corrective actions are implemented and have an inferred effect on the probability of occurrence or severity of consequences of each hazard.

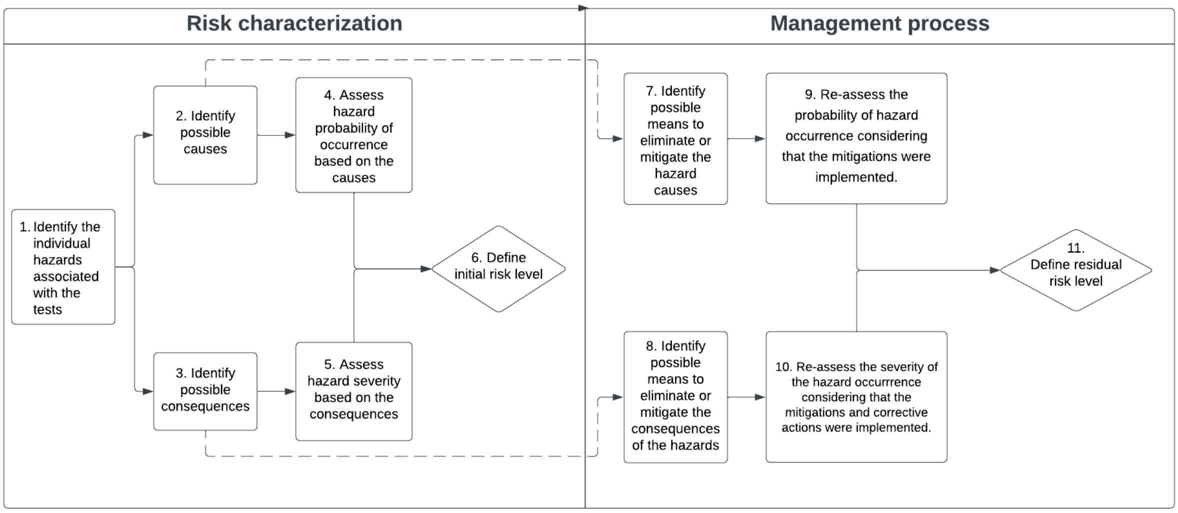

Figure 1 – Test Hazard Analysis Process adapted from [3]

The understanding of a system that has acceptable risk was initially associated with the absence of undesirable outcomes. This approach implies a condition where the management system is excited only when unwanted events happen. Aven [5] says that science is based on knowledge, and a science that must wait for an undesirable outcome to be studied seems to be counterintuitive and frustrating in terms of academic rationale. Hollnagel [6] suggests that “being safe” should be attributed to situations when things occur as planned and measurements of safety should not only be based on activities that did not have success but also on the successful ones.

Since the aircraft flight test activity deals with new systems and limited knowledge about the expected behavior of the system, monitoring the performance of every test event and comparing it to the predictions from simulations or ground test data is a common and wide strategy used by several organizations and entrepreneurs [1]. The similarity with Hollnagel’s theory on monitoring everyday activity is not a coincidence since flight test systems are extremely complex and innovative which require the involvement and coordination of several actors. This article will discuss how simple strategies, such as the THAs, can be enhanced to generate indicators that monitor and control aircraft flight test sociotechnical systems safety performance.

2. System Performance Variability (SPV)

Hollnagel [7] [8] introduces the concept of activities function with a degree of uncertainty which means that the sociotechnical system (Human, machine, and environment) will normally operate inside the designed limitations, but small deviations are normal and acceptable. When these variations become sufficiently high, the system is exceeding its accommodation capacity leading to a loss (Figure 2).

Acceptance of this concept leads to monitoring system status and constant control of SPV as the key to managing resilience in sociotechnical systems, thus providing safety. The flight test activity could be classified as a sociotechnical system since encompasses humans, machines, and the environment and therefore their interactions [7].

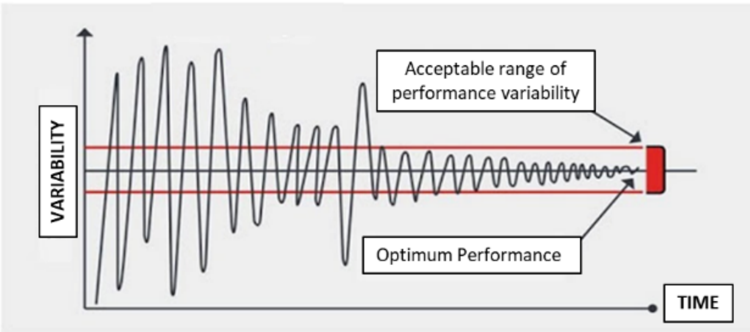

Figure 2 – Graphical representation of how SPV is monitored [Author, 2024]

Rasmussen [9] [10] exposes a metaphorical representation of system functioning. He says that systems are exposed to environmental and individual pressures. The environmental boundary is threatened when organizations increase pressure for efficiency and productivity. Individual boundaries are threatened when less individual effort is applied by human actors. To protect the system, Safety Management Systems (SMS) normally uses artificial boundaries drawn by safety regulations or norms. To realistically manage the everyday activity, an additional safety boundary should exist beyond the design safety boundary representing an acceptable level of variability, accounting for buffer, margins, flexibility, and tolerances used by the system. Figure 3 represents a degraded flight test system functioning resulting in an accident, adapted from [9].

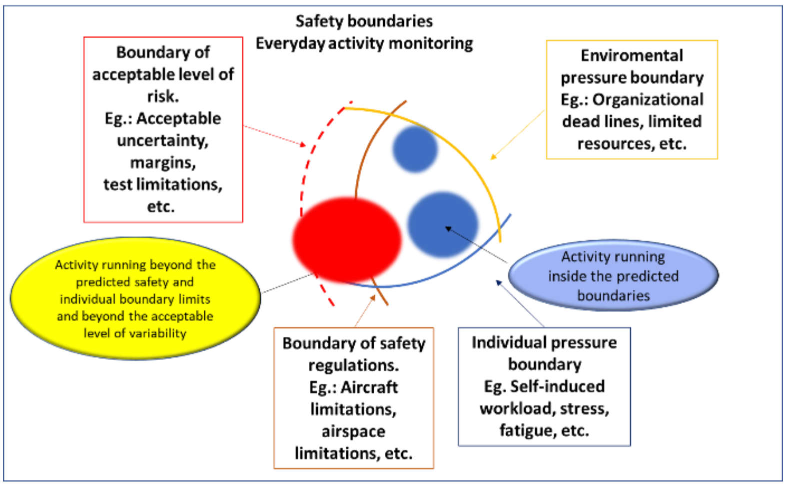

Figure 3 – Representation of a degraded flight test system functioning resulting in an accident adapted from [9]

[10] brought to attention the importance of modeling system behavior to scrutinize its structural elements, its external influences and communication. By doing so, a visualization of the information flow in the organization as long as its rules and subprocesses. They suggest a proactive approach by decomposing the system into elements and state that describing their function by causal-effect interactions is not a reliable approach because it does not consider scenario dynamicity and its integrations with the elements of an organization. Causal interactions are not sufficient to visualize the natural adaptative capacity of the society, it is necessary to map the information flow structure that is used to control hazards. This model can be used to analyze

system key subprocesses and elements even for unpredictable variations of the designed functioning. Moreover, the management occurs through control over closed-loop feedback instead of open-loop casual strategy. They conclude and suggest three principles:

- Endure identifying boundaries of safe performance.
- Make these boundaries visible to decision makers.
- Measures to counter react pressures that drive decision-makers towards the boundaries.

Woods et al. [11] comments that a huge challenge for the organization is to balance its acute and chronic goals. Acute goals are short-term elements and related to pointed productivity parameters, whereas chronic goals are long term and systemic which emerge from the interaction of system elements. Acute and chronic goals normally are difficult to achieve simultaneously. Normally, reaching acute goals means putting the chronic goals aside. [12] suggests four concepts to characterize resilience of a system and better monitor these scenario changes. The first concept is called *rebound* which is the system capacity to return to normal activities against a disrupting or traumatic event. The second is robustness, which is the system ability to absorb variations or disturbances. Next is graceful extensibility, referring to how a system adapts to handle rare, unexpected events forcing it beyond the designed margins. Lastly, sustained adaptability, meaning how the system handles trade-offs and constraints when interacting with other systems and the environment. He discusses that rebound and robustness focus is the knowledge and learning from previous undesirable situations. This knowledge should feed the learning about disturbances providing support to graceful extensibility which will explore how the system will adapt to handle disturbances instead of focusing on how or why the system bounces back to normal. This last concept should be able to limit the system performance and will be adjusted according to inputs from sustained adaptability fed by scenario pressures and interactions. He emphasizes that current tools and theories are well prepared to tackle known disturbances, but discusses how they could handle model changes while they are happening. He stimulates inquiries that should be done to overcome historical rebound and robustness practices focused on past occurrences, veering to knowledge-based improvements on proactive graceful extensibility and sustained adaptability. He offers these questions to explore how resilient the systems are:

- How do adaptive systems fail in general and across scales?
- How systems can be prepared for inevitable surprises while still meeting pressures to improve efficiency of resource utilization?
- What mechanisms allow a system to manage the risk of brittleness at the boundaries of normal function?
- What architectures allow systems to sustain adaptability over long times and multiple cycles of change?

These inquiries combined with the basic inquiries stated by Rasmussen [10] picture the theoretical basis of a resilient system, however, the interactions among them still need some scrutinization to bring the concepts to a more practical stage. Hollnagel, Woods and Leveson [13] state that the SMS needs to provide knowledge to organizations about when to relax pressure to reduce risk, sacrificing sometimes the productivity to achieve more productive results in the future and adjusting to new scenarios. The resilient process needs to control the on-time performance of the system to be able to apply a real time corrective action based on the predicted maximum defined system variability. They suggest a process based on monitoring the quality of system performance as a manner of coping with complexity and maintaining control of the risk. Controlling a system in this case is to perceive, reduce and avoid undesirable performance variability. To have control, the knowledge about what happened, what happens and what may happen is necessary. At the same way, the knowledge about what to do when unwanted variability happens and have the resources to do it is fundamental. More important than losing control is to know when control is going to be lost which, hence, is associated with lack of time, knowledge, competence, or resources.

Figure 4 shows a representation adapted from [13] of what would be the qualities of a resilient system using an approach and landing operational situation. The dynamicity of the environment is exemplified by the wind, runway conditions or the organization pressures over the airliner to land the aircraft at the correct time. The system knowledge, competence, time, and resources would interact with the environment through its anticipation, attention and response capacity. At the end, the goal is to constantly learn and update those system qualities to control the variability of the system performance.

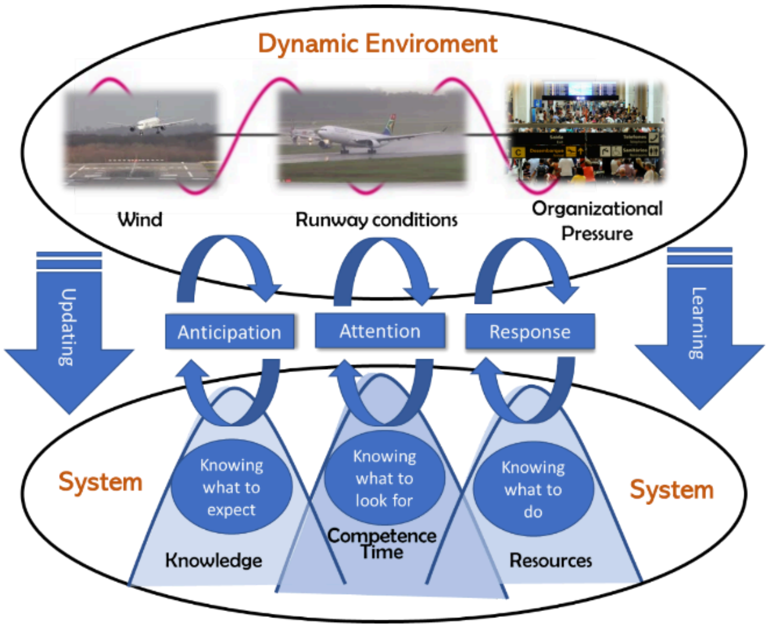

Figure 4 – Qualities of a resilient aircraft landing procedure adapted from [13]

Recalling that traditional SMSs normally aim to reduce the number of undesirable events and need to be complemented by the RE approach where the objective is to improve system ability to have success in several conditions, [14] suggests that "a system cannot be resilient but, a system can have a potential for resilient performance". In other words, productive safety (Safety II) needs to complement protective safety (Safety I). Proactive adjustments are necessary to control the SPV, by monitoring daily normal performance variations, setting the boundaries, commented by [9], responding to those variations, learning from potential and finally anticipating potential future disruptions. Respond, monitor, learn and anticipate are the four dimensions [14] in the Resilience Analysis Grid (RAG), which is a questionnaire intended to identify these four abilities of resilience which could be recognized in any event analysis. He states that the choice of abilities is pragmatic, and no strong theory supports the idea that a distinct set should be chosen. According to him, resilience points to something that the system does rather than to something that the system has, meaning that resilience performance must be understood by the conditions that make a resilient system instead of what is a resilient system. Consequently, assessing the four abilities makes more sense than assessing resilience performance as one unique ability. The abilities are defined as follows:

Ability to respond or know what to do: This ability is intimate to the survivability of the system. The response should be timely and feasible to ensure the desired objective without incurring harmful conditions. The process should encompass the ability to detect that something has happened, recognize what is the condition and whether is necessary to respond and an action measure with defined procedure and start and stop criteria. The proper resources should be available and ready to be consumed and recovered to the normal system state level. A classification rationale should be present to differentiate levels of response.

Ability to monitor or know what to look for: This ability is related to how the system can perceive its own performance and the environment. It relies on indicators to perceive the performance system and environment boundary. Two types of indicators are mentioned. Leading indicators, which are intended to anticipate the closure to the boundaries and lagging indicators, which are based on past statistics. Leading indicators require more modeling and system understanding, consequently requiring more effort (financial, personnel, resources, etc.) and time for simulation and prediction. Lagging indicators are more tangible and applicable to the scenario from where they were produced, but normally just protect against specific environment and system conditions.

Ability to learn or know what has happened: this ability is importantly relevant when the environment is non predictable and unstable. The systematic and efficient learning process from everyday activities will enhance safety and the knowledge about how system works and thus feeding the models and prediction capacities. Normally, learning from near misses and why things went right will improve learning ability.

Ability to anticipate or know what to expect: consists basically the ability to anticipate the future outcomes, either disruption, constraints or novel opportunities or modifications on the environment conditions.

These control principles perfectly align with the flight test scenario where a number of variables and unprecedent conditions happen differently for each new airplane designed and a great database of lessons learned is available.

In summary the resilience engineering system qualities, according to Hollnagel, Woods and Leveson (2006), can be enumerated as an interaction process of:

- Anticipation (Knowing what to expect)
- Attention (Knowing what to look for)
- Response (Knowing what to do)
- Time (Quality variable)
- Learning and Updating

3. The method

The methodology aimed to indicate areas on the THAs process that could have adherence to resilience engineering principles and areas that could be enhanced by these principles.

I. Description of a case study where the adapted THA process was used

II. Construction of questionnaires to verify if the adapted THA process can control the performance variability of a flight test sociotechnical system

III. Control phase: Questionnaire application on the flight test community

IV. Experiment phase: Questionnaire application on the flight test community that was exposed to the adapted THA process

3.1 Case study research

The experiment consisted in reviewing the planning, execution, analysis, and reporting activities that occurred during Flight Test Program and recognizing in these activities subprocesses that had characteristics of system performance variability control. The flight test program aimed to certify a twin engine, medium category aircraft to operate in a runway which had slope characteristics beyond the certified normal. The study started by initially examining the THA process recorded in the flight test planning. On a second stage, SPV control tools and features incorporated into the standard THA process were identified and included in the mapping, resulting in the adapted THA process. The outcomes, lessons learned, and tools applied to the program were collected to rationalize the possible effects of the THA process. The aim is to clarify how the adapted THA process contributed to visualizing, absorb, and control the performance variability of the sociotechnical flight test system. Few SPV control tools were identified during the flight test program, in this article we will highlight the two main control tool processes developed during the program.

3.1.1 Structured information flow to constantly update the system

[10] divide society in levels where the information objectives flow downwards, and the actual system state propagates upwards. The analysis of system safety is executed by judging how close the actual information is to the objective information. To make this comparison, the mechanisms should be based on the system performance indicators defining limits and margins for staffing, management, and system operation. Figure 5 shows a representation of the flight test program under study.

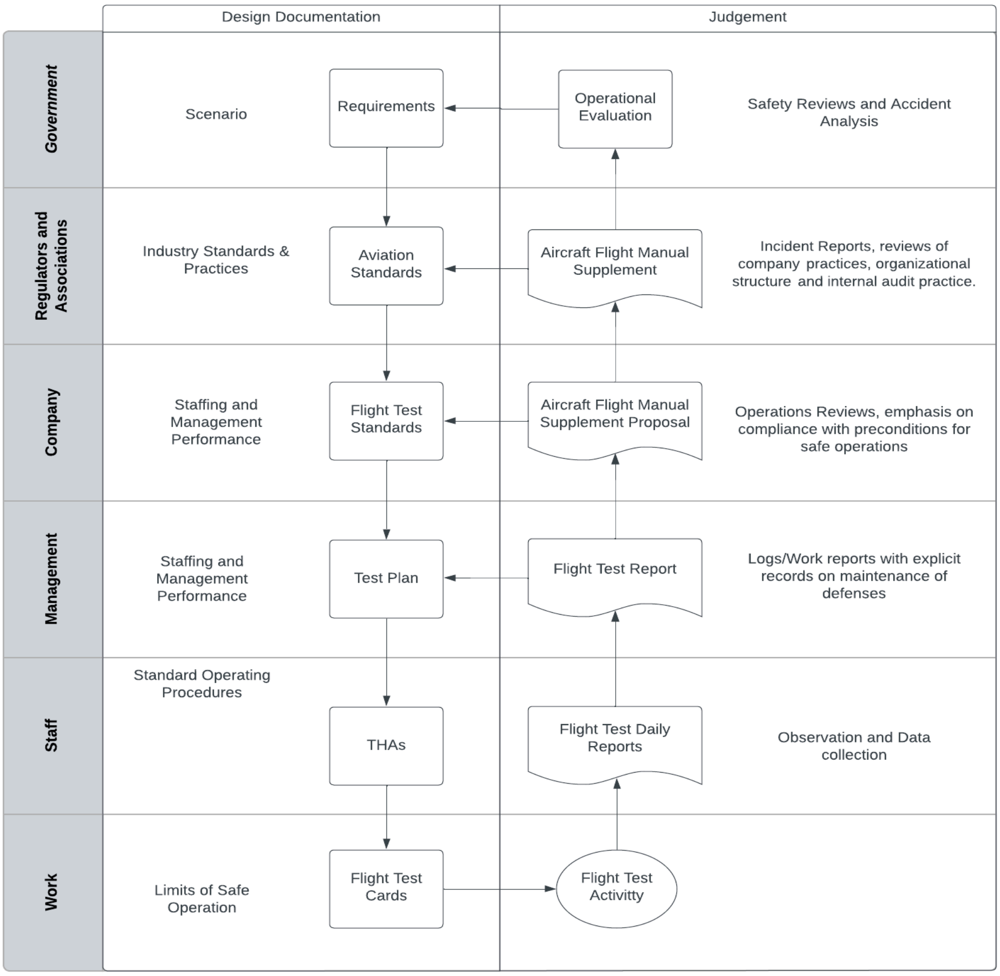

Figure 5 – Flight test information flow [Author, 2024]

When looking to the downward flow, the program identified that the guidance [15] provided by the regulatory agency did not contain clear safety boundaries leaving to the team the responsibility of managing its efficiency and risk. The strategy adopted was a buildup approach process to acquire knowledge about the system using smalls gradual steps. This strategy is well-known in the flight test community, but how large or small these steps are, relies heavily on flight test crew and engineering judgment. Following the guidance and using the standard THAs was run, the team identified that it was necessary to build knowledge about how the test subject (Aircraft, pilot, scenario) would behave in certain conditions. For example, in many cases, the aircraft manufacturer data was not available, the knowledge about the environmental conditions was not accurate, the pilot's experience in landing on short runways was limited, the conditions of the runway were inaccurate, and others. As a result, the team started to associate hazard probability to lack of knowledge or prediction capacity about what would be the flight test results. To better measure the risk, a set of indicators were constructed and positioned on specific areas that could cause the hazard. This was the first add-on of this experience, where the team associated the hazard probability to the epistemic uncertainty about the phenomena or system behavior. [16] suggested that a resilient system cannot rely only on safety barriers and new procedures, it requires a monitoring architecture of how things are done to understand the complexity and enhance the ability to maintain control. Even though the probability of occurrence used in the THAs is an inference, in this project, the association to lack of knowledge to predict the outcomes was providing support it. To make this parameter more readable and to continuously feedback the knowledge state of the program, the program incorporated a few mechanisms to the standard THA as illustrated in yellow on Figure 6:

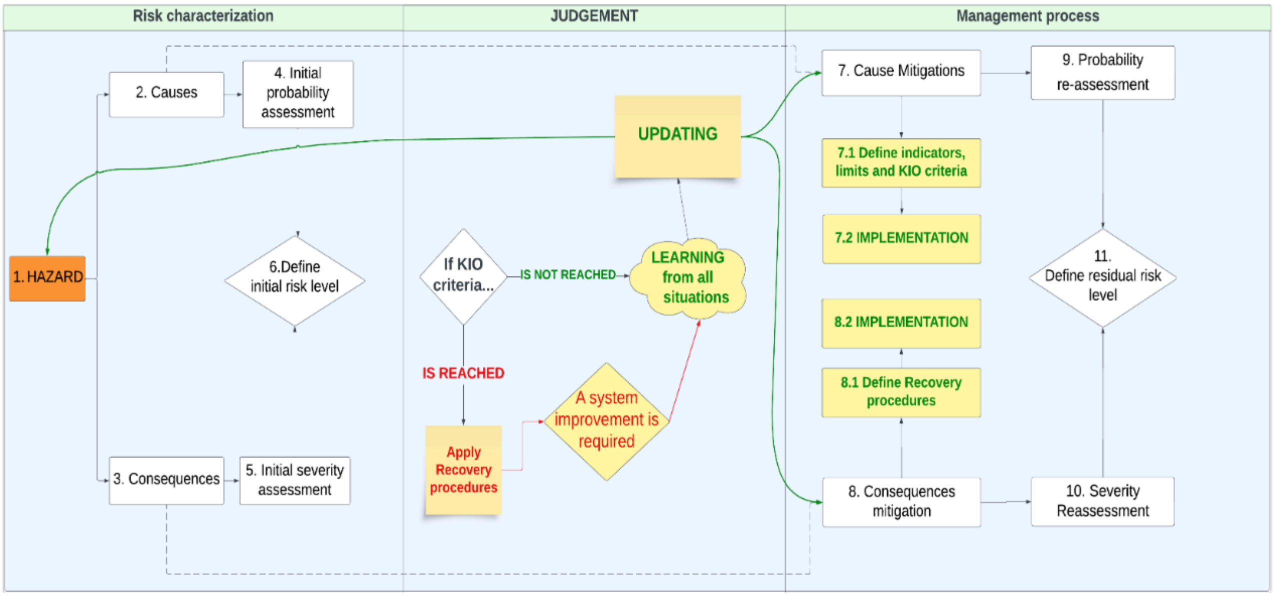

Figure 6 – The Adapted THA process flow [Author, 2024]

An enlightening example was the usage of the Cooper-Harper (CH) scale [17] as an indicator, which is a 10-level scale where pilots assign rates to a predefined task. The scale was found suitable for the task since it contains aircraft parameters combined with human workload measurements resulting in a system performance indicator. However, a third variable needed to be included in this measurement, the scenario. In that, the team designed a buildup approach to monitor and obtain knowledge about the system. They executed the tasks in a simulator initially, moving to tests on runaways with slopes of 0%, 0.8%, and 1.4%. As the operations occurred in these runways, data was being gathered and processed, improving prediction capacity, and continuously updating the knowledge about the system. The control was accomplished by assigning a Knock-It-Off (KIO) to the indicator, in this case, if a score 7 was assigned in any flight, a recovery procedure was initiated as established on the THA and the operation would come to a pause. Following this, a reassessment was mandatory and must result in a system modification that improved either aircraft performance parameters, human workload, or scenario conditions. This way, two feedback loops were created, one that flowed when the KIO or limits were not reached, which continuously updated system safety margins, mitigations, recovery procedures and hazards, and when KIO was reached requiring a stop and a mandatory system modification.

Another example of indicator usage was structural loads measurements on landing gear during the landings associated with CH scale scores on the same task. This combination was found to have correlation which led the team to tradeoff margins in between these two dimensions to manage the closeness to KIO limits. Before even reaching KIO, the team was able to execute small system adjustments, such as, reduction of the maximum weight of the aircraft (aircraft performance) or limitation of wind/weather conditions (scenario restriction) as a tradeoff technique among indicator safety margins. The team decided to maintain the human performance (CH) as minimum as possible, since previous accidents indicated that the workload to operate the aircraft was already above normal. This technique is supported by Woods (2014) where he discusses economic-safety tradeoffs. Therefore, the performance of the aircraft was traded by more margin to pilot error, making the balance heavier on the safety side instead of the available payload that the aircraft can deliver.

Using these real examples, some associations are worthy of being made. Firstly, the process focused on determining what is the knowledge edge about the phenomena, setting visible limits and operational margins (KIO). A clear information and decision process flow in case KIO is reached facilitated decision making in flight especially when unexpected outcomes occurred and providing sufficient margin to the crew. Secondly, as the knowledge was being processed, a proactive tradeoff in between margins was executed based on the objectives of the program. This was possible because indicator information also had meaning to organizational objectives. [10] say to identify system boundaries, make it visible and be able to counteract pressures. These subprocesses allowed the flight test crew to always visualize limits and progress to next steps as knowledge was being built.

Another contribution made was the control over mitigation measures, indicators, and recovery procedures implementation. It consisted of creating, in parallel, a sub process to ensure minimum resources and skills to support a reliable functioning of these three actions. Examples could be as simple as inserting a note on the test card (mitigation measures), executing simulator training to recover from hazardous conditions (recovery procedures), or executing error analysis on aircraft sensors and set on time KIO criteria (indicators). [18] talk about “Slack” and define it as an asset to mitigate variability, providing extra space for performance adjustments during predicted and unpredicted events. Thus, the implementation step came to ensure that the control of the safety margin/Slack was being properly installed and had reliable operability.

3.1.2 Improving communication across system organizational levels

Sanders and McCormick [19] already suggested that a combination of design and behavior approaches is usually necessary to maintain a system safe. Risk management is mainly executed by humans; therefore, human behavior needs to be closely observed and guided. A proper design will reduce hazards and create mitigations to avoid them and, a proper human behavior process through training, reinforcement and feedback can control and absorb unpredicted situations. They say that we cannot wait for accidents to happen to feed hazard likelihood, instead multiple control strategies of system behavior must be implemented. As the THA was being executed for all hazards identified, the author identified extra steps that needed to be implemented at the organizational level. To clarify the results of this continuous process, an illustration of the “Work as Done” during the program is exposed on Figure 7. The mechanisms included in the system were represented by yellow tags in the design column. This process guided the team towards controlling not only system performance in flight but also project performance to effectively direct the efforts to understand what to expect, what to look for and what to do [13]. In the design column of Figure 7, it is found subprocesses were included to build safety boundaries and margins, including control tools to monitor human, machine, and project performance variability.

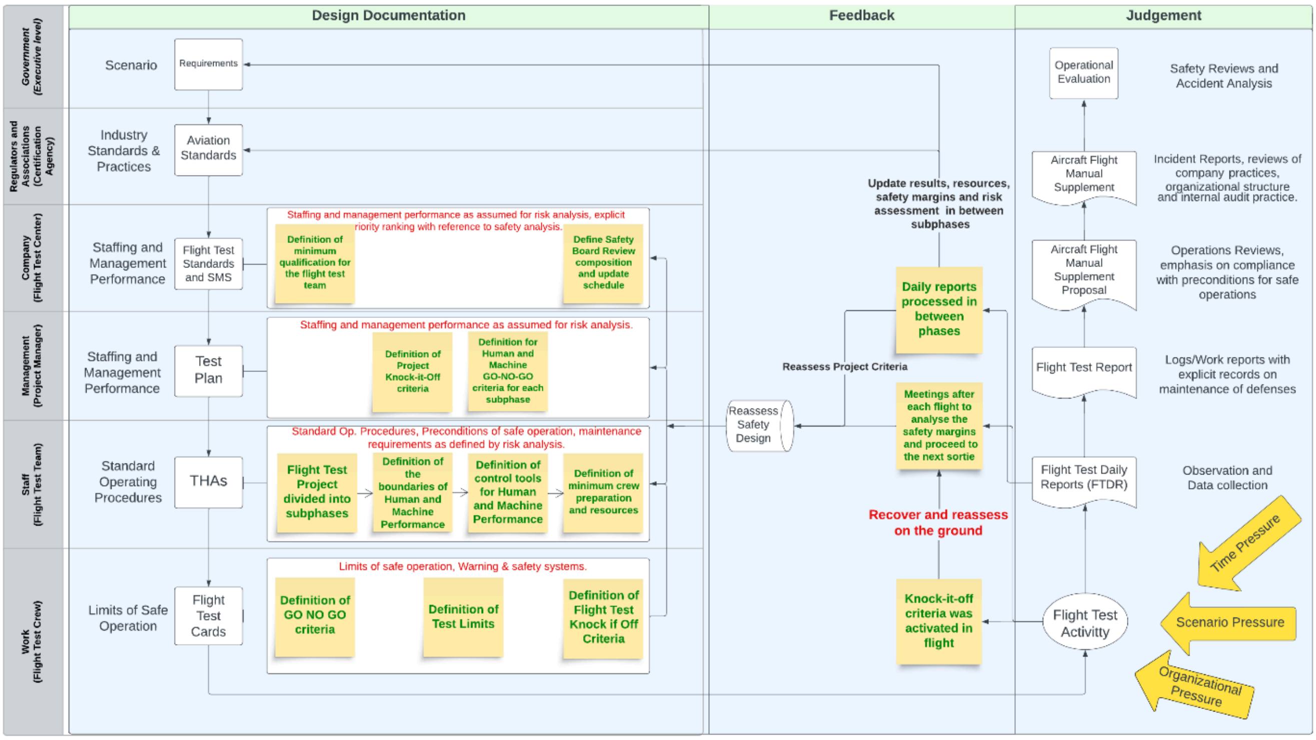

Figure 7 – Structure information flow with management tools [Author, 2024]

(Author, 2024)

3.1.3 Learning and updating process

[19] say that the training needs to stress the learning of safe behaviors instead of avoidance of unsafe acts, it needs to be executed in suitable settings to enhance transferability to the real scenario and to have an indicator that shows effectiveness and timely and proper feedback. Thus, the timely and combined feedback provided by daily meetings, Flight Test Daily Reports (FTDR) and KIO criteria activation identified the necessity of improving crew training by the knowledge gathered during flight, assessed during the daily meetings, and updated to all organizational levels through the FTDRs. Figure 8 shows how this learning point was updated. Observe that phase 2, Figure 8, was subdivided into two subphases because of KIO activation occurred during phase 1. The system was modified and as a result more training and verification of the indicators implementation was reassessed on phase 2.1 before proceeding to the pre-planned phase 2.2. On phase 2.1 (, the KIO criteria were again activated, this time the team decided to execute tradeoff in between safety margins as explained earlier when aircraft performance was compromised in lieu of pilot performance.

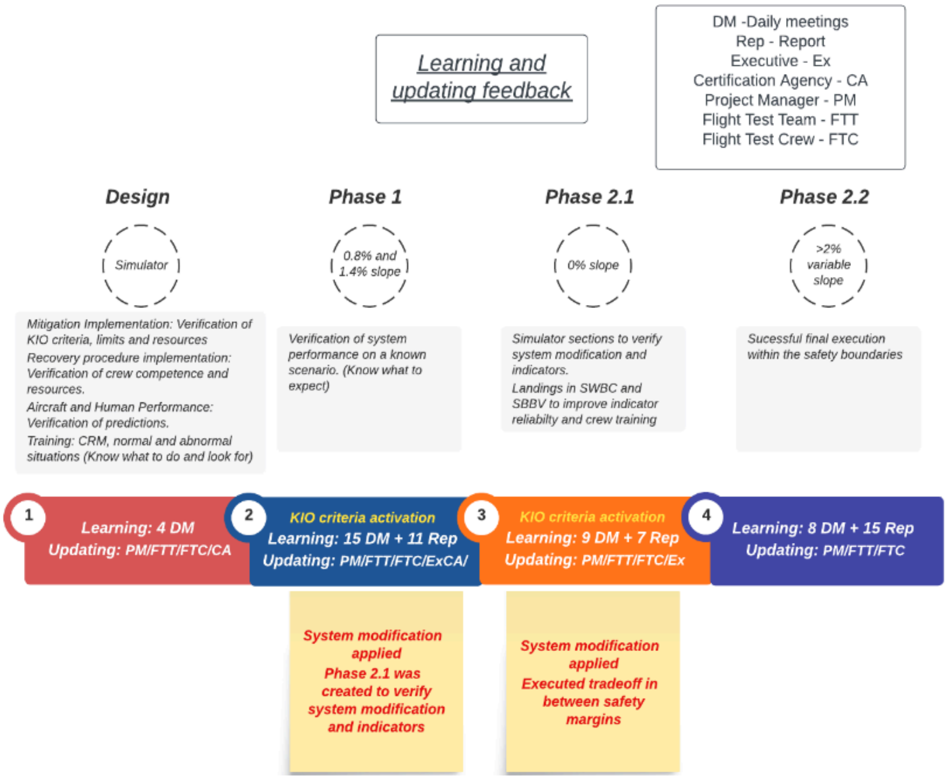

Figure 8 – Flight test program subphases and feedback results [Author, 2024]

The discussed aspects contributed to maintaining a sustained update and progress towards the unknown area of system behavior. It is possible to infer that the graceful extensibility (as per the request to test in an additional runway due to the constant indication of high-performance variability CH7) and sustained adaptability (as per the safety margin trade-offs accomplished) were embedded in the process during the operation. In addition, it was also observed that the process dispended attention to verify if the knowledge, competence, resources, and time were being properly monitored and kept in a sufficient level to control the SPV as suggested by Hollnagel, Woods and Levenson (2006).

In summary the following characteristics were identified in the adapted THAs: definition of system boundary and margins by the implementation of indicators that control the everyday SPV; timely, realistic and communication feedback loops to enhance learning and damp time and organizational pressures through daily meetings and FTDRs which contributed to monitor safety margins by lower (daily meetings) and top (FTDR) organizational levels; an adaptable sociotechnical system as a result of the learning and updating from the communication feedback loops; Clear procedures associated to the indicators which enhanced the decision process under unexpected behavior; Procedural THA checklist to confirm mitigations recovery procedure and indicator implementation; Competence, resource and knowledge monitoring to maintain indicator safety margins.

3.2 Construction of questionnaires to verify if the adapted THA process can control the performance variability of a flight test system.

After highlighting the SPV control tools, questionnaires data collection was chosen to verify how a control group composed by flight test pilots and engineers understand the standard THA process when they are inquired about the resilience engineering principles, and how an experiment group, also of flight test pilots and engineers which had the experience in using the adapted THA process, understand the same concepts.

Humans are excellent integrators and asking them about the performance of a system could be a cost-effective solution and may provide useful indication to the examiner. Therefore, a questionnaire was the tool chosen to collect data from subjects that are currently living and using the standard THA process. According to [20], the construction process may follow the steps and principles below:

- Select the questionnaire type based on the sample size and expertise. Principle: a statistically representative sample size;
- Select a response scale and a descriptor set for the responses. Principle: a parametric rating scale that is balanced and has equal intervals among the levels;
- Construct the wording of the questions. Principle: questions that narrow down as much as possible the evaluation areas; In that, the vocabulary should be speaking the language of the interviewed individuals. Principle: questions that are objective and related to a performance measure. In that, positive phrases should be used, avoid double and/or leading questions, avoid emotionality, keep the questions short and keep the relevance of the questions related to the objective of the test;
- Structure your questionnaire with clear instructions, consistent format, brevity, and appropriate materials; and
- Review the questionnaire for grammatical mistakes and execute a quality check with subject matter experts. Principle: an acceptance criterion defined beforehand.

A combination of multiple choice with open space for justification was chosen. According to [21], Subject Matter Expert (SME) is an expert person that can consistently differentiate about a subject or an inquiry. Experience, certification, and qualification, up to a certain extent, could indicate the expertise but consistency and discrimination drive a more solid indication of expertise level. Following these lines, a selection of questions was used to characterize the level of expertise. The first one intended to verify consistency and discrimination by presenting a standard THA table from [22] to the respondent and asking about the level of agreement to the process (Figure 9). If the subject had agreed to the process, he was classified as a SME, since the THA presented is widely known and approved by several certification agencies and flight test schools.

Test Hazard Analysis										
Risk characterization						Management				
Hazard	Cause	Effect	Initial Risk			Mitigation	Corrective Action	Residual Risk		
			Probability	Severity	Risk			Probability	Severity	Risk
Describe "what might happen that could adversely affect safety"	Describe "what might cause the hazard"	Describe the effect in case the hazard happens	Describe the hazard probability of occurrence	Describe the severity of the hazard effect	State the initial risk based on the probability and severity associated	Describe how or what will be done to minimize the risk.	Describe what will be done in case the hazard occurs to reduce the severity.	Describe the hazard probability of occurrence considering that the mitigations were implemented.	Describe the severity of the hazard effect considering that the mitigations and corrective action were implemented.	State the overall risk based on the latest probability and severity associated

Figure 9 – Test Hazard Analysis Process adapted from [22]

Other subject characteristics such as flight test hours, flight test graduation school, flight test instructor qualification and if the subject was a pilot or engineer were used to characterize the subjects. Sample size prediction was accomplished using Eq.1:

$$\eta = \frac{NZ^2}{\frac{sE^2(N-1)}{P(1-P)}+Z^2} \qquad \text{(Eq. 1)}$$

Where: sE = Sampling error
Z = Z score for desired confidence level
P = Percent of sample in one category (SME)
η = Sample size
N = Population size

Two five-level Likert ratings scales were offered to the subjects during the questionnaire construction phase (agreement and frequency). Questions from the academic authors were translated to the flight test community language to allow clear and direct communication. The process consisted of transcribing each question to affirmative format and associating the academic concepts to situations or scenarios that could occur in flight tests. Next, interviews were executed one by one, face-to-face, with SMEs which were initially briefed about the objective of the section, to verify the understanding of the statements. The SME were asked to answer the question and were inquired about why they chose that specific answer. Next, the correlated academic question was presented and an explanation about the academic concept was exposed to the SME. Finally, the SME was questioned about the semantical alignment in between the academic and translated version. In case it was identified a misalignment, the SME was asked to suggest a modification on the translated version. The process was repeated to all questions and subsequently with different SMEs, until no modifications were required.

3.3 Control phase: Questionnaire application on the flight test community

The last version questionnaire was applied to the flight test community using google forms. This questionnaire did not have open-ended questions to facilitate data reduction. It consisted of 5 questions for subject characterization and 16 questions related to the research. The time to answer was estimated at 15 minutes total. The subjects answered the questionnaire via Google Forms which were sent via email and social media to achieve maximum reach. Subject identification was not required to stimulate participation and improve reliability of the answers and to deal with confidentiality constraints. The data collected was organized in a tabular form presenting the percentage of subjects that chose each option for each question. Next an aggregation process was applied to each question to group agreement (Strongly Agree and Agree), neutral (Neutral) and disagreement (Strongly Disagree and Disagree) answers. Mode, average and standard deviation analysis was conducted to both conditions (aggregated and non-aggregated). Following, based on the mode and the average results, the questions were escalated again onto five levels of agreement (Strong Agreement, Agreement, Neutral, Disagreement and Strong Disagreement). Finally, the scaled questions were correlated with the authors that generated the questions.

A MultiDimensional Scaling (MDS) analysis was executed to validate the aggregation process of the respondents and utilizing a cluster analysis perspective. MDS is a statistical technique used to analyze complex datasets that have multiple dimensions. In this case, similarities and dissimilarities between questions and respondents are processed by using the least squares technique on the Euclidean distances calculated between dimensions. The main objective is to represent, in a two-dimensional space, relationships between questions and respondents [23]. To verify if the aggregation strategy, mentioned in the previous paragraph, was acceptable, the MDS was run using the respondents as dimensions to be reduced. If the resultant clusters had a correlation with the aggregation, the data reduction was deemed valid. A second utilization of the MDS was accomplished using the questions as dimensions to be reduced, in this case, if clusters were identified, the clustered respondents indicated similarities in their responses. The RStudio software was used to execute the calculations. The analysis was based on the inference regarding a possible correlation between standard THA process characteristics and SPV characteristics, tools and concepts.

3.4 Experiment phase: Questionnaire application on the flight test community that was exposed to the adapted THA process.

The same questionnaire structure was used on 5 subjects that were exposed to the adapted THA process. The difference was that the subjects were asked to answer the questions only taking into consideration the adapted THA process used during Surucucu flight test campaign. The distribution and application method were as the control phase. The same data reduction and analysis strategy used to the control group was applied to the experiment group. A comparison between the control and experiment group results is exposed to present in what extent the adapted THA process fills the knowledge gap proposed in this thesis

4. Results and Discussions

4.1 Questionnaire construction

Questionnaire construction was the foremost step to avoid polluted data or biases. The flight test community is somewhat a small community. Nowadays, only 8 flight test schools are recognized by the Society of Experimental Test Pilots and the Society of Flight Test Engineers [24] [25]. Although the population tends to be limited, these professionals are highly qualified and normally acquire extensive experience on real flight test operations in a short term, which facilitates their qualification as SMEs. The two international societies had respectively 2504 and 1300 signatory members around the world. Therefore, this work considered a population of 4000 professionals to calculate the sample size. The graduation syllabus for flight test pilots and engineers are fairly similar across the schools, which normally are verified by these societies when the recognition process is applied. Taking into consideration the strict and very regulated work scenario of these professionals. On the control phase, 28 questionnaires were collected. The respondents presented to be mostly composed by SMEs (Table 1) since 89.3% of the subjects Strongly Agreed or Agreed to the standard THA process presented in the questionnaire. 89.3% of the subjects were instructors and had experience above 250 hours indicating solid experience on executing THAs throughout their careers. P value was assumed to be 0.89 (SME proportion) resulting in a confidence level of 95% with a sample error of 11% for the control group. All subjects graduated in recognized test pilot schools reinforced the expertise about the THA process.

Table 1 – Sample characterization

Characteristic	Options	Control		Experiment	
		%	#	%	#
Test Crew	Test Pilot	64.3	18	40	2
	Test Engineer	35.7	10	60	3
Graduation School	Brazilian Test Pilot School (EFEV)	78.6	22	80	4
	French Test Pilot School (EPNER)	7.1	2	-	-
	International Test Pilot School (ITPS)	7.1	2	-	-
	National Test Pilot School (NTPS)	3.6	1	-	-
	United States Naval Test Pilot School (USNTPS)	3.6	1	20	1
Flight Test Hours	Less than 100 hours	7.1	2	0	0
	Between 100 and 250 hours	3.6	1	40	2
	Between 250 and 500 hours	28.6	8	20	1
	More than 500 hours	60.7	17	40	2

Flight Test Instructor	Yes	89.3	25	60	3
	No	10.3	3	40	2
THA SME	Strongly Agree	46.4	13	80	4
	Agree	42.9	12	20	1
	Neutral	3.6	1	0	0
	Disagree	3.6	1	0	0
	Strongly Disagree	3.6	1	0	0

Figure 10 exposes five control group dissimilarities characteristics presented in Table 1. Engineers appeared to be grouped on the left side and pilots on the right side of dimension 1, but still with no definitive cluster. The few non-instructors, non-EFEV graduation school subjects, and subjects with experience below 250 hours were scattered around the graph area. Again, no tendencies for cluster were perceived. The SME assessment graph showed that non-SMEs subjects (red, yellow, or gray triangle) were scattered around the graph area indicating even though they have more probability to enhance result sampling error, they still are part of the group common sense. The analysis did not bring any tendency for clustering supporting the argument that the control group was uniform and had a prominent level of consensus, reliability, and discriminability, which are characteristics of SMEs.

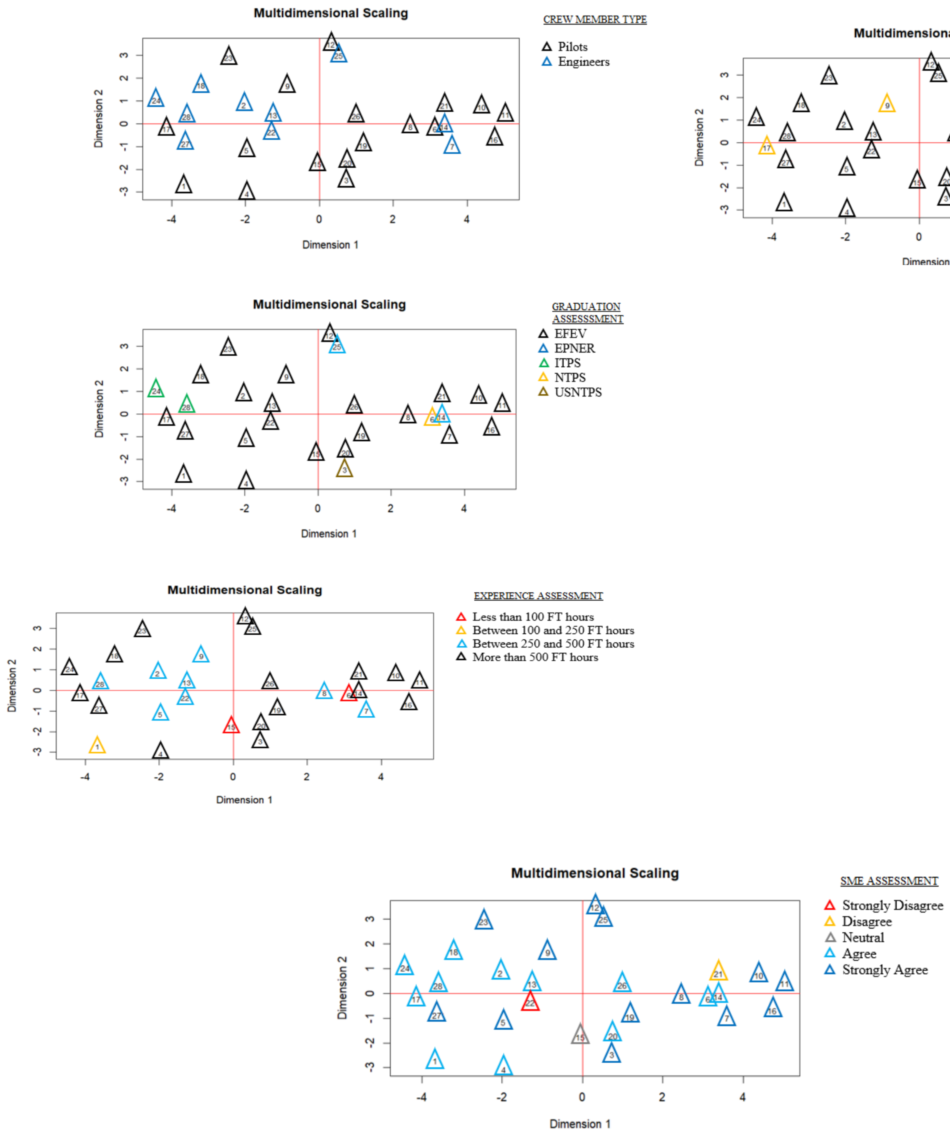

Figure 10 – Control Group MDS of the subjects [Author, 2024]

The experiment group was composed of 5 subjects, 2 pilots and 3 engineers of a universe of 6 which was the number of professionals exposed to the adapted THA process. One of the subjects did not answer the questionnaire. These professionals were the flight test pilots and engineers of Case Study flight test operation. It was observed a close balance between pilots and engineers ‘quantity, a heavy presence of EFEV graduates, a fairly even distribution of flight test experience and a balanced instructor qualification presence. The SME criteria was achieved for all subjects where 100% of the subjects agreed with the characterization of the standard THA process.

4.2 Statement construction and quality check

The type of questionnaire chosen for statement construction phase was a combination of rating scale with open-ended option since it was necessary to verify if the statements, structure, and answers would be understood by the subjects and to mitigate unanticipated answers that would happen during control and experiment phases. The combined rating scale and open-ended questionnaire type used during the construction phase provided feedback to the author about the semantics of the statements and mitigated the possibility of having unexpected answers during the control and experiment phase increasing the chances of having consistent and reliable data. SMEs addressed understanding of the academic terms, new concepts for the community and breaking down generic statements. The agreement scale was chosen suitable since the statements were in an affirmative construct and provided objective answers about the veracity of the statement. They also added that the agreement scale conducted a faster and more logical association between the statements and the scale levels when comparing to frequency scale. Table 2 shows the academic statement conversion accomplished by SMEs.

Table 2 – Academic statements conversion to flight test community language

#	Rasmussen and Svedung (2000) statement	#	Statements converted by SMEs
1	Endure in identifying the boundaries of safe performance	1a	THAs leads the tester to identify Organizational pressures and the associated boundaries.
		1b	THAs leads the tester to identify Individual pressures and the associated boundaries.
		1c	THAs leads the tester to identify boundaries of safety regulations (test limits) for the predictable hazards.
2	Make these boundaries visible to decision makers	2	THAs always leads the tester to operationally visualize test limits (boundaries) for each specific hazard.
3	Measures to counter react pressures that drives decision-makers towards the boundaries	3	THAs provide measures to counter react pressures that drive the flight test crew beyond the acceptable identified limits.
	Woods (2015a) statement		
4	How adaptive systems fail in general and across scales?	4a	THAs are able to identify the interactions between hazards.
		4b	THAs are routinely reassessed and updated after each test flight.
5	How systems can be prepared for inevitable surprise while still meeting pressures to improve on efficiency of resource utilization?	5	THAs has mechanisms to identify unpredicted hazards during flight test campaign (limited resources).
6	What mechanisms allow a system to manage the risk of brittleness at the boundaries of normal function?	6	THAs have mechanisms to verify the effectiveness of test limits.
7	What architectures allow systems to sustain adaptability overlong times and multiple cycles of change?	7	THAs has provisions to routinely reassess test limits during flight test campaign.
	Hollnagel (2015) statement		
8	Anticipation (Knowing what to expect)	8	THAs leads the tester to predict what would be the expected aircraft/system behavior
9	Attention (Knowing what to look for)	9	THAs leads the tester to set test limits when the data is not sufficient for a reliable prediction.
10	Response (Knowing what to do)	10	THAs leads the tester establish procedures to recover when the flight test limits are exceeded.
11	Time (Quality variable)	11	THAs consider the impact of time availability during the flight testing to maintain the acceptable level of risk.
12	Learning	12	THAs have mechanisms to learn from failures and success.
13	Updating	13	THAs have mechanisms to indicate that an update in flight test team knowledge, competence or resources is required.

4.3 Control group results

After aggregating control group raw results, they exposed the extremes in terms of agreement and disagreement keeping the neutrality isolated which reduced the answers to 3 categories: Agreement, Neutrality and Disagreement (Table 3). To analyze the behavior of the group and verify its tendency related to each question, an aggregation between the average and the mode was made. The mode shows how frequently an option was chosen while the average indicates where the sample opinion is leaning. If a question had a mode of 4 and an average of 4, it was inferred that the group had a strongly agreement on that question. If the mode was 4 and the average was 3, the question had an agreement. In case questions 6 and 12, we have two modes (2 and 4) with the average in 3 which led the author to classify this result as a group neutrality of lack of understanding about the question. To better visualize the agreement level of the aggregated questions to the academic authors, the questions were rearranged in Table 4

Table 3 – Control group raw results aggregation

Question #	1a	1b	1c	2	3	4a	4b	5	6	7	8	9	10	11	12	13
Strongly Disagree & Disagree	43%	32%	4%	14%	29%	39%	50%	50%	39%	43%	7%	14%	7%	54%	39%	57%
Neutral	7%	21%	0%	11%	11%	25%	21%	18%	21%	25%	14%	7%	4%	18%	21%	7%
Agree & Strongly Agree	50%	46%	96%	75%	61%	36%	29%	32%	39%	32%	79%	79%	89%	29%	39%	36%
Mode	4	4	4	4	4	2	2	2	2/4	2	4	4	4	2	2/4	2
Average	3	3	4	4	3	3	3	3	3	3	4	4	4	3	3	3
STD DV	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 4 – Control group academic author's association

Question aggregation	Rasmussen and Svedung	Woods	Hollnagel, Woods and Levenson
Strong Agreement	1c, 2,	-	8, 9, 10
Agreement	1a, 1b, 3	-	-
Neutral	-	6	12
Disagreement	-	4a, 4b, 5, 7	11, 13
Strong Disagreement	-	-	-

To support this data manipulation process, MDS analysis was executed on control group data. Observe in Figure 11 that correspondent clusters were correlated to the categories aggregated in Table 4. The consequent grouping indicates that the aggregation and categorization applied has positive relation when represented in two dimensions. Questions 9 and 12 a located slightly outside the disagreement clusters towards to the center which indicates the trend for neutrality.

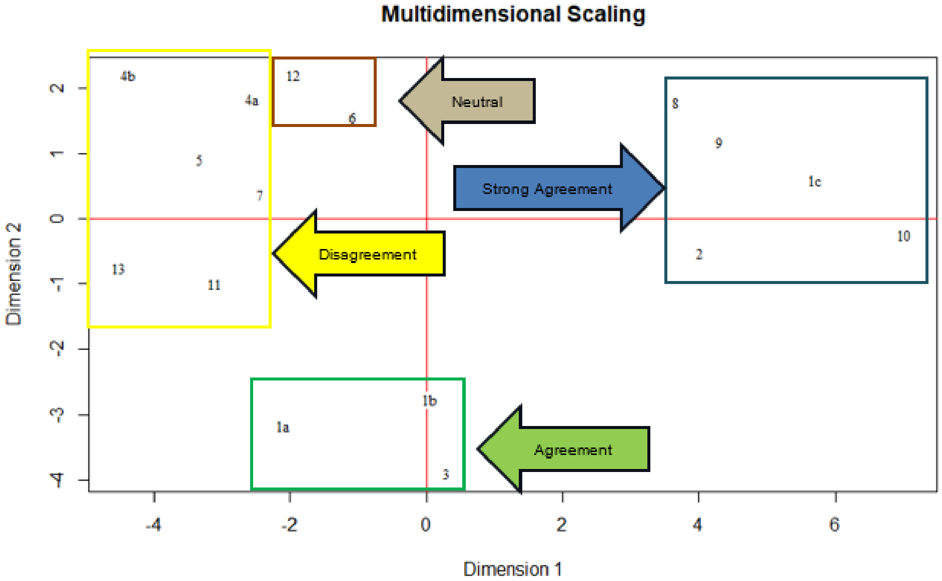

Figure 11 – Control Group Multidimensional analysis [Author, 2024]

The results indicated an agreement of the control group that the standard THA presents characteristics of making the boundaries identifiable, but mostly when the aviation regulation provides guidance (#1c). Likewise, defining safety limits for flights, which is a safety measure widely taught by the schools, is a common strategy used by all the community. However, when boundaries are associated with individual and organizational pressures (#1a, #1b), the level of agreement was lower indicating that the process is not as clear as the flight boundaries. This rationale aligns with the lack of standardization across the regulatory agencies and industry when controlling organizational and individual pressures [26]. #2 talks about test limits identification which is a process that is not normally clear in the THA, but it is commonly present in the Test Plan. These two processes are intricately linked, and the test envelope limits are established during the planning phase, describing the flight conditions that the crew is authorized to fly. However, no clear update mechanism in this process is provided and no specific limits for each hazard is stimulated. The management to counteract the pressures (#3) is not a strong agreement, since different organizations have different processes associated at this level. The standard THA process only asks to list mitigations to the hazard causes on the management phase but does not clearly define the process to implement and control the mitigation efficiency.

[12] states that an activity needs to be seen as a system. The disagreement opinion about #4a indicates that the SMEs do not clearly understand this systemic view. [27] say that a complex system is composed of diverse groups of humans, technologies, and organizations in which the central characteristic is its non-linearities and feedback. The interactions between the different components are dynamic and difficult to predict. The standard linear interaction is important to understand some cause-effect chains, but it is important to assume that these linear strings interact with each other and can have different influences on the overall risk. [28] proposes the Functional Resonance Analysis Method (FRAM) to model complex systems, since this tool aims to model the complexity of the system. The bibliographic research made by [27] about the application of this method in aviation points some works that used FRAM with indexation to quantify the variability, but the results still show inconsistence making this line of research a promising field for future studies. #4a, #4b responses represent the lack of ability to understand and measure non-linear interactions, therefore neglecting a systemic view. # 4b talks about how the system is able to sustain a constant monitoring of its performance by recognizing brittleness, adapt the limits or boundaries to this new status and self-prepare to absorb these new variations. For the standard THA case, it is almost consensus that the process does not routinely reassess effectiveness of process, limits, or mitigations during the WD phase of the flight test activity. Without this systemic view, the Standard THA is not clearly able to acknowledge neither interactions between hazards nor routine reassessments and updating which indicates lack of ability in being adaptive.

[11] suggests a portfolio where safety indicators should be implemented to increase awareness of the manager. He classifies and stimulates the usage of well-balanced pairs of safety indicators: proactive-reactive and economic-safety. The standard THAs philosophy addresses management by the elimination or reduction of the probability/severity but does not clearly concern about system' status quo which makes managers lean towards reactive indicators. Additionally, no procedure in the standard THA obligates the user to set KIO, indicator, or margins to the hazards. The disadvantage of this approach is that it is not possible to have a visualization of how safe or effective the mitigations and limits (#5) are, especially during unexpected outcomes (#6).

The linear cause-effect rationale brings a safety paradox perspective into play, as discussed by [29]. He says that the presence of barriers or mitigations undoubtedly makes the system safer through redundance, automation, etc. but also impinges more complexity on it, which increases its opacity for those who manage and control it. When the standard THA process asks to list the hazards, it is looking for undesirable outcomes that the stakeholders experienced before in their careers. Nevertheless, the crew would have a visualization of causes and consequences related to this hazard. The next step is to find out ways to avoid hazard occurrence by implementing barriers or some way to break the causality thus reducing hazard probability or severity. This exercise brings the sensation that the process can predict the expected behavior by drawing several causality strings that could occur. This is true to some extent since it may stimulate the crew to build knowledge about each case. However, the process of setting KIO criteria and executing a buildup approach are the real players of the prediction capacity. The standard THA process is not clear about how large the buildup approach steps should be and how much margin is necessary to set the KIO criteria. Ideally, these two subprocesses should build up the limits of known expected behavior and as the data collected is processed a next step can be achieved. Statements #8 and 9 responses indicate that control group strongly agree that the standard THA is able to predict system behavior and set limits, which is probably influenced by the common use of KIO and build up approach strategies. #10 response, on the other hand, is very clear to the SME since the THA has a step specifically dedicated to this action which is illustrated in Figure 1 as step 8 or as corrective actions/recovery procedures in Figure 9.

#11, 12 and 13 talk about time pressure, learning and updating. These three subprocesses are not clearly stated on the process and it is very dependent on the processes attached by each organization to the THAs. Again, the lack of standardization on the SPV control strategies mentioned by [26] indicates that an improvement in the processes could be explored by the managers. [13] mentioned that time pressure could lead to lack of foresight which could have a significant effect on prediction capacity and flight preparation. Most of the time, this variable is not mentioned in the THA process, since it normally falls in an organizational category, but it could be easily listed as a hazard cause. Thus, by using indicators attached to the THA process, the effects of time pressure could be indirectly sensed and managed.

The learning and updating is again very dependent on organizational rules and processes. Since the THA can predict some scenarios, constant learning from small deviations from “Work as imagined” could be used to update the process, but one more time, the process does not have a specific mechanism for learning and updating subprocesses.

4.4 Experiment group results

The aggregated experiment group results are shown in Table 5 and Table 6.

Table 5 – Experiment group raw results aggregation

Question #	1a	1b	1c	2	3	4a	4b	5	6	7	8	9	10	11	12	13
Strongly Disagree & Disagree	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Neutral	20%	0%	0%	20%	0%	40%	20%	60%	0%	20%	0%	20%	0%	0%	0%	0%
Agree & Strongly Agree	80%	100%	100%	80%	100%	60%	80%	40%	100%	80%	100%	80%	100%	100%	100%	100%
Mode	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4
Average	4	4	4	4	4	4	4	3	4	4	4	4	4	4	4	4
STD DV	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0

Table 6 – Experiment group academic author’s association

Question aggregation	Rasmussen and Svedung	Woods	Hollnagel, Woods and Levenson
Strong Agreement	1a, 1b, 1c, 2, 3	4a, 4b, 6, 7	8, 9, 10, 11, 12, 13
Agreement		-	-
Neutral	-	5	-
Disagreement	-	-	
Strong Disagreement	-	-	-

Observing Rassmussen and Svedung column in Table 6, the level of agreement increased. #1a, 1b and 3 results jumped up to the Strong Agreement level when compared to Table 4. This effect on #1a and #1b could be associated to the organizational management tools exposed in Figure 7 and introduced in the design phase such as GO-NO-GO and KIO criterion not only for flights but also for Staff and Management. The timely, realistic, and cross communication feedback loops contributed to enhance learning and damp time and organizational pressures. The individual pressures were controlled using the Organizational Resource Risk (ORM) matrix which signalized to the crew when the individual risk was above normal triggering feedback to the management system. #1c (100% of agreement) and 2 had no disagreement which could be associated to the indicator implementation step in the adapted THA. This addition contributed to having more control over mapped hazard which may have increased the likelihood of avoiding unpredictable outcomes by sensing the side effects of non-mapped cause-effect lines. The combination of CH, landing gear structural load sensor and environmental indicators was an example of this strategy. Finally, #3 was a consensus, which could be influenced by the procedural usage of KIO criteria during flight test and on project actions.

Most results on Woods column (Table 6) had a strong agreement which were evidently disparate when comparing to the control group. The presence of a timely feedback loop (Figure 7) through the daily meetings with the entire team contributed to maintain the system in a sustainable paste. In addition, the daily reports in between subphases were useful to update resources necessary to proceed to the next phase across organization levels. These feedback loops associated to the indicator implementation process and the KIO criteria decision process contributed to increase the learning and updating rate, therefore explaining #4a, 4b and 7 strong agreements regarding adaptative capacity across the organization. #6 responses brought a strong agreement which was a result of a consistent implementation of indicators across the hazards. The association of KIO to each indicator based on the knowledge and competence improved safety margins monitoring, allowing better limits and boundaries management, therefore improving awareness regarding SPV status and the possibility to sense closeness to unexpected outcomes. #5 had a neutral tendency but with no disagreement, which could be associated to the adaptation to the new method and the understanding about what mechanisms would control SPV which in this case were again the combination of learning and updating feedback loops across the organization, implementation of procedural hazard-based indicators and KIO criteria setting based on knowledge to control safety margins. A real case could be introduced here when even using the most aggressive pilot and using the most disadvantageous aircraft configuration, the flight test program was successful with all indicators within the designed margins indicating system robustness and sustained adaptability.

#8 and 10 had no disagreement or neutrality showing assuredness of what to expect and what to do. Only one subject returned neutrality regarding #9. Again, the procedural learning process established by the adapted THA checklist, feedback loops and KIO criteria led to identification of lack of training and data (Figure 8), which contributed to enhance prediction competence on these 3 dimensions.

11, 12 and 13 also have a strong agreement on Hollnagel, Woods and Leveson questions. This was a result of the constant learning exercise provided by the judgment subprocess during the execution of the flights followed by the learning process occurred during the daily meetings. The time pressure was damped and controlled part by the designing phase where the division of the operation in subphases and the GO-NO-GO project and flight criterion definition allowed an alignment of expectations across the organization. In complement, the processing of daily reports in between subphases enabled an acceptable update of results, resources, safety margins and risk

assessment not only to Staff level, but also to all levels above. It is also important to mention that the Organizational pressures and Scenario pressures were indirectly monitored by the KIO criteria installed during the design phase. The adapted THA process forced the crew to generate indicators in order to set KIO criteria (Figure 6). This practice consequently influenced the project level to use and create indicators to set KIO and GO-NO-GO criteria on that level. Nevertheless, the use of indicators as long as KIO criterion process (Figure 7) decisively contributed to monitoring SPV providing sufficient learning capacity from everyday activity to anticipate and dump unexpected system outcomes before they became unmanageable (Figure 8).

Rationalizing the results from control and experiment group (Figure 12), while the control group had 8 questions on the agreement dimension (5-Strong Agreement and 3–Agreement) and 8 on the disagreement dimension (8-Disagreement), the experiment group had 15 questions on the agreement dimension (15-Strong Agreement), 1 on the neutral dimension and none on the disagreement dimension. As a result, the experiment group agreed on 94 % of the questions while the control group agreed to 50% of the questions, a difference of 44%.

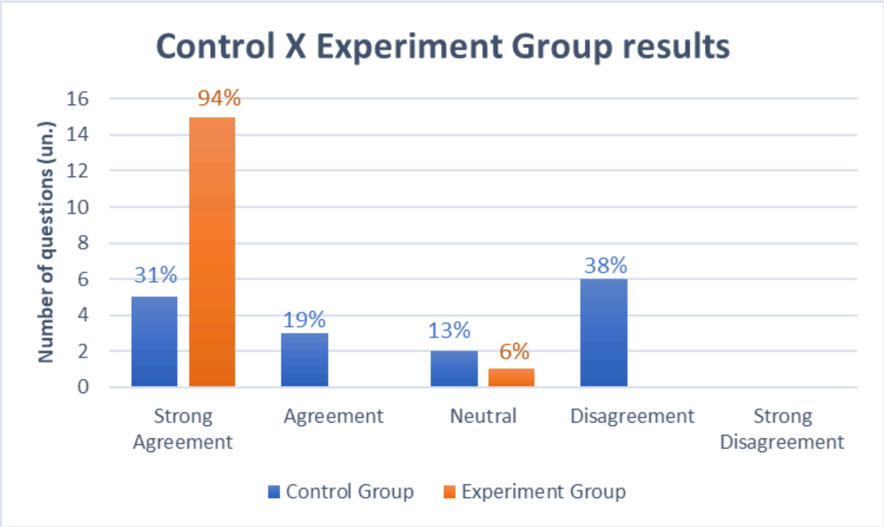

Figure 12 – Control versus Experiment group results [Author, 2024]

Following these results, and according to the option of flight test SMEs it was evident that the adapted THA process was able to better control SPV in flight test systems when compared to the standard THA process.

5. Limitations and conclusions

This article conclusions of this research were stated considering the following limitations: All technical documentation regarding the flight test operation are classified documents, therefore quantitative information about operation did not have open disclosure status. This article was authorized to expose only aspects related to the safety management process or information that was already disclosed in media channels. All information used to characterize the standard THA process was based on documents with non-classified disclosure. Internal classified organizational subprocesses were not considered in the analysis. Control group sample size was limited by the volunteer willing of the flight test community. 78.6% of the control group graduated from EFEV, which could have tailored the results towards the standardization provided by this institution. Experiment group sample was limited to only one flight test operation, envelope expansion, and 5 subjects.

The standard THA process is a linear cause-effect risk management tool based on preexistence

SME experiences which lean the process towards a reactive management [4] [2]. Aviation industry and organizations have subprocess attached to standard THAs indicating some level SPV control by using KIO and test limits techniques. However, they are neither standardized across the organizations, nor have a procedure of when or where they should be implemented [26]. The control group agreement on only 40% of the questions indicated that the standard THA does not completely and proactively manage SPV whereas the adapted THA process was able to connect its outputs to other organizational levels through communication feedback loops organized by daily meetings and FTDRs (Figure 8). The adapted THA process implemented hazard-based indicators that incorporate KIO criteria supported by a continuous learning from all work events. As a result, the control of safety margins proactively sensed closeness to SPV limits before an unexpected catastrophic event, thus managing SPV (Figure 7). Learning from the indicator data that were timely updated to the pertinent levels of the organization, proactively and timely triggered modifications in the system to bridge knowledge, competence, training, and resources gaps which prevented catastrophic outcomes during the operation showing an evident shift on the effective capacity of SPV management provided by the adapted THA. And according to flight test SMEs, the adapted THA process is 44% more capable to account for SPV in flight test systems since it kept the flight test team proactively safe during flight envelope expansion operation.

This work can also suggest some additional learning taken during the research process, as follows: Although all aviation regulatory agencies enforce the usage of THAs, they are not regular in describing the process for active control of system performance, some talk about mitigation implementation, while others focus on information feedback. Additionally, they are not consistent in describing how either KIO or communication feedback (Daily meetings and FTDRs) should occur during flight test execution phase. The adapted THAs contributed by offering a validated SME option to the community providing subprocess to incorporate these strategies. According to the SMEs, the adapted THA process allowed the procedural usage of epistemic-base data in risk science [30] through the usage of KIO and communication feedback tools. Safety I and Safety II are complimentary, the adapted THA provided a path to change the risk management human behavior where not only looking to hazards that happened is important but also the control the SPV on all organizational levels through feedback and indicators has a contribution to enhance safety. Moreover, all organizational levels had a better view and standard understanding of the risk management process contributing to easier communication to justify investments in training. And finally, the questionnaire construction technique used to transcribe the academic to the flight test community language facilitated the communication in between the academia and the target public being evaluated as useful technique to enhance the reliability of this type of data.

Future work is suggested by applying this work in their flight test operation and using other strategies of control tools and indicators.

6. Copyright Statement

The authors confirm that they, and/or their company or organization, hold copyright on all of the original material included in this paper. The authors also confirm that they have obtained permission, from the copyright holder of any third party material included in this paper, to publish it as part of their paper. The authors confirm that they give permission, or have obtained permission from the copyright holder of this paper, for the publication and distribution of this paper as part of the ICAS proceedings or as individual off-prints from the proceedings.

References

[1] North Atlantic Treat Organization, *AGARDograph 300: Introduction to flight test engineering,* vol.

14, Washington: NATO, 2005, pp. 1-3.
[2] H. J. Pasman, W. J. Rogers e M. S. Mannan, “How can we improve process hazard identification? What can accident investigation methods contribute and what other recent developments? A brief historical survey and a sketch of how to advance.,” *Journal of Loss Prevention in the Process Industries,* vol. 55, pp. 80-106, 2018.
[3] J. Reason, Managing the risks of organizational accidents, vol. 1, Ashgate, 1997, p. 251.
[4] A. Rujiter e F. T. Guldenmund, “The bowtie method: a review,” *Safety Science,* nº 88, pp. 211-218, 2016.
[5] T. Aven, “Risk assessment and risk management: Review of recent advances on their foundation,” *European Journal of Operational Research,* nº 253, pp. 1-13, 2016.
[6] E. Hollnagel, Safety–I and Safety–II: the past and future of safety management, Ashgate, 2014.
[7] E. Hollnagel, The ETTO principle: why things that go right sometimes go wrong, Farnham, UK: Ashgate, 2009.
[8] E. Hollnagel, “Resilience engineering: a new understanding of safety.,” *Journal of Ergonomics Society of Korea,* vol. 35, pp. 185-191, 2016.
[9] J. Rasmussen, “Risk management in a dynamic society: a modelling problem,” *Safety Science,* vol. 27, pp. 183-213, 1997.
[10] J. Rasmussen e I. Svedung, Proactive risk management in a dynamic society., Karlstad: Swedish Rescue Service Agency, 2000.
[11] D. D. Woods, M. Branlat, I. Herrera e R. Wolter, “Where is the organization looking in order to be proactive about safety? A Framework for revealing whether it is mostly looking back, also looking forward or simply looking away,” *Journal of contingencies and crisis management,* vol. 23, nº 2, pp. 97-105, 2015.
[12] D. D. Woods, “Four concepts for resilience and the implications for the future of resilience engineering,” *Reliability engineering and system safety,* nº 141, pp. 5-9, 2015.
[13] E. Hollnagel, D. D. Woods e N. Levenson, Resilience engineering: concepts and precepts, Hampshire: Ashgate, 2006, p. 410.
[14] E. Hollnagel, “Introduction to the resilience analysis grid (RAG),” Denmark, 2015.
[15] Federal Aviation Administration, *AC-25-7D: Flight test guide for certification of transport category airplanes,* Washington: FAA, 2018.
[16] E. Hollnagel e D. D. Woods, Joint cognitive systems: foundations of cognitive systems engineering, Boca Raton: Taylor and Francis/CRC Press, 2005.
[17] G. E. Cooper e R. P. H. jr, “AGARD report 567: the use of pilot rating in evaluation of aircraft handling qualities,” NATO, USA, 1969.
[18] T. A. Saurin e A. J. B. Werle, “A framework for the analysis of slack in socio-technical systems,” *Reliability engineering and system safety,* vol. 167, pp. 439-451, 2017.
[19] M. S. Sanders e E. J. McCornmick, Human factors in engineering and design, New York: McGraw-Hill, 1993.
[20] S. M. Charlston e T. G. O’Brien, Handbook of human factors testing and evaluation, New Jersey: Lawrence Erlbaum Associates, 2008.
[21] J. Shateau, J. W. David, P. T. Rickey e C. P. Julia, “Performance-based assessment of expertise: how to decide if someone is an expert or not,” *European Journal of Operational Research,* vol. 136, pp. 253-263, 2002.
[22] Federal Aviation Administration, “ORDER 4040.26C: aircraft certification service flight test risk management program,” FAA, Washington, 2021.
[23] T. Hastie, R. Tibshirani e J. Friedman, The elements of statistical learning, New York: Springer, 2017.
[24] Society of Flight Test Engineers, “Member Map - Society of Flight Test Engineers,” SETP, 2023. [Online]. Available: sfte.org. [Acesso em 01 Aug 2023].
[25] Society of Experimental Test Pilots, “SETP members statistics,” SETP, [Online]. Available: setp.org. [Acesso em 01 Aug 2023].

[26] North Atlantic Treat Organization, *AGARDograph 300: flight test safety and risk management,* Washington: NATO, 2021.

[27] W. Tian e C. Caponecchia, “Using the functional resonance analysis method (FRAM) in aviation safety: a systematic review,” *Journal of advanced transportation,* vol. 2020, p. 14, 2020.

[28] E. Hollnagel, FRAM: functional resonance analysis method., London: CRS Press, 2012.

[29] J. Reason, “Safety paradoxes and safety culture,” *Injured control and safety promotion,* vol. 7, nº I, pp. 3-14, 2000.

[30] S. O. Hansson, S. O. e T. Aven, “Is risk analysis scientific?,” *Risk Analysis Journal,* vol. 34, nº 7, pp. 1173-1183, 2014.