RBAC&CiA
Revista Brasileira de Aviação Civil
& Ciências Aeronáuticas
Artigos

SUBJECTIVITY CHALLENGES IN EVALUATING OBSERVABLE BEHAVIORS IN CBTA

Felipe Gontijo Guimarães [1]
Rodrigo Luiz de Souza Rodrigues [2]

ABSTRACT

This study investigates the challenges of subjectivity in the evaluation of observable behaviors (OBs) within the Competency-Based Training and Assessment (CBTA) framework applied to civil aviation. The assessment of competencies through OBs is a fundamental pillar of CBTA, allowing the measurement of pilots' knowledge, skills, and attitudes in real or simulated operational scenarios. However, subjectivity in evaluators' judgments can compromise the standardization and reliability of results, leading to interpretative variations that affect crew training and certification. This research analyzes the main factors influencing this variability, including evaluators' experience, clarity of assessment criteria, and task complexity. Additionally, different competency assessment models and their limitations in mitigating subjectivity are discussed. Finally, a new theoretical framework is proposed to enhance OB evaluation, reducing interpretative discrepancies among evaluators and ensuring greater consistency in the assessment process.

Keywords: CBTA; Observable Behaviors; Competency Assessment; Subjectivity; Operational Safety.

[1] Bacharel em Ciências Aeronáuticas pela Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS). Pós-graduado em Gestão, Empreendedorismo e Desenvolvimento de Negócios pela PUCRS. Primeiro Oficial na GOL Linhas Aéreas há 5 anos, atuando na frota Boeing 737 NG (-700, -800 e MAX), com mais de 3.100 horas de voo, incluindo 2.500 horas em aeronaves à reação. Foi instrutor de Voo (INVA). Possui diversos treinamentos e certificações, incluindo CBTA pela ANAC, CRM, LOFT, EC-PREV e AVSEC. ORCID: https://orcid.org/0009-0000-1887-2083 E-mail: felipe.fgg@hotmail.com

[2] Bacharel em Ciências Aeronáuticas pela Universidade Veiga de Almeida (UVA). Mestre em Engenharia Aeronáutica pelo Instituto Tecnológico de Aeronáutica (ITA). Auditor Líder (IRCA) ISO 9001. Facilitador de CRM pela ANAC. Primeiro Oficial na GOL Linhas Aéreas há menos de 1 ano, atuando na frota Boeing 737 NG (-700, -800 e MAX), com mais de 650 horas de voo. Foi instrutor de Voo (INVA). Possui experiência em operações de Aerolevantamento Geofísico, Manutenção de Aeronaves e Táxi Aéreo. ORCID: https://orcid.org/0009-0009-1013-3310 E-mail: fay.rodrigues@gmail.com

DESAFIOS DA SUBJETIVIDADE NA AVALIAÇÃO DE COMPORTAMENTOS OBSERVÁVEIS NO CBTA

RESUMO

Este estudo investiga os desafios da subjetividade na avaliação de comportamentos observáveis (OBs) no Treinamento e Avaliação Baseados em Competências (CBTA) aplicado à aviação civil. A avaliação de competências por meio de OBs é um dos principais pilares do CBTA, permitindo medir conhecimentos, habilidades e atitudes dos pilotos em cenários operacionais reais ou simulados. No entanto, a subjetividade no julgamento dos avaliadores pode comprometer a padronização e a confiabilidade dos resultados, gerando variações interpretativas que impactam a formação e certificação dos tripulantes. A pesquisa analisa os principais fatores que influenciam essa variabilidade, incluindo a experiência dos avaliadores, a clareza dos critérios avaliativos e a complexidade das tarefas observadas. Além disso, são discutidos diferentes modelos de medição de competências e suas limitações na mitigação da subjetividade. Por fim, propõe-se um novo referencial teórico para aprimorar a avaliação dos OBs, reduzindo as diferenças interpretativas entre avaliadores e garantindo maior consistência no processo avaliativo.

Palavras-chave: *CBTA; Comportamentos Observáveis; Avaliação de Competências; Subjetividade; Segurança Operacional.*

1 INTRODUCTION

Commercial aviation is widely recognized as one of the safest and most technologically advanced modes of transportation (Borucka; Romele, 2024), the result of decades of continuous efforts in regulation, training, and the development of safety systems. However, challenges remain, especially regarding human factors, which continue to play a critical role in Operational Safety Events (OSEs), such as aviation incidents and accidents. Between 2019 and 2023, investigations conducted by the European Aviation Safety Agency (EASA, 2024, p. 23) indicated that crew errors or confusion during the management of adverse conditions, such as procedural errors or entering severe meteorological conditions, were predominant causes of fatal accidents. Additionally, issues such as loss of control in flight (LOC-I), runway excursions, and ground collisions continue to be the most common outcomes of these events, reinforcing the need for continuous improvement in crew training and evaluation (EASA, 2024, p. 24).

The advancement of automation and the integration of systems in commercial aircraft cockpits have highlighted the limitations of traditional training methods, which were based on the repetition of specific situations, making them less effective in preparing pilots to handle unexpected events (Sun *et al.*, 2023b, p. 345).

In this context, CBTA emerges as an innovative approach designed to address the growing complexity of aircraft operations and to align training with the operational demands of modern aviation (ICAO, 2020, p. I-2-1). This approach focuses on the development and evaluation of critical competencies through the analysis of observable behaviors (OBs), which serve as parameters to measure knowledge, skills, and attitudes in real or simulated situations (Sun *et al.*, 2023b, p. 346). However, as indicated by IATA (2025), OB-based evaluation is not the only criterion adopted in training programs.

In addition to observing and classifying OBs, other parameters are used to ensure a comprehensive evaluation of pilot performance. Among these, the analysis of the number of OBs demonstrated (HOW MANY) allows quantification of the manifestation of competence in operational situations, and the frequency with which these behaviors are presented (HOW OFTEN) provides information about performance consistency over time (IATA, 2025, p. 18). The effectiveness of the behavior in the context of Threat and Error Management (OUTCOME of TEM) is also considered, assessing whether the pilot's actions result in the adequate mitigation of operational risks (IATA, 2025, p. 19).

Furthermore, the evaluation of pilots in CBTA may include both formative and summative assessments, which allow continuous monitoring of pilot development and verification of meeting the required standards (IATA, 2025, p. 13). The competency evaluation matrix provides a detailed framework to classify the level of pilot performance based on descriptive scales and contextual factors (IATA, 2025, p. 17). Additionally, specific evaluation guides for single- and multi-crew operations ensure that the training is adapted to the particularities of the operational environment (IATA, 2025, p. 34-50).

Although these complementary parameters provide a broader view of pilot development, OB-based evaluation remains the central element of CBTA. The detailed analysis of observable behaviors enables precise measurement of essential competencies, allowing identification not only of what the pilot did, but also how and why certain actions were executed. In this way, the integration of OBs with other metrics enhances the objectivity and reliability of evaluations, ensuring that competency certification is based on structured criteria aligned with the operational demands of modern aviation. In some cases, training curricula may use OBs as the sole parameters for competence evaluation (IATA, 2025, p. 29). CBTA has been progressively integrated into international regulations, promoting significant changes in commercial aviation training (ICAO, 2020, p. I-2-1).

Despite its advancements, the inherent subjectivity in the CBTA evaluation process remains a challenge. Evaluations frequently depend on human judgment, which may vary according to the experience and perception of evaluators, compromising the consistency and reliability of the results (Sun *et al.*, 2023b, p. 346).

These limitations become especially relevant in a sector where precision and objectivity are critical to operational safety. The lack of standardized criteria or more objective tools may lead to inconsistent evaluations, reducing the effectiveness of CBTA in its core objective of enhancing practical competencies and mitigating risks. Thus, the need to explore solutions that minimize the impacts of subjectivity in the judgment of OBs becomes a priority.

Based on this context, this article aims to investigate the factors that influence subjectivity in the evaluation of observable behaviors within the CBTA framework. It seeks to identify variables that directly affect evaluative judgment and to propose strategies to mitigate these effects, promoting greater standardization and reliability in the evaluation processes. This study also aims to contribute to the formulation of new benchmarks that can be incorporated into training programs, enhancing their effectiveness and alignment with sector needs.

This work is particularly relevant because CBTA is a relatively new type of training and lacks studies that can practically guide its fair and effective application.

This article is organized as follows: Section 2 presents the theoretical foundation of CBTA, addressing human factors in aviation, the evolution of training methods, and the importance of observable behaviors (OBs) in competency evaluation. In addition, the structural competencies of CBTA and the challenges associated with subjectivity in evaluation are discussed. Section 3 details the methodology adopted for the study, including a systematic literature review, analysis of evaluation models, and the proposal of a new theoretical framework to mitigate interpretative variability in the observation of OBs. In Section 4, the evaluation methods in CBTA are discussed, with an emphasis on structured observation, quantitative approaches, and the application of the Venn competency model. The factors that influence subjectivity in evaluation, as well as the limitations of existing models, are also analyzed. Finally, Section 5 presents practical examples of OB evaluation in different competencies, illustrating the applicability of the proposed theoretical framework and its implications for standardizing the evaluation process.

2 THEORETICAL REVIEW

Human factors are widely recognized as one of the main components of operational safety in aviation (Lázaro *et al.*, 2024). The interaction among pilots, automated systems, and the operational environment directly influences risk mitigation and decision-making under high cognitive load conditions (Yilmaz, 2024). According to Biede *et al.* (2023a), an evaluation system for aviation professionals' competencies has a significant impact on improving operational safety, as it contributes to the verification of both technical and non-technical skills that are essential for the effective management of adverse situations, such as in the prevention of Operational Safety Events (OSEs).

Technical skills refer to operational knowledge, the application of procedures, manual control, and the use of aircraft automation. In contrast, non-technical skills include competencies such as situational awareness, communication, leadership, workload management, and decision-making (IATA, 2021). The importance of non-technical skills has been widely recognized, as errors in these areas are frequently associated with aviation incidents and accidents (Biddle, 2023, p. 71). In this sense, adopting training approaches that encompass both categories is essential for the formation of professionals capable of operating in a highly complex and multifactorial environment.

However, conventional pilot training has historically been based on task repetition and the accumulation of flight hours – a model that, despite being effective for the development of motor and operational skills, presents limitations in preparing pilots for unexpected events (Biddle *et al.*, 2023). Evaluation in traditional training is often based on fixed criteria and subjective observation, which can result in inconsistencies in standardizing pilot performance (Biede *et al.*, 2023a). With the evolution of air operations and the growth of cockpit automation, the study of new, more dynamic and adaptive methods for pilot training and evaluation has become evident.

In this context, CBTA emerged as a response to the limitations of the traditional model, promoting an approach focused on the development of essential competencies for operational safety (IATA, 2021). Unlike conventional methods, CBTA emphasizes continuous evaluation through observable behaviors (OBs), allowing instructors to make more objective measurements of pilot performance (Biddle, 2023, p. 73). Moreover, this approach enables the personalization of training, adjusting it to individual needs and reducing the influence of subjective judgments in professional certification (IATA, 2021).

CBTA is structured around nine fundamental competencies that encompass both the technical and non-technical skills necessary for safe and efficient pilot performance. These competencies include: Knowledge Application (KNO), Procedure Application and Regulatory Compliance (PRO), Flight Path

Management with Automation (FPA), Flight Path Management with Manual Control (FPM), Communication (COM), Situational Awareness and Information Management (SAW), Leadership and Teamwork (LTW), Workload Management (WLM), and Problem Solving and Decision-Making (PSD) (ICAO, 2020, p. II-1-1-1). Each of these competencies must be evaluated through a set of OBs, enabling a detailed analysis of pilot performance in real or simulated operational scenarios (Biddle, 2023, p. 73).

More specifically, evaluation in CBTA occurs through the collection and analysis of evidence regarding pilot performance relative to established competency standards. During training and proficiency assessments, instructors and evaluators observe a selection of desired behaviors and classify them according to the defined performance criteria (IATA, 2021, p. 27). This process seeks to ensure that pilots demonstrate minimum proficiency in the required competencies and are not approved if any behavior is not observed. This principle can be verified in practice during simulator training. In an instrument approach under low-visibility conditions, a Boeing 737 pilot demonstrates situational awareness (SAW) by correctly monitoring the minimum altitude of the procedure and efficiently distributing tasks, ensuring effective workload management (WLM). However, if the pilot exceeds the maximum descent ratio and requires instructor correction by misapplying procedures (PRO), and upon reaching the minimum altitude does not decide to execute a go-around within the appropriate time – thus compromising decision-making (PSD) – the pilot will not be approved and will need reinforcement in both technical and decision-making aspects before re-assessment. This process ensures that only truly prepared pilots advance, reducing subjectivity in certification.

This situation highlights one of the main challenges of competency-based evaluation: subjectivity in the observation of expected behaviors (OBs). Although some behaviors are easily verifiable – such as correctly monitoring the minimum altitude (SAW) or clearly distributing tasks among crew members (WLM) – others may be interpreted more subjectively, depending on the evaluator's perception.

According to IATA (2021, pp. 33–36), in the case of procedure application (PRO), an objective OB would be to correctly follow the approach parameters, something that can be precisely measured through instruments and regulatory limits. However, decision-making (PSD) involves more subjective aspects, such as the perception of the need for a go-around within the appropriate time frame. One evaluator may consider that the pilot hesitated excessively, while another might interpret this hesitation as a valid attempt to confirm the minimums before aborting the approach.

Furthermore, competencies such as leadership and teamwork (LTW) or communication (COM) frequently involve OBs that are difficult to quantify. For example, a pilot may give clear and firm instructions to the co-pilot, but the evaluation of "clarity" can vary according to the instructor's perception. Similarly, situational awareness (SAW) can be inferred from gestures, glances, or interactions, elements that are not always recorded in a standardized manner.

Thus, although CBTA aims for greater objectivity in evaluation, the subjectivity in interpreting certain OBs can still generate inconsistencies in results, making it essential to develop more standardized and precise methods to minimize variability among evaluators. There is still no specific standardized model for evaluating observable behaviors (OBs). The subjectivity in interpreting these behaviors can lead to variations in evaluation, compromising the reliability and consistency of the results (IATA, 2021, p. 15). The absence of a consolidated model for OB evaluation represents a challenge for the full implementation of CBTA, making it necessary to investigate more objective and standardized methods to ensure the effectiveness of the evaluation system.

3 METHODOLOGY

This study adopts a qualitative methodological approach to investigate subjectivity in the evaluation of OBs within the framework of CBTA. The adopted methodology is based on references from the academic literature on CBTA and pilot performance evaluation models, with emphasis on methods proposed by Sun *et al.* (2023a, p. 346), as well as guidelines from the International Civil Aviation Organization (ICAO, 2020, p. I-2-1).

3.1 STUDY DESIGN

The research is developed in two main stages: (i) document analysis and systematic literature review on evaluation models in CBTA, and (ii) critical analysis of existing qualitative and quantitative methods for pilot evaluation. The document review included publications from regulatory bodies, such as ICAO (2020, p. II-1-1-1), and empirical studies on competency evaluation in aviation (Sun *et al.*, 2023a, p. 347).

3.2 QUALITATIVE AND QUANTITATIVE METHODS FOR PILOT EVALUATION

Qualitative methods in pilot evaluation often involve direct observation of performance during simulator training or actual operations. Instructors and evaluators use descriptive scales to classify performance based on predefined behavioral parameters (ICAO, 2020, p. II-1-1-1). However, this process is subject to the inherent subjectivity of the evaluator's perception, since the interpretation of an observable behavior (OB) may vary according to the evaluator's experience and personal bias (Sun *et al.*, 2023a, p. 348).

Quantitative methods, although not the focus of this study, are mentioned to contextualize existing approaches. Studies indicate that subjectivity can be minimized through structured evaluation protocols, such as the standardization of training scenarios and the use of detailed evaluation guides. Sun *et al.* (2023a, p. 349) propose a behavior-based model in which the observation of OBs is complemented by objective criteria extracted from operational parameters, such as altitude tolerance and trajectory variation rate. In addition, metrics such as attitude control indices and response time to unexpected events have been explored (ICAO, 2020, p. II-1-1-1).

3.3 DEVELOPMENT OF A NEW THEORETICAL FRAMEWORK

This study does not propose a quantitative model for pilot evaluation, but rather a new theoretical framework for analyzing OBs, while preserving the evaluator's right to subjectivity. The objective is to guide evaluators to consider whether the pilot has effectively utilized all available resources before classifying a behavior as observed or not. Since different evaluators may identify distinct behaviors in the same pilot and on the same mission, the theoretical framework seeks to minimize these variations by proposing an evaluation based on the effective use of the pilot's available resources.

In this way, this study contributes to the construction of more standardized guidelines that increase the reliability of CBTA training, ensuring that the evaluator's subjectivity is grounded in a more objective and verifiable criterion.

4 DEVELOPMENT

In this chapter, different methods for evaluating competencies in aviation are explored, including structured and quantitative approaches based on operational and physiological data. In addition, the application of competency measurement models, such as the VENN diagram model, and their limitations in capturing the complexity of pilot performance are discussed. The factors that influence subjectivity in the evaluation of observable behaviors (OBs), as well as the interpretative variations among evaluators, are analyzed. Finally, a new theoretical framework is proposed that seeks to reduce subjectivity, ensuring a more structured evaluation aligned with the best practices of CBTA.

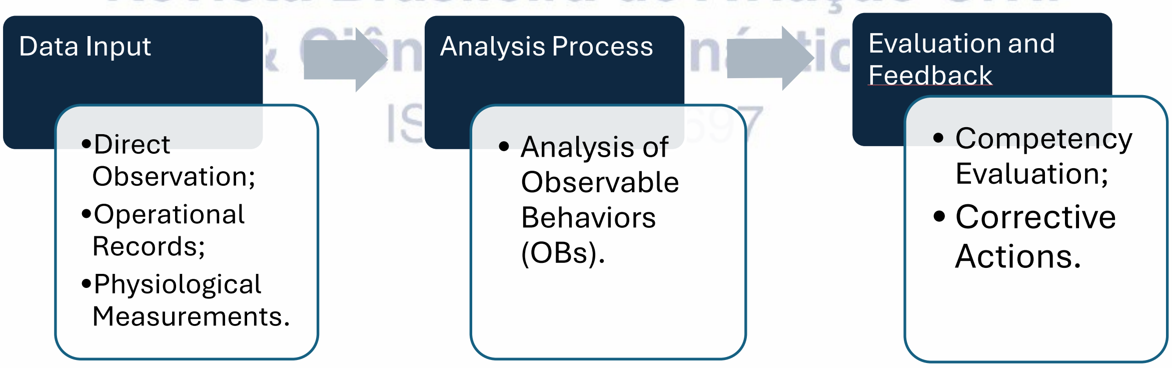
4.1 CBTA EVALUATION PROCESS

Evaluation in CBTA occurs in a structured manner, as illustrated in Figure 1, following guidelines established by ICAO (2020, p. II-1-1-1) and adopted by various training organizations. The process begins with the definition of specific operational scenarios, in which pilots are observed in simulated or real contexts. These scenarios are designed to assess both technical and non-technical competencies, considering aspects such as the application of procedures, flight path management, and decision-making (Sun *et al.*, 2023a, p. 346).

Data collection for the evaluation of OBs is carried out through direct observation by instructors, the recording of flight parameters, and, in some experimental cases, physiological measurements of pilots during task execution (Sun *et al.*, 2023a, p. 348). The evaluation is based on predefined behavioral indicators that help standardize evaluators' judgment and ensure greater objectivity in the process.

Pilot scoring is assigned based on the observed performance, with the common use of descriptive scales that classify the level of proficiency in each evaluated competency. The validation of the evaluation is done by comparing it with previously established standards, allowing the identification of performance gaps and directing corrective actions in training.

Figure 1 - Conceptual Diagram of the CBTA Evaluation Process



Source: Prepared by the authors.

4.2 QUANTITATIVE METHODS IN PILOT COMPETENCY EVALUATION

A Pilot competency evaluation can be carried out using quantitative methods that employ operational records – whether from actual flights or simulators – physiological data, and advanced statistical techniques to measure performance. These methods offer an objective approach to analyzing pilot performance, minimizing the influence of evaluator subjectivity and ensuring greater accuracy in identifying behavioral and operational patterns. Among the main quantitative approaches are the analysis of flight data via the Quick Access Recorder (QAR), the monitoring of physiological parameters, and the application of Principal Component Analysis (PCA) to correlate competencies.

The analysis of flight data obtained from the QAR has been used to quantify the quality of pilot operations. Researchers have developed methods that utilize QAR data to correlate flight parameters with established operational standards, allowing for objective performance evaluations (Sun *et al.*, 2023a, p. 16). However, the variability in pilot skills and the complexity of training scenarios hinder standardized data analysis, necessitating the development of more robust models for interpreting these parameters (Sun *et al.*, 2023a, p. 17).

The inclusion of physiological parameter analysis in pilot evaluation has been another approach in aviation research. Studies demonstrate that factors such as heart rate and heart rate variability can indicate levels of cognitive load and stress during aircraft operation (Sun *et al.*, 2023a, p. 19). Combined with flight data, these insights can provide a more complete picture of the pilot's operational competence, enabling more accurate, evidence-based evaluations (Sun *et al.*, 2023a, p. 20).

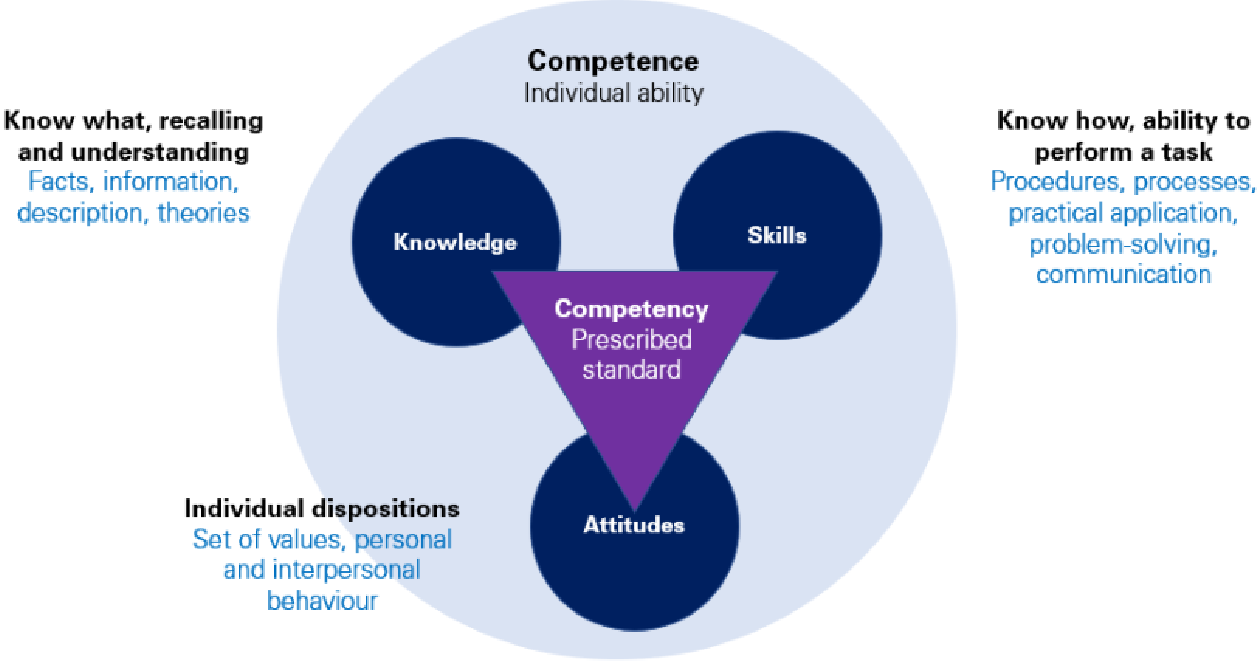
Competency analysis through the technique of Principal Component Analysis (PCA) has been employed to map the relationship among different essential pilot skills. This approach allows for the identification of competency patterns and gaps, guiding more effective training aligned with the specific needs of pilots (Sun *et al.*, 2023a, p. 15). Mansikka, Harris, and Virtanen (2017, p. 79) applied PCA to analyze pilot performance data obtained during Operator Proficiency Checks (OPCs), identifying four main components that structure essential competencies in an Input-Process-Output (IPO) model. This model demonstrated that certain competencies, such as Situational Awareness (SAW) and Problem Solving and Decision Making (PSD), play a critical role during the initial phase of task execution, directly influencing Workload Management (WLM) and Communication (COM), which in turn impact final operational actions, such as

Flight Path Management (FPM/FPA). However, despite PCA's potential in competency evaluation, challenges remain in defining objective criteria for interpreting results, as the interdependence among different competencies can complicate the extraction of unequivocal patterns (Sun *et al.*, 2023a, p. 22). Furthermore, because PCA depends on the availability of a large volume of data to establish robust statistical correlations, its application in the continuous evaluation of pilots may be limited by the variability in operational conditions and by the evaluation methods adopted by different instructors (MANSIKKA; HARRIS; VIRTANEN, 2017, p. 80). Thus, while PCA provides a type of quantitative framework, its use as the sole approach to competency evaluation in CBTA may not fully capture the qualitative aspects of pilot performance, requiring supplementation with structured observational methods.

4.3 CBTA MEASUREMENT MODELS AND THEIR LIMITATIONS

Within the CBTA framework, pilot competency evaluation is structured around different models, aiming to capture both the technical and behavioral aspects of in-flight performance. Among the main models used are the traditional checklist-based model – with direct observation by instructors – and the continuous competency-based evaluation model, as well as more integrated approaches such as the VENN competency model, as shown in Figure 2, which is embedded in the CBTA structure. Each of these methods presents advantages and limitations, particularly regarding the objectivity of the evaluation and the influence of evaluator subjectivity.

Figure 2 – Simplified VENN Diagram for CBTA Competency



Source: Biede *et al*., 2023b.

The VENN competency model proposes that the evaluation of pilot competencies be carried out based on the intersection of three fundamental elements: knowledge, skills, and attitudes. This approach assumes that competent performance depends not only on technical-operational mastery but also on the pilot's ability to effectively integrate these three aspects in a dynamic and highly complex environment (BIDDLE, 2023, p. 74). In the aviation context, Colonese (2022, p. 45) emphasizes that the application of this model should be structured by considering three main dimensions: the number of observable behaviors (OBs) demonstrated by the pilot, the frequency with which these OBs are exhibited when required, and the impact of the specific competency in the context of Threat and Error Management (TEM).

The first dimension, the number of OBs demonstrated, allows evaluation of whether the pilot's competency is consistently applied during operational tasks (COLONESE, 2022, p. 46). The second dimension, the frequency of OBs, verifies whether the pilot can repeatedly reproduce the expected behavior across different operational contexts. The third dimension, the impact of the competency within

TEM, assesses whether the behavior presented by the pilot serves as an effective barrier against operational threats and errors (COLONESE, 2022, p. 47). These three dimensions enable a more structured and objective evaluation, reducing interpretative variations among evaluators and promoting greater alignment between training and operational safety.

For example, a pilot may demonstrate theoretical knowledge of approach procedures in adverse meteorological conditions, but if they lack the practical skill to apply them correctly or the appropriate attitude to manage situational stress, their overall competency will be compromised. Thus, the intersection of these three factors is essential to ensure that competency evaluation is not limited to isolated aspects, but rather to the integrated performance of the pilot (BIEDE *et al.*, 2023a, p. 100).

The main limitation of the VENN competency model in CBTA is that it does not necessarily address the subjectivity in the observation of OBs, as it depends on the evaluator's interpretation of how the three elements interact in the pilot's performance. Moreover, the model can be challenging to implement in operational evaluations, as the objective measurement of a pilot's attitude, for example, can be complex and subjective (ICAO, 2020, p. II-1-1-1). However, Colonese (2022, p. 48) proposes that combining the three aforementioned dimensions can minimize these limitations by providing a more standardized criterion for competency certification.

The literature points to different models for pilot performance evaluation, with the main ones being the traditional checklist-based model and the continuous competency-based evaluation model (BIDDLE, 2023, p. 73). The traditional model presents limitations because it relies on isolated observation of behaviors and often results in a binary evaluation approach (presence or absence of the behavior). On the other hand, the competency-based model, widely adopted in CBTA, allows for a more detailed performance analysis but still suffers from the influence of evaluator subjectivity (SUN *et al.*, 2023a, p. 350). Biede *et al.* (2023a, p. 99) highlight that the adoption of technological tools and advanced statistical methods can complement human observation and reduce variations in judgments. Methods such as the use of operational data analysis and digital performance records have been considered to increase evaluation reliability (ICAO, 2020, p. II-1-1-1).

4.4 FACTORS INFLUENCING VARIABILITY IN THE EVALUATION OF OBSERVABLE BEHAVIORS (OBs)

The subjectivity in the evaluation of OBs, as detailed in Table 1, can be attributed to a number of factors, including evaluator experience, clarity of evaluation criteria, and the complexity of the observed task (Sun *et al.*, 2023a, p. 347). Studies indicate that different evaluators may interpret the same behavior in a pilot in different ways, resulting in discrepancies in evaluation outcomes (ICAO, 2020, p. II-1-1-1).

Chart 1 - Factors Influencing Variability in the Evaluation of OBs

Factor of Variability	Description	Impact on Evaluation	Mitigation Strategy (Practical Innovative Application)
Evaluator Experience	Differences in training level and practical experience among evaluators.	Divergent interpretations may compromise consistency in evaluations.	Conduct regular calibration sessions using simulation video reviews, fostering group discussions to harmonize evaluations and provide mutual feedback.
Clarity of Criteria	Clear communication of evaluation parameters and exemplification of expected observable behaviors (OBs).	Ambiguous criteria can lead to subjective assessments even with checklists.	Develop interactive digital modules featuring practical examples and demonstration videos, enabling evaluators to practice applying the criteria in varied scenarios, with real-time feedback.

Task Complexity	Variety and difficulty level of the operational scenarios being evaluated.	Multifaceted scenarios can lead to inconsistent interpretation of OBs.	Implement a dynamic simulation tool that adjusts scenario complexity and allows for performance comparisons across different levels, with statistical analyses to identify patterns.
Use of Available Resources by the Pilot	Verification of effective utilization of all technical, cognitive, and interpersonal resources available to the pilot.	Inadequate detailed verification can lead to incomplete assessments of competence.	Integrate the use of digital records (such as logs and video annotations) to map the pilot's sequence of actions, allowing for comparison against a validated reference performance and identification of critical deviations.

Source: Prepared by the authors.

Additionally, the absence of standardized guidelines for the evaluation of OBs can contribute to inconsistent judgments. Sun *et al.* (2023a, p. 349) suggest that the use of a structured framework based on behavioral indicators and objective evidence can reduce variability and increase the precision of evaluations.

4.5 PROPOSAL OF A NEW THEORETICAL FRAMEWORK FOR OB EVALUATION

The proposal of this study aims to develop a theoretical framework for the evaluation of OBs that preserves the evaluator's role while introducing criteria to reduce subjectivity in judgment. Unlike approaches based on QAR flight data, physiological parameters, or competency analysis using PCA, the new framework proposed by the authors focuses on the qualitative evaluation of the effective use of all technical, cognitive, and interpersonal resources available to the pilot before classifying a behavior as observed or not.

In this way, it is proposed that evaluators verify – prior to assigning an observable behavior – whether the pilot has effectively utilized all the resources at their disposal. This analysis involves the proper use of the aircraft's automation systems, effective crew communication, and decision-making based on the situational information available.

This method aims to reduce differences in judgment among evaluators and ensure a more reliable process aligned with the best practices of CBTA. Furthermore, to exemplify the practical application of this framework, Chapter 5 will present two concrete situations of OB evaluation in specific competencies. These situations will be analyzed with methodological rigor and based on structured techniques, ensuring the replicability and scientific grounding of the examples discussed. Thus, the transition from the theoretical framework to its practical application will be demonstrated clearly and objectively, reinforcing the importance of standardization in the evaluation of OBs within CBTA.

5 RESULTS AND DISCUSSION

Subjectivity in the evaluation of observable behaviors (OBs) can be reduced through a structured approach that takes into account the effective utilization of the resources available to the pilot. These resources include, but are not limited to, communication with air traffic control (ATC), collaboration with other crew members, proper use of the aircraft's monitoring systems, operational redundancies, procedural alternation, and normative references. The following sections present two hypothetical examples that illustrate how the use of these resources can influence the evaluation of pilot performance.

5.1 EXAMPLE 1: ADJUSTMENT OF APPROACH PROCEDURE DUE TO A DIVERGENCE IN ANP

During an approach using a Performance-Based Navigation (PBN) procedure, a pilot en route to a major airport receives an indication from the Flight Management System (FMS) that the Actual Navigation Performance (ANP) is higher than the Required Navigation Performance (RNP).

This condition compromises navigation accuracy and may necessitate an alternative procedure. In this situation, the pilot can utilize several resources to mitigate the problem, including:

- Coordination with ATC: Request an alternative procedure, such as vectoring or an approach based on ground aids.
- Collaboration with the Co-pilot: Discuss the ANP condition and jointly analyze potential solutions.
- Verification of Aircraft Systems: Confirm the integrity of navigation sensors and check for possible calibration failures.
- Consultation of the Operations Manual: Review standard procedures for addressing ANP discrepancies.
- Use of Redundancies: If the aircraft is equipped with alternative navigation systems (such as VOR/DME or ILS), the pilot may consider their use.

The evaluation of the OB corresponding to this situation should not be based solely on the pilot's final decision, but rather on an analysis of how they managed the available resources. If the pilot ignored the ANP indication without taking appropriate mitigating actions, they may be evaluated negatively. Conversely, if they effectively utilized the available resources to make an informed and safe decision, the corresponding competency can be considered demonstrated.

5.2 EXAMPLE 2: CONFIRMATION OF LANDING GEAR CONFIGURATION BEFORE TOUCHDOWN

During an approach for landing under low-visibility conditions, a pilot receives an audible alert indicating a possible failure in the extension of the landing gear. In this situation, the immediate action should not be simply to execute an instinctive go-around; instead, the pilot should utilize all available resources to confirm the status of the landing gear.

The resources the pilot may use include:

- Visual and Sensor Checks: If the aircraft model allows, the co-pilot can verify external mirror indicators or camera systems.
- Use of Alternative Indicators: Cross-check hydraulic pressure, selector positions, and warning lights.
- Coordination with ATC: Request that the control tower or another aircraft visually confirm that the landing gear is down and locked.
- Consultation of the Checklist: Follow the Quick Reference Handbook (QRH) procedures to verify and, if necessary, execute corrective actions.
- Collaboration with Other Crew Members: In operations involving a flight engineer, this crew member may assist in verifying the system – for instance, checking if a single burnt-out indicator light is causing the failure signal.
- Planning Alternatives: Assess whether it is necessary to divert to an airport with better landing infrastructure if the landing gear's status remains uncertain.

The evaluation of the OB in this situation should consider whether the pilot efficiently utilized the available resources before deciding to continue the approach or to execute a go-around. If a hasty decision was made without seeking additional confirmation, the pilot's competency may be questioned. However, if the pilot demonstrated a structured process and consulted all available resources to validate the landing gear's condition, the evaluation may indicate proficiency in resource management and decision-making.

6 CONCLUSIONS

This study analyzed the challenges of subjectivity in the evaluation of OBs within the CBTA framework, highlighting the need to improve evaluation criteria to ensure greater consistency and reliability in the process. The proposed approach emphasized the efficient use of the resources available to the pilot, providing a more structured benchmark for evaluators' decision-making.

The results indicate that subjectivity in OB evaluation can be mitigated through a more standardized framework, focusing not only on the outcome of the pilot's action but also on the process by which the decision was made. Evaluations should consider whether the pilot appropriately utilized available resources – such as coordination with air traffic control (ATC), aircraft redundancies, crew interaction, and consultation of operational manuals. This model allows for a more detailed performance analysis, reducing interpretative discrepancies among evaluators.

The central hypothesis of this study was that standardizing evaluation criteria, based on the effective use of resources and proper management of operational variables, could reduce the influence of evaluator subjectivity. The findings confirm this hypothesis, demonstrating that the structuring of an evaluation model with more objective parameters increases the reliability of competency certification and aligns with international best practices.

In addition to theoretical support, empirical evidence plays an essential role in enhancing CBTA. Systematic observation of pilots in training and the collection of operational data are indispensable tools for validating the proposed models. The accumulated experience of instructors and examiners, combined with the analysis of real data, contributes significantly to the construction of a more robust evaluation model that meets the operational needs of aviation.

Given that instructors and examiners are at the forefront of the evaluation system, this study recommends further research aimed at improving the training and standardization of evaluation criteria for Flight Instructors (INVA), Designated Pilot Examiners (DPE), Type Rating Instructors (TRI), and Type Rating Examiners (TRE). Continuous professional development for these professionals is fundamental to ensuring that the principles of CBTA are consistently applied and aligned with the demands of modern aviation.

Furthermore, it is recommended to expand studies on the impact of subjectivity in the practical evaluation of pilots, exploring the feasibility of using new technologies – such as artificial intelligence and machine learning-based data analysis – to support evaluators' decision-making. The creation of databases with structured evaluation records can also help identify patterns and trends that facilitate the continuous refinement of the CBTA model.

Finally, this study reinforces the need to continuously evolve competency evaluation processes in aviation. The adoption of more structured theoretical frameworks, combined with empirical methodologies and advanced technologies, can provide a fairer, more precise evaluation environment that meets the demands of the aviation sector. In this way, CBTA is consolidated as an essential model for the future training and certification of pilots, ensuring an increasingly high level of operational safety.

REFERENCES

BIDDLE, E. Adaptive learning considerations for commercial pilot pipeline. In: SINATRA, A. M.; GRAESSER, A. C.; HU, X.; TOWNSEND, L. N.; RUS, V. (ed.). **Design recommendations for intelligent tutoring systems: professional career education**. Orlando: US Army Combat Capabilities Development Command – Soldier Center, 2023. v. 11, p. 69.

BIEDE, S. *et al*. A competency framework for aviation psychologists and human factors specialists in aviation. In: **INTERNATIONAL SYMPOSIUM ON AVIATION PSYCHOLOGY**, 22., 2023, Wright State University: CORE Scholar, 2023a. p. 99. Available at: https://corescholar.libraries.wright.edu/isap_2023/62. Accessed on: 16 feb. 2025.

BIEDE, S. et al. **Handbook on competencies for endorsement as an aviation psychologist and an aviation human factors specialist: including requirements for continuous professional development**. 2. ed. [S.l.]: European Association for Aviation Psychology, 2023b.

BORUCKA, A.; ROMELE, M. **Examining Airplane Accidents through the Lens of Fatalities in Such Catastrophes**. [*s. l.*], p. 84–87, 2024.

COLONESE, J. **Competency Based Training and Assessment in Brazilian Commercial Pilot Licensing: New methodology updating theoretical and practical**. Dissertação (Mestrado em Operações de Transporte Aéreo) - ISEC Lisboa, Lisboa, 2022. Available at: http://hdl.handle.net/10400.26/40921. Accessed on: 07 nov. 2024.

IATA - International Air Transport Association. **Competency-Based Training and Assessment (CBTA) Expansion within the Aviation System**. Montreal: IATA, 2021.

IATA - International Air Transport Association. **Competency assessment and evaluation for pilots, instructors/evaluators: guidance material**. 4 ed. Geneva: IATA, 2025. Available at: https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/competency-assessment-and-evaluation-for-pilots-instructors-and-evaluators-gm.pdf. Accessed on: 19 feb. 2025.

ICAO - International Civil Aviation Organization. **Doc 9868 – Procedures of Air Navigation Services - Training PANS-TRG**. Third Edition, 2020.

LÁZARO, F. et al. Human factors as predictor of fatalities in aviation accidents: a neural network analysis. **Applied Sciences**, v. 14, n. 2, p. 640, 2024. Available at: https://doi.org/10.3390/app14020640. Accessed on: 16 feb. 2025.

MANSIKKA, H.; HARRIS, D.; VIRTANEN, K. An input-process-output model of pilot core competencies. **Aviation Psychology and Applied Human Factors**, Göttingen, v. 7, n. 2, p. 78-85, 2017. Available at: https://dx.doi.org/10.1027/2192-0923/a000120. Accessed on: 19 feb. 2025.

SUN, H.; YANG, F.; ZHANG, P.; HU, Q. Behavioral indicator-based initial flight training competency assessment model. **Applied Sciences**, v. 13, n. 10, p. 6346, mai. 2023. 2023a. Available at: https://doi.org/10.3390/app13106346. Accessed on: 17 feb. 2025.

SUN, H.; ZHOU, X.; ZHANG, P.; LIU, X.; LU, Y.; HUANG, H.; SONG, W. Competency-based assessment of pilots' manual flight performance during instrument flight training. **Cognition, Technology & Work**, v. 25, p. 345–356, nov. 2023. 2023b. Available at: https://doi.org/10.1007/s10111-023-00737-3. Accessed on: 17 feb. 2025.

YILMAZ, A. Flight in transition: navigating the skies of automation and human judgement. **Journal of Aviation**, v. 8, n. 3, p. 305–314, 2024. Available at: https://doi.org/10.30518/jav.1486331. Accessed on: 16 feb. 2025.