applied sciences

Article

Behavioral Indicator-Based Initial Flight Training Competency Assessment Model

Hong Sun [1], Fangquan Yang [2], Peiwen Zhang [3,*] and Qingqing Hu [2]

[1] Flight Technology and Flight Safety Research Base, Civil Aviation Flight University of China, Guanghan 618307, China
[2] School of Airport, Civil Aviation Flight University of China, Guanghan 618307, China
[3] School of Economics and Management, Civil Aviation Flight University of China, Guanghan 618307, China
* Correspondence: zhangpeiwen@cafuc.edu.cn

Abstract: Ensuring training safety is paramount to flight schools. In response to the inadequacy of traditional flight training assessment for comprehensive quantitative evaluation of cadet competency, an initial flight training competency assessment standard based on behavioral indicators was developed and optimized using the VENN model. Firstly, the Assessor Score Measurement Form (ASMF) was constructed according to the requirements of the Training Evaluation Worksheet specification, such as typical subjects, observations, and completion criteria. Secondly, based on the basic principles of the experience of the flight expert and the Competency-Based Training and Assessment (CBTA), a matrix of correlations between the observations and each competency-based behavioral indicator was created to construct a competency assessment matrix. In addition, a two-dimensional model for representing competency items characterized by behavioral indicators was established and an optimization model for competency assessment criteria was constructed. Finally, through combining actual flight training data, the proposed method was validated in the flight screening check phase. The results show that the optimized flight training competency assessment scheme can be well quantified and matched to real instructor ratings with an accuracy of 84%. The assessment worksheet, the assessment matrix, and the VENN competency rating model can be adapted to the different teaching requirements of each flight phase, achieving a perfect match between the behavioral indicators and the competency items, which is highly versatile. The proposed model can more accurately reflect the core competencies of flight trainees, enable quantitative assessment of behavioral indicators and competency items, and provide support for subsequent training of trainees.

Keywords: behavioral indicators; initial flight training; competency; evaluation criteria

Citation: Sun, H.; Yang, F.; Zhang, P.; Hu, Q. Behavioral Indicator-Based Initial Flight Training Competency Assessment Model. *Appl. Sci.* **2023**, *13*, 6346. https://doi.org/10.3390/app13106346

Academic Editors: Jérôme Morio, Luís Santos, Rui Melicio and André Silva

Received: 6 April 2023
Revised: 16 May 2023
Accepted: 18 May 2023
Published: 22 May 2023

1. Introduction

The Airline Transport Pilot License theory test (ATPL) is a test that students must pass to work in airline transport. The initial flight training is a key stage in helping students build the comprehensive skills needed to enter airline transport flight in the future. The purpose of the training evaluation of the students in the ATPL flight school is to evaluate the skills that they show during the training, finding their skill structure deficiencies and optimizing the training program. Significant safety risks may arise on subsequent flights if the training is not tailored to the relevant technical requirements and training characteristics. For example, on 29 October 2018, a B-737 Max aircraft operated by Indonesia's Lion Air Airlines with call sign PK-LQP crashed into the Java Sea at 23:31:53 UTC. All 189 people on board were killed, and the aircraft was destroyed. In this case, the pilot's lack of training in the specific Maneuvering Characteristics Augmentation System (MCAS) technique was the cause of the accident. Therefore, a comprehensive and accurate flight capability assessment is essential to ensure the quality and progress of flight training. The current training evaluation of flight students in flight schools mainly relies on the experienced judgment of highly qualified flight instructors. However, with the rapid development of civil aviation, the traditional training evaluation method has two major flaws. On one hand, the trainees remain only at the pass or fail stage, with the assessment lacking a comprehensive picture of the cadet's ability structure, which is not in line with the Pilot Skills Life Cycle Management (PLM) concept proposed by the civil aviation industry; on the other hand, the implementation of standards by different flight instructors may lead to different assessment criteria, making it difficult to ensure the objectivity and stability of the assessment's results. Therefore, there is an urgent need to change the initial training quality assessment method from the old "tick-box" and "fixed subject" training quality assessment to a core Competency-Based Training Assessment (CBTA) based on Observable Behaviors (*OB*) [1]. For this purpose, it is necessary to rely on the core competency assessment index system established by the International Civil Aviation Organization (ICAO) in order to optimize the flight training competency assessment scheme, achieve refinement and precision in assessing flight competence, and solve the practical problems associated with the traditional instructor assessment, which is more subjective and less stable.

In connection with the development of sensors, the Internet of Things, and artificial intelligence, in the field of aviation, aircraft have a large number of different types of sensor devices. The research on evaluating based on sensing data has attracted much attention worldwide. Researchers conducted numerous studies on aviation safety [2–4], flight quality [5], and flight monitoring [6–8] with the help of sensor flight parameter data, and provided a large amount of data in support of flight operation quality assessment [9–11]. From a psychological perspective, the evolution of human-based behavioral research was accompanied by the development of human-monitoring devices. In the field of civil aviation, many researchers investigated the relationship between pilots' physiological factors and flight operation behavior [12–14]. However, as research on "pilot competence" intensified, the research on flight quality shifted to the core competencies of pilots via analyzing industry characteristics and behavioral traits [15]. The research on flight quality based on competencies, therefore, became of the main focuses of research.

Until now, the evaluation of flight operation quality in China and abroad was conducted from three perspectives: flight parameter data, pilot physiological parameter data, and pilot core competency data. Firstly, in terms of flight parameter data, researchers utilized big data from the Quick Access Recorder (QAR) to quantify the operational quality of pilots. For example, Liu S. et al. [16] developed a system to evaluate flight operation performance and a quantitative evaluation method model based on QAR data. One or more flight parameters were selected for combination to objectively evaluate the pilot's flight operation performance. Sembiring J. et al. [17] estimated the parameters of some aircraft parameters for subsequent flight maneuver evaluation using aircraft quick access recorder data via the output error method based on the maximum likelihood principle and the classical least squares method. Wang L. et al. [5,11,18] analyzed the QAR data to extract and characterize the flight parameters of the aircraft during the landing and propose preventive measures from the perspective of the pilot operation. In summary, from the perspective of flight parameter data, existing methods are able to overcome the subjectivity of instructor evaluation, rely on flight data, and match data to operational capability through the knowledge transformation paradigm to achieve digital quantitative evaluation, fully exploiting the obvious advantages of flight big data in measuring pilot operational capability. However, the variability in cadets' flight skills and the complexity of training scenarios lead to poor quality of actual flight data, making it difficult to conduct scientific and standardized data analysis. Furthermore, flight data are only a concrete demonstration of behavioral actions in flight operations, which does not include the assessment of deeper physiological and psychological factors behind the iceberg theory.

In terms of physiological parameters, related researchers use relevant physiological data, such as the pilot's heart rate, heartbeat, and respiration rate, to make a comprehensive and holistic evaluation of flight operation quality. For example, Lahtinen et al. [19] showed that heart rates reflect the magnitude of cognitive load during simulated flight through recording Electrocardiograms (ECGs) and calculating individual Incremental Heart Rates

(IHR) at rest during each flight phase and using them for statistical analysis. Maciejewska et al. [20] analyzed pilots' psychophysiological states according to their cardiovascular work, examining Heart Rate Variability (HRV) parameters. In summary, in the quantitative research of flight maneuver quality evaluation, scholars mostly used QAR data or psychological signal data as the basis; selected internal indicators; and established corresponding evaluation models, which can evaluate the overall training quality. Through the changes in physiological factor indicators, this method can reflect the pilot's flight physiological state to a certain extent, though it is difficult to objectively analyze and evaluate their specific ability level. In addition, none of the above studies include the study of cadet competence development in flight academies, the correspondence between physiological information and specific competencies is lacking, and the study of flight cadet training competencies is still in its infancy. Therefore, from the point of view of competencies, the study of all aspects of flight cadets' comprehensive competencies from the core competencies is the focus of current flight training research, which can expand the evaluation of ideas regarding cadet operational skills.

In the context of competency assessment research, Jirgl et al. [21] discussed the change or progressive development of pilot competencies during training and based their hypothesis on a corresponding behavioral model against which pilot competencies can be initially assessed. Mansikka et al. [15] used Principal Component Analysis (PCA) of pilot performance scores for different competencies to construct a path analysis model to examine the relationship between different flight-related core competencies of professional airline pilots. Sarkar et al. [22] focused on the process of skills mapping and the use of skills mapping for training needs assessment to validate changes in skills gaps for needs-driven training. To summarize, research to date on competence assessment tends to focus on the theoretical level, which has provided a more comprehensive theoretical underpinning, but lacks more detailed assessment criteria. Behavioral indicators of students' level of competence, as demonstrated in specific subjects and training, are also relatively unexplored.

In conclusion, the specific characteristics of the above research are shown in Table 1. It is clear that current research is mainly confronted with the complexity of data processing, the limitations of the assessment scope, and the subjectivity of the assessment criteria. Moreover, current research on student handling quality in initial flight training mainly focuses on quantitative analysis of QAR and physiological data, though there is a lack of in-depth research on core competency assessment. Although the ICAO has proposed the nine core competencies of pilots and a theoretical framework for competency assessment based the *OB*, it lacks quantitative definitions of competency assessment criteria, especially for the implementation of CBTA in the initial flight training phase. In particular, there is a lack of quantitative definitions of competency assessment criteria and manipulable recommendations for the implementation of CBTA in the initial flight training phase [16]. Therefore, based on the traditional operational model of flight training performance assessment, using specific flight training practices as the research object and relying on regulatory documents, this paper proposes an optimized solution for the initial flight training competency assessment criteria based on behavioral indicators. For the first time, specific flight data, flight maneuver characteristics, and competency items are mapped to form a more comprehensive, refined, and objective assessment process through the Venn competency assessment model. This approach ensures the scientific and objective nature of the assessment and promotes the further development of pilot core competency-based assessment research, providing suggestions and directions for subsequent targeted competency training improvements. In Section 2, the basic concept of core competency and the observable terms are introduced. In Section 3, the proposed competency-based assessment optimization model is presented in detail. In Section 4, the validity of the design is verified through comparing the *OB*-based evaluation criteria to the examiner's evaluation, using "screening check" as an example. Finally, Section 5 summarizes the contributions and implications of this paper and draws conclusions.

Table 1. Related literature features.

Research Category	Related Literature	Difficulty of Data Processing	Comprehensiveness of the Assessment	Nature of Assessment
Flight data aspects	[5,11,18,20]	Extremely complex	Singularity	Objectivity
Physiological data aspects	[19,23,24]	Complex and less relevant	Singularity	Objectivity
Competence aspects	[15,21,22]	Lack of specific criteria	Comprehensive but not detailed	Subjective
The proposed model	-	Easy to access and understand	Comprehensive and detailed	Objectivity

2. Research Method

Flight training is the basis for flight safety assurance and high-quality development in civil aviation. A scientific and standardized training quality assessment scheme is an important linkage to control training quality and improve training efficiency. In the increasingly complex civil aviation system, the ICAO proposed that flight training and assessment focus on nine core competencies for pilots, including Application of Knowledge (KNO), Application of Procedures and Compliance with Regulations (APK), Communication (COM), Airplane Flight Path Management—Automation (FPA), Airplane Flight Path Management—Manual control (FPM), Leadership and Teamwork (LTW), Problem Solving and Decision-Making (PSD), Situation Awareness and Management of Information (SAW), and Workload Management (WLM). As a strategy to continuously improve global aviation safety, the Global Aviation Safety Program (GASP) emphasizes the Competency-Based Training System (CBTA). On 21 June 2019, the Civil Aviation Administration of China (CAAC) issued the "Guidance on the Comprehensive Deepening of Transport Airline Flight Training Reform" document, which explicitly puts forward the new era of flight training reform guidelines for "implementing flight training based on core competence". Among them, nine competencies of mature pilots [25] and the dimensions of behavioral indicators are defined, as shown in Table 2 below.

Table 2. Competencies and descriptions.

Competency	Description	Observable Behavior (*OB*)
0. Application of knowledge	Demonstrates knowledge and understanding of relevant information, operating instructions, aircraft systems, and the operating environment.	OB0.1–OB0.7
1. Application of procedures and compliance with regulations	Identifies and applies appropriate procedures, in accordance with published operating instructions and applicable regulations.	OB1.1–OB1.7
2. Communication	Communicates through appropriate means in the operational environment, in both normal and non-normal situations	OB2.1–OB2.10
3. Airplane flight path management—automation	Controls the flight path through automation.	OB3.1–OB3.6
4. Airplane flight path management—manual control	Controls the flight path through manual control.	OB4.1–OB4.7
5. Leadership and teamwork	Influences others to contribute to a shared purpose. Collaborates to accomplish the goals of the team	OB5.1–OB5.11
6. Problem solving and decision-making	Identifies precursors, mitigates problems, and makes decisions.	OB6.1–OB6.9

Table 2. *Cont.*

Competency	Description	Observable Behavior (*OB*)
7. Situation awareness and management of information	Perceives, comprehends, and manages information and anticipates its effect on the operation.	OB7.1–OB7.7
8. Workload management	Maintain available workload capacity through prioritizing and distributing tasks using appropriate resources.	OB8.1–OB8.8

Competencies must be demonstrated through a set of "behaviors" that can be observed and assessed; ICAO and IATA outlined specific "Behavioral Indicators—*OB*" for each "competency" based on extensive research. If a pilot has a sufficient number of OBs in flight training, the pilot can be judged to have the appropriate performance level. Taking as an example the most important and fundamental skill of initial flight training, i.e., Flight Path Management (FPM), the FPM skills of a mature pilot should include the dimensions of *OB* shown in Table 3. Existing competency assessment guidelines are mainly based on the VENN model proposed by the ICAO, which assesses the level of competency in three dimensions: quantity, frequency, and results of hazard and error management. VENN model assessment essentially uses the minimum of the three-dimensional assessment scores as the final competency level.

Table 3. Competence "Flight Track Management—Manual Flight" *OB* item.

Observable Behavior (*OB*)	Description of the Observable Behavior (*OB*)
OB4.1	Controls the aircraft manually with accuracy and smoothness as appropriate to the situation.
OB4.2	Monitors and detects deviations from the intended flight path and takes appropriate action.
OB4.3	Manually controls the airplane using the relationship between airplane attitude, speed and thrust, and navigation signals or visual information.
OB4.4	Manages the flight path safely to achieve optimum operational performance.
OB4.5	Maintains the intended flight path during manual flight while managing other tasks and distractions.
OB4.6	Uses appropriate flight management and guidance systems, as installed and applicable to the conditions.
OB4.7	Effectively monitors flight guidance systems including engagement and automatic mode transitions

From the perspective of the competency evaluation index system and guidelines, there are specific provisions for core competency evaluation indexes and situational elements for identifying competencies; however, there is only a principled approach for evaluating each competency, as well as a lack of operational quantitative criteria. Therefore, implementation is particularly dependent on the experienced judgment of highly qualified instructors, and there is a lack of specific description of core competency training requirements and evaluation criteria for initial flight training. In practice, there are some deficiencies, such as the standardized definition of behavior indicators *OB*, as well as how to measure and quantitatively classify the number and frequency of *OB* displays given the lack of quantitative standards.

3. Optimization Model of Competency Assessment Criteria Based on VENN Criteria

Combining the VENN criteria with the core concept of competency, this paper proposes an optimization model of competency evaluation criteria based on the VENN criteria with reference to the traditional flight training performance evaluation operation model. The proposed model framework includes designing training assessment worksheets, constructing measurement vectors, constructing correlation matrices between observations and behavioral indicators, and creating a CBTA Competency Assessment Matrix that can be applied to all phases of initial flight training with four core modules. The specific model framework is shown in Figure 1.

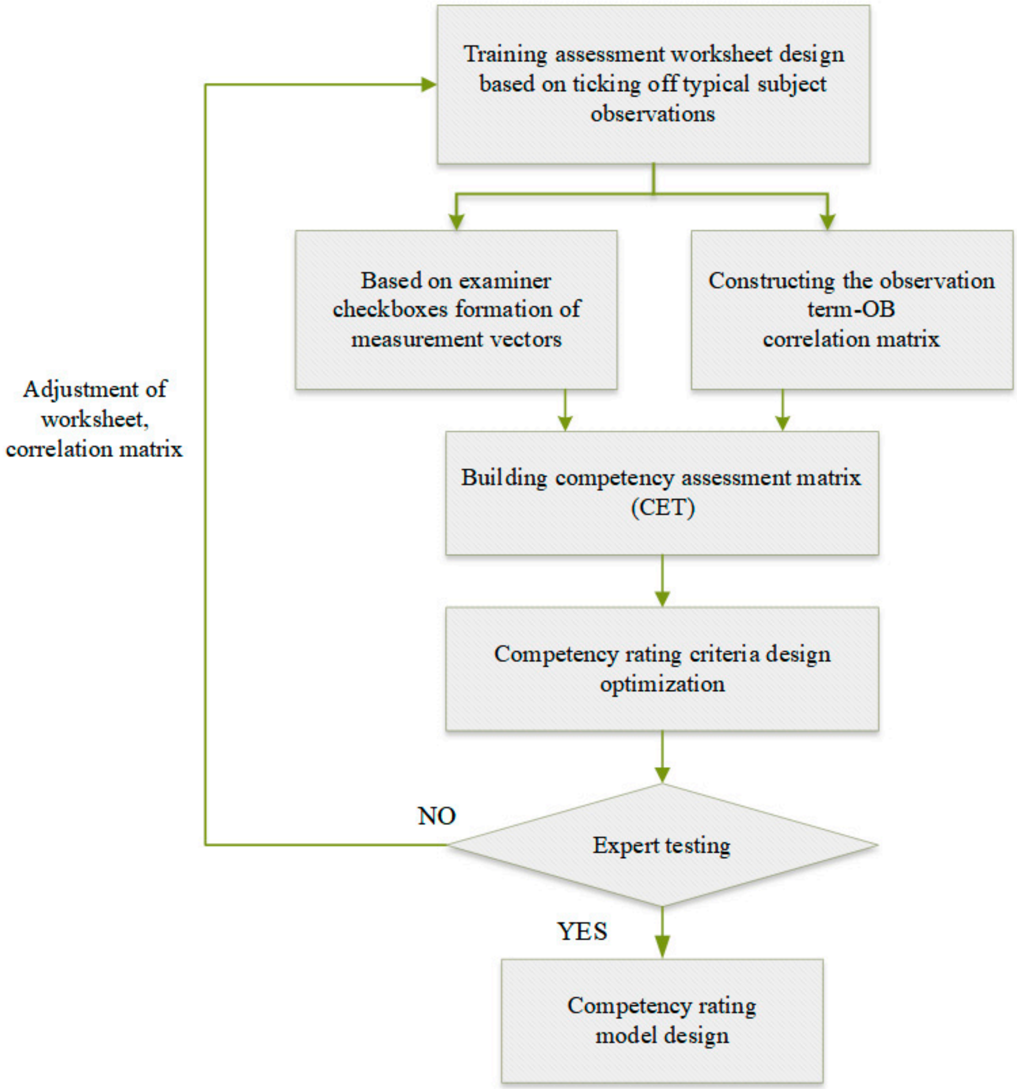

Figure 1. Initial flight training CBTA evaluation process.

3.1. Training Evaluation Worksheet

Subjects are an important tool for organizing and conducting initial flight training. Thus, students are trained through a series of typical subjects to develop relevant skills, and the quality of the training is assessed through examining the students' performance in each subject. Based on the characteristics of traditional training evaluation implementation, a uniform evaluation worksheet was designed by flight experts for each inspection item. The typical examination topics, as well as the observation items and completion criteria for each topic, were standardized in the modified worksheet to provide a consistent quantitative measure of the cadet's skill mastery, as detailed in Table 4. The initial training assessment worksheet is designed. During the assessment process, the assessor scores the trainee's observations in each subject based on the completion criteria of the training assessment worksheet.

Table 4. Design of Initial Flight Training Evaluation Worksheet.

Subject	Observation (*OB*)	Scoring Criteria	Examiner Scoring
	OB. 1	4: . . . ; 3: . . . ; 2: . . . ; 1: . . .	. . .
Subject 1	*OB*. 2	4: . . . ; 3: . . . ; 2: . . . ; 1: . . .	. . .
	. . .	. . .	. . .
. . .	. . .	. . .	. . .
	. . .	. . .	. . .
Subject K	*OB*. m − 1	4: . . . ; 3: . . . ; 2: . . . ; 1: . . .	. . .
	OB. m	4: . . . ; 3: . . . ; 2: . . . ; 1: . . .	. . .

According to the flight training practical examination standards issued by the Civil Aviation Administration of China (CAAC) and the requirements of each training institution's curriculum, the observation items and evaluation criteria for each training subject can be analyzed. The evaluation criteria focus on refining the evaluation scale, scoring performance as 4, 3, 2, or 1. The assessor obtains a score observation vector through scoring the student's completion, as shown in Equation (1).

$$\begin{cases} A_s = (a_i)_{m\times 1} = (a_1, a_2, \cdots, a_m) \\ i = 1, 2, 3 \cdots m \\ s = 1, 2, 3 \cdots q \end{cases}, \tag{1}$$

where A_s is the observation vector of the S-th sample participant, a_i is the score of the i-th observation, and its maximum value $a_i^{\max}$ is the full score of the observation. If all observations have full scores, the observation vector can be obtained as follows. $A^{\max} = \left(a_1^{\max}, a_2^{\max}, \cdots a_m^{\max}\right)^T$.

For example, the landing attitude subject contains three observations, namely the pull start height, the pull level height, and the grounded attitude. The Section Landing Attitude Training Evaluation Worksheet is shown in Table 5.

Table 5. Section Landing Attitude Training Evaluation Worksheet.

Subject	Observation	Scoring Criteria	Examiner Scoring
Landing position	Pulling start height	4: Conform to the regulations. 3: Within 1 m. 2: Within 2 m. 1: Beyond 2 m.	4
	Leveling height	4: Points: conform to the regulations. 3: Within 0.25 m, slightly pulled, corrected correctly. 2: Points; within 0.5 m, slightly pulled, corrected correctly. 1: Points: within 0.5 m.	3
	Grounding gesture	4: Three points smoothly grounded. 3: Slightly tilted when grounded. 2: Significant tilt when grounded. 1: Jump when grounded.	3

3.2. Correlation Matrix of Observations Corresponding to Behavioral Indicators

To avoid the disadvantages of traditional initial flight training assessments, i.e., "only subjects, only results, but not ability", the mapping relationship between data and ability should be further established using the concept of CBTA. There are corresponding behavioral indicators *OB* for each of the nine core competencies, and each observation corresponds to the behavioral indicator of a particular competency that is used in the training evaluation. Using the Delphi survey method to solicit the opinions of flight experts, an association can be constructed between any observable *i* and the behavioral indicator of competence OBJ, and an association matrix B between the observable and the behavioral indicator can be constructed, as shown in Equation (2).

$$B = [B_1, B_2, \cdots B_n] = \begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}, \tag{2}$$

where b_{ij} denotes the association property of the i-th observation with the j-th OB, and if b_{ij} = 1, it means that the i-th observation is associated with the n-th OB; otherwise, it takes 0. $i = 1, 2, \ldots, m; j = 1, 2, \ldots, n$.

3.3. Modeling of VENN Criteria Based on Competency Assessment Matrix

According to the VENN guidelines, a student's competency level can be measured through counting the number and frequency of OBs demonstrated in the assessment and constructing a competency assessment matrix using the observation vector A_s and the association matrix B, as shown in Equation (3).

$$Y = [Y_1, Y_2, \ldots, Y_n] = \begin{bmatrix} a_1 b_{11} & a_1 b_{12} & \cdots & a_1 b_{1n} \\ a_2 b_{21} & a_2 b_{22} & \cdots & a_2 b_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_m b_{m1} & a_m b_{m2} & \ldots & a_m b_{mn} \end{bmatrix}, \tag{3}$$

where $a_o b_{ij}$ represents the contribution of the i-th observation to m. Using the properties of vector (or matrix) parameterization [19], which have the length of the space of the metric vector (or matrix), the frequency and number of observations can be shown through the parametric characterization of the behavior indicator (OB) of the Y-matrix. Firstly, it is agreed that if the frequency of the OB displayed exceeds 25% of the maximum, the OB is displayed, while the opposite is not displayed. Through calculating the rating matrix paradigm using Equations (4) and (5), the number (f_{mny}) and frequency (f_{ofn}) of OB presentations based on the competency rating matrix are obtained.

$$f_{mny} = count\left\{ Y_j, ||Y_j||_1 \geq \frac{1}{4} A^{\max} B_j, j = 1, 2, \ldots, n \right\}, \tag{4}$$

$$f_{ofn} = ||Y||_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_{ij}, \tag{5}$$

3.4. Competency Assessment Criteria Optimization Model

This grading method was used in the design of the OB-based Competency Rating Criteria to continue the traditional training evaluation using the four rating categories of excellent, good, fair, and poor. Here, p_{exa} indicates the grade the proctor gave the student, i.e., $P_{exa} = \{\text{Excellent, good, medium, poor}\} = \{4, 3, 2, 1\}$.To facilitate further calculation and comparison, it is necessary to convert Equations (4) and (5). Indeed, if all observations are taken as full values, i.e., $A^{\max} = \left(a_1^{\max}, a_2^{\max}, \cdots a_m^{\max}\right)^T$, the maximum value of the number (f_{mny}) and frequency (f_{ofn}) of OB presentations is obtained according to the evaluation matrix as follows.

$$f_{mny}^{\max} = count\left\{ Y_j, \left|\left| Y_j^{\max} \right|\right|_0 > 0, \forall j = 1, 2, \ldots, n \right\}, \tag{6}$$

$$f_{ofn}^{\max} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i^{\max} b_{ij}, \tag{7}$$

Given the different training institutions and training courses involved in observing and completing the standard set of differences in the situation, as well as the need to facilitate the unification of competency assessment standards, the need to show the number of *OB* (f_{mny}), frequency (f_{ofn}) for normalization, is as follows. $\overline{f_{ofn}}, \overline{f_{mny}} \in [0,1]$

$$\overline{f_{mny}} = \frac{f_{mny}}{f_{mny}^{\max}}, \tag{8}$$

$$\overline{f_{ofn}} = \frac{f_{ofn}}{f_{ofn}^{\max}}, \tag{9}$$

where $\overline{f_{ofn}}, \overline{f_{mny}} \in [0,1]$. p_{ofn} denotes the evaluation of *OB* presentation frequency (f_{mny}), p_{mny} is the evaluation of *OB* presentation quantity, and (f_{ofn}) is the final competency evaluation based on VENN criteria. Equations (9)–(11) show the competency evaluation model.

$$P_{OB} = \min(P_{ofn}, P_{mny}), \tag{10}$$

$$P_{ofn} = \begin{cases} 1, 0 \le \overline{f_{ofn}} < \partial_1 \\ 2, \partial_1 \le \overline{f_{ofn}} < \partial_2 \\ 3, \partial_2 \le \overline{f_{ofn}} < \partial_3 \\ 4, \partial_3 \le \overline{f_{ofn}} \le 1 \\ \partial_1 \le \partial_2 \le \partial_3 \end{cases}, \tag{11}$$

$$P_{mny} = \begin{cases} 1, 0 \le \overline{f_{mny}} < \gamma_1 \\ 2, \gamma_1 \le \overline{f_{mny}} < \gamma_2 \\ 3, \gamma_2 \le \overline{f_{mny}} < \gamma_3 \\ 4, \gamma_3 \le \overline{f_{mny}} \le 1 \\ \gamma_1 \le \gamma_2 \le \gamma_3 \end{cases}, \tag{12}$$

where $\partial_1, \partial_2, \partial_3$ and $\gamma_1, \gamma_2, \gamma_3$ denote the hierarchical frequency ($\overline{f_{ofn}}$) and quantity ($\overline{f_{mny}}$) thresholds for presenting the *OB*, respectively. These threshold calculations are first obtained via solving, based on sample data from the flight expert's evaluation of cadet training quality, an integer optimization problem consisting of Equations (9)–(12).

$$\min \sum_{s=1}^{q} |P_{OB} - P_{exa}|, \tag{13}$$

where the objective function Equation (13) represents the minimum mean deviation based on the VENN criterion from the ratings given by the examiner.

4. Case Research

4.1. Screening Check Phase Competency Assessment

In this section, based on the overall training syllabus for an air transport pilot course, the 'screening check' phase of the single engine airplane private pilot training program is selected as an example. The screening check, which focuses on the fourth of the nine competencies, i.e., "Flight Path Management—Manual Flight" (FPM), is an important assessment of a student's "flying talent" in the early stages of initial flight training. In addition, the screening check focused mainly on the selection of basic piloting skills; thus, the two higher-order FPM pilot competency indicators, i.e., OB4.6 and OB4.7, were not addressed in the behavioral indicators, while the behavioral FPM competency indicators demonstrated by the trainees in the screening phase focused on the five dimensions of OB4.1–OB4.5. The study selected a sample of 93 trainees in 2020, of whom 74 were trained and 19 were tested, and conducted a statistical analysis of the trained samples' performances on the screening exam. The overall program implementation process consisted of four steps. Using a participant as an example, the specific steps are as follows:

(1): Construct the Observation Vector

Firstly, according to the main assessment items of the screening check [17], the training assessment worksheet was determined, which consisted of 24 typical subjects, 99 observations, and corresponding scoring criteria, as shown in Table 6 (See Appendix A for detailed criteria). In additional, the observation vector can be derived from the examiner scoring column in Table 6: $A = (4, 2, \cdots, 3, 3, 3)^T$.

Table 6. Screening Check Training Evaluation Worksheet.

Subject (Sub)	Observation (No)	Scoring Criteria	Examiner Scoring
Sub 1: Up Down	No. 1: Direction of navigation	4: Maintain accuracy. 3: Within 5 degrees. 2: Within 10 degrees. 1: Beyond 10 degrees.	4
	No. 2: Speed	4: Maintain accuracy. 3: Within 5 knots. 2: Within 10 knots. 1: Beyond 10 knots.	2
...	...	...	...
Sub 24: Landing position	No. 97: Pulling start height	4: Conform to the regulations. 3: Within 1 m. 2: Within 2 m. 1: Beyond 2 m.	3
	No. 98: Leveling height	4: Conform to the regulations. 3: Within 0.25 m, slightly pulled, corrected correctly. 2: Within 0.5 m, slightly pulled, corrected correctly 1: Within 0.5 m.	3
	No. 99: Grounding gesture	4: Three points smoothly grounded. 3: Slightly tilted when grounded. 2: Significant tilt when grounded. 1: Jump when grounded.	3

(2): Construct the Observation and *OB* Correlation Matrix

The research first used a Delphi survey to solicit input from flight professionals to correlate the 99 screening checklist observations with the 5 FPM behavioral indicators. Here, 1 indicates correlation between an observation and behavioral measure, while 0 indicates no management relation. The final analysis of the flight expert's opinion is aggregated to obtain the Observation–Behavior Correlation Matrix B.

$$B = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}$$

(3): Construction of a competency assessment matrix

From the observation vector A and the association matrix B, the evaluation matrix Y can be constructed in combination. The number ($f_{mny} = 5$) and frequency ($f_{ofn} = 240$) of the participant's *OB* performances based on the competency rating matrix can be obtained via calculating the rating matrix paradigm.

$$Y = \begin{bmatrix} 0 & 4 & 4 & 0 & 0 \\ 0 & 2 & 2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 3 & 3 & 0 & 0 & 0 \\ 3 & 0 & 0 & 3 & 0 \\ 3 & 0 & 0 & 3 & 0 \end{bmatrix}$$

When all observations are fully scored using $A^{\max} = (4, 4, \cdots 4)^T$, the maximum values for the number (f_{mny}) and frequency (f_{ofn}) of *OB* presentations for this participant are $f_{mny}^{\max} = 5$ and $f_{ofn}^{\max} = 428$, respectively, according to the scoring matrix. In order to facilitate the unification of competency evaluation criteria, this paper normalizes and takes the relative norm to obtain $\overline{f_{mny}} = \frac{5}{5} = 1, \overline{f_{ofn}} = \frac{240}{428} = 0.56$.

(4): Optimization results of competency evaluation criteria

Through repeating the above steps, the FPM competency rating matrix of 74 sampled trainees was obtained, and the relative values of *OB*, which show the quantity and frequency, were derived. The VENN criterion shows that the *OB*-based evaluation is determined through both frequency and quantity, i.e., the minimum value of both is taken as the *OB*-based evaluation. According to Section 3.4, the evaluation model is solved considering the non-linearity of the approximation model; thus, this paper adopts a grid-based search method to derive the optimal approximation solution, as shown in Table 7, where the evaluation results for the trainees are shown in Appendix B.

Table 7. FPM competency grading criteria based on frequency and number of *OB* presentations.

Grade	1	2	3	4
***OB* frequency classification interval ($\overline{f_{ofn}}$)**	[0, 0.56]	(0.56, 0.71]	(0.71, 0.86]	(0.86, 1]
***OB* number of classification interval ($\overline{f_{mny}}$)**	/	/	/	1

Notes: $\overline{f_{ofn}}$ represents numerical division of the four levels of a specific *OB*; $\overline{f_{mny}}$ represents specific score of student in batch showing *OB*; / represents no score in level range; 1 represents in level 4. *OB* is shown in full and level is four.

4.2. Comparing Evaluation Results

Based on the evaluation criteria shown in Table 6, the FPM competency rating of the examinee based on *OB* can be obtained. To verify the feasibility of the competency rating scheme designed in this paper, the rating is tested for consistency with the rating given by the examiner. Firstly, the monotonic relationship with the data was verified via SPSS data analysis software. On the basis of the monotonic relationship, the Spearman's rank correlation coefficient was calculated using a non-parametric hypothesis test to analyse the correlation between the *OB*-based ratings and the surveyors' ratings. The correlation coefficient was 0.846 (as shown in Table 8), which was obtained via testing whether there was a correlation between the two variables from the perspective of whether they were synergistically consistent. The results indicated that there was a significant correlation between the *OB*-based scores and the surveyor scores.

Table 8. Spearman's correlation analysis results.

_	-	Based on *OB* Rating	Examiner Rating
Based on *OB* rating	Correlation coefficient	1.000	0.846
	Significance	–	0.000
	number	19	19

Furthermore, the statistical analysis of the performance of the participants in the screening check can be carried out according to the steps shown in the previous section, and the results of the evaluation are shown in Table 9.

Table 9. Test sample rating results.

Sample Serial No.	f_{mny}	$\overline{f_{mny}}$	p_{mny}	f_{ofn}	$\overline{f_{ofn}}$	p_{ofn}	p_{ob}	p_{exa}
1	5	1	4	307	0.717	2	2	3
2	5	1	4	276	0.645	2	2	2
3	5	1	4	329	0.769	3	3	3
4	5	1	4	341	0.797	3	3	3
5	5	1	4	315	0.736	3	3	3
6	5	1	4	310	0.724	3	3	3
7	5	1	4	307	0.717	2	2	3
8	5	1	4	403	0.942	4	4	4
9	5	1	4	273	0.638	2	2	2
10	5	1	4	342	0.799	3	3	3
11	5	1	4	310	0.724	3	3	3
12	5	1	4	255	0.596	2	2	1
13	5	1	4	299	0.699	2	2	2
14	5	1	4	328	0.766	3	3	3
15	5	1	4	237	0.554	1	1	1
16	5	1	4	320	0.748	3	3	3
17	5	1	4	304	0.710	2	2	2
18	5	1	4	321	0.600	3	3	3
19	5	1	4	306	0.572	3	3	3

According to Table 9, after further comparing the similarities and differences between *OB*-based ratings and examiner ratings, a comparative analysis of the two ratings can be performed. According to Figure 2, 84% of the samples based on *OB* ratings are exactly the same as examiner ratings, and 16% of the samples have rating differences within one level. Level 1 deviation indicates the acceptable range of deviation in the rating. This outcome is because examiner ratings are subjective and there is some uncertainty about the boundaries between two adjacent levels.

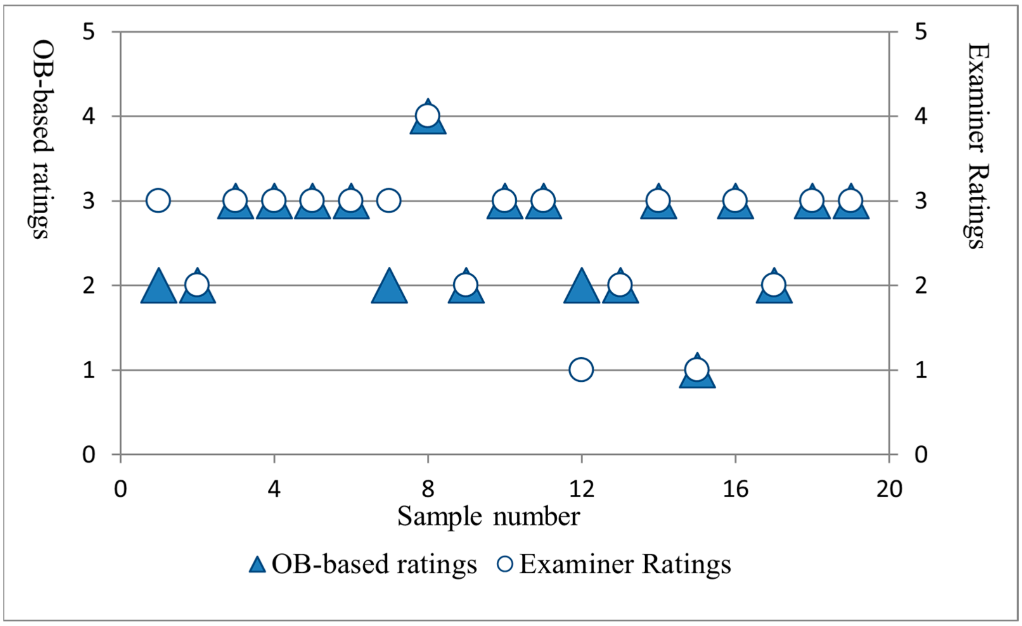

Figure 2. *OB* rating vs. examiner rating analysis.

It can be seen that through constructing a competency assessment matrix to measure the number and frequency of observable items within an acceptable range of assessment bias, a quantitative assessment of competencies can be achieved with a high degree of consistency. This result fully validates the effectiveness of the proposed competency assessment model.

5. Conclusions

Through integrating experts' experience, constructing a typical subject observation and *OB* correlation matrix, and establishing a competency optimization evaluation model based on the VENN criterion, this research solves the shortcomings of the traditional initial flight training evaluation procedure based on the traditional flight training work order evaluation model. The conclusions are as follows:

(1) The optimized Training Assessment Worksheet highlights the core competencies for manual control at this screening stage. The specific behavioral indicators in the subject under this competency are presented in the form of a Training Assessment Worksheet, which allows a straightforward correlation between the behavioral indicators and the competency items, resulting in a more refined and scientific quantitative assessment, and providing important data support for the subsequent targeted training of trainees.

(2) Through combining the data of flight trainees in the screening stage of the case, the optimal solution of the objective function was obtained, the threshold of the optimal skill evaluation model was derived according to the steps of the evaluation model, and the skill evaluation criteria based on behavioral indicators could be further obtained. In addition, test samples were selected to validate the scheme, and the results showed that 84% of the 19 test samples agreed with the examiner's scores based on the above skill evaluation criteria, thus validating the feasibility of the scheme.

(3) An optimized evaluation scheme of competency assessment criteria for the initial flight training phase is designed. For the student, this scheme provides a quantitative assessment of the quality of flight training and a competency level for this phase, which can provide suggestions and directions for subsequent targeted training improvements; for the flight instructor, the use of the new Training Assessment Worksheet provides the ability to quantify the assessment and track the data, facilitating the implementation of "individualized" training for students.

(4) In the whole competency evaluation model and evaluation scheme study, the subject-based teaching organization characteristics of initial flight training are well utilized. On one hand, the traditional subject-based assessment is continued; on the other hand, the shortcomings of subject-based assessment, which seems to be generalized and not refined enough, are improved, and core competencies are added to assess the overall training quality of trainees, thus providing a comprehensive picture of trainees' competencies. In addition, this scheme can be extended to the CBTA assessment of all phases of initial flight training, such as the instrument rating training phase, commercial pilot license training, etc. The difference is that the corresponding assessment worksheet, assessment matrix, and associated competency rating model must be designed according to the instructional requirements and characteristics of each phase of the training course.

Author Contributions: Conceptualization, F.Y. and H.S.; methodology, F.Y. and P.Z.; software, F.Y. and P.Z.; validation, F.Y. and Q.H.; formal analysis, F.Y. and H.S.; writing—original draft preparation, F.Y.; writing—review and editing, F.Y. and P.Z.; supervision, H.S.; project administration, H.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (U2033213); supported by "the Fundamental Research Funds for the Central Universities": (FZ2021ZZ01).

Institutional Review Board Statement: Informed consent was obtained from all subjects involved in the study.

Informed Consent Statement: Not applicable.

Data Availability Statement: All the data contained in this study can be obtained upon request from the corresponding author. Readers can also request part of the original data and the results of data processing outlined in this paper.

Acknowledgments: We would like to thank those who contributed to our research. We are particularly grateful to flight instructors Yunsong Lu, Hong Huang and Wuyang Song from the Civil Aviation Flight University of China for their expert support and validation work in this research.

Conflicts of Interest: The authors declare that they have no conflict of interest to report regarding the present study.

Appendix A

Table A1. Part of the assessment details.

Subject (Sub)	Observation (No)	Scoring Criteria
Up down	Direction of navigation	4: Remains accurate. 3: Within ±5°. 2: Within ±10°. 1: Beyond ±10°.
	Speed	4: Remains accurate. 3: Within ±5 knts. 2: Within ±10 knts. 1: Beyond ±10 knts.
Horizontal flight	Direction of navigation	4: Remains accurate. 3: Points: within ±5 knts. 2: Points: within ±10 knts. 1: Point: beyond ±10 knts.
	Speed	4: Remains accurate. 3: Within ±5 knts. 2: Within ±10 knts. 1: Beyond ±10 knts
	Height	4: Within ±15 ft. 3: Within ±30 ft. 2: Within ±50 ft. 1: Outside ±50 ft
Swerve	Slope	4: Remains accurate. 3: ±2° or less. 2: Within ±5°. 1: Beyond ±5°.
	Compatibility	4: Maintained accuracy without side slippage. 3: Within half a frame. 2: More than half a frame. 1: More than one frame.
	Speed	4: Within +5 knts. 3: Within +10/−5 knts. 2: Within +15/−10 knts. 1: Outside +15/−10 knts
	Change course	4: Within ±2°. 3: Within ±4°. 2: Within ±6°. 1: Outside ±6°

Table A1. *Cont.*

Subject (Sub)	Observation (No)	Scoring Criteria
Grounding gesture	Pulling start height	4: Conform to the regulations. 3: Within 1 m. 2: Within 2 m. 1: Beyond 2 m.
	Leveling height	4: Conform to the regulations. 3: Within 0.25 m, slightly pulled, corrected. 2: Within 0.5 m, slightly pulled, corrected. 1: Within 0.5 m.
	Grounding gesture	4: Three points smoothly earthed. 3: Slightly tilted or slightly heavily earthed, but no secondary earthed. 2: Jumps of up to 0.25 m or more and pronounced tilting when earthed, corrected. 1: Mark: jumps of more than 0.25 m when earthed, corrected

Appendix B

Table A2. Results of trainee ratings.

Sample Serial No.	f_{mny}	$\overline{f_{mny}}$	p_{mny}	f_{ofn}	$\overline{f_{ofn}}$	p_{ofn}	p_{ob}	p_{exa}
1	5	1	4	240	0.561	1	1	1
2	5	1	4	303	0.708	2	2	2
3	5	1	4	326	0.762	3	3	2
4	5	1	4	299	0.699	2	2	3
5	5	1	4	310	0.724	3	3	2
6	5	1	4	312	0.729	3	3	3
7	5	1	4	332	0.776	3	3	3
8	5	1	4	301	0.703	2	2	3
9	5	1	4	342	0.799	3	3	3
10	5	1	4	312	0.729	3	3	2
11	5	1	4	280	0.654	2	2	2
12	5	1	4	353	0.825	3	3	3
13	5	1	4	328	0.766	3	3	3
14	5	1	4	331	0.773	3	3	3
15	5	1	4	282	0.659	2	2	3
16	5	1	4	328	0.766	3	3	2
17	5	1	4	282	0.659	2	2	3
18	5	1	4	336	0.785	3	3	3
19	5	1	4	291	0.680	2	2	2
20	5	1	4	357	0.834	3	3	3
21	5	1	4	365	0.853	3	3	3
22	5	1	4	321	0.750	3	3	2
23	5	1	4	300	0.701	2	2	2
24	5	1	4	318	0.743	3	3	3
25	5	1	4	280	0.654	2	2	2
26	5	1	4	339	0.792	3	3	3

Table A2. *Cont.*

Sample Serial No.	f_{mny}	$\overline{f_{mny}}$	p_{mny}	f_{ofn}	$\overline{f_{ofn}}$	p_{ofn}	p_{ob}	p_{exa}
27	5	1	4	277	0.647	2	2	2
28	5	1	4	367	0.857	3	3	3
29	5	1	4	305	0.713	3	3	3
30	5	1	4	308	0.720	3	3	3
31	5	1	4	321	0.750	3	3	3
32	5	1	4	342	0.799	3	3	3
33	5	1	4	262	0.612	2	2	2
34	5	1	4	333	0.778	3	3	3
35	5	1	4	337	0.787	3	3	3
36	5	1	4	353	0.825	3	3	3
37	5	1	4	345	0.806	3	3	3
38	5	1	4	310	0.724	3	3	2
39	5	1	4	219	0.512	1	1	1
40	5	1	4	327	0.764	3	3	2
41	5	1	4	316	0.738	3	3	2
42	5	1	4	313	0.731	3	3	3
43	5	1	4	325	0.759	3	3	2
44	5	1	4	280	0.654	2	2	2
45	5	1	4	237	0.554	1	1	1
46	5	1	4	321	0.750	3	3	3
47	5	1	4	302	0.706	2	2	2
48	5	1	4	331	0.773	3	3	3
49	5	1	4	378	0.883	4	4	3
50	5	1	4	313	0.731	3	3	1
51	5	1	4	331	0.773	3	3	2
52	5	1	4	282	0.659	2	2	2
53	5	1	4	276	0.645	2	2	2
54	5	1	4	350	0.818	3	3	2
55	5	1	4	289	0.675	2	2	2
56	5	1	4	328	0.766	3	3	2
57	5	1	4	294	0.687	2	2	2
58	5	1	4	324	0.757	3	3	3
59	5	1	4	330	0.771	3	3	3
60	5	1	4	311	0.727	3	3	3
61	5	1	4	349	0.815	3	3	3
62	5	1	4	342	0.799	3	3	4
63	5	1	4	303	0.708	2	2	2
64	5	1	4	334	0.780	3	3	3
65	5	1	4	331	0.773	3	3	2
66	5	1	4	336	0.785	3	3	2

Table A2. *Cont.*

Sample Serial No.	f_{mny}	$\overline{f_{mny}}$	p_{mny}	f_{ofn}	$\overline{f_{ofn}}$	p_{ofn}	p_{ob}	p_{exa}
67	5	1	4	317	0.741	3	3	2
68	5	1	4	330	0.771	3	3	2
69	5	1	4	249	0.582	2	2	2
70	5	1	4	306	0.715	3	3	3
71	5	1	4	323	0.755	3	3	3
72	5	1	4	314	0.734	3	3	2
73	5	1	4	402	0.939	4	4	4
74	5	1	4	376	0.879	4	4	4

References

1. Ouyang, T.; Sun, H.; Li, F. Researches on the Education Reform for the Core Competencies -oriented Flight Training of Civil Aviation Pilots. In Proceedings of the Proceedings of 2021 6th International Conference on Education Reform and Modern Management (ERMM2021), Beijing, China, 11 April 2021; pp. 399–402.
2. Holbrook, J. Exploring methods to collect and analyze data on human contributions to aviation safety. In Proceedings of the 49th International Symposium on Aviation Psychology, Online, 5 January 2021; pp. 110–115.
3. Rose, R.L.; Puranik, T.G.; Mavris, D.N. Natural language processing based method for clustering and analysis of aviation safety narratives. *Aerospace* **2020**, *7*, 143. [CrossRef]
4. Pate, J.; Adegbija, T. AMELIA: An application of the Internet of Things for aviation safety. In Proceedings of the 2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 12–15 January 2018; pp. 1–6.
5. Wang, L.; Zhang, J.; Dong, C.; Sun, H.; Ren, Y. A method of applying flight data to evaluate landing operation performance. *Ergonomics* **2019**, *62*, 171–180. [CrossRef] [PubMed]
6. Feigl, F. *Combination of ADS-B and QAR Data for Mid-Air Collision Analysis*; Technische Universität München: München, Germany, 2018.
7. Li, W.-C.; Nichanian, A.; Lin, J.; Braithwaite, G. Investigating the impacts of COVID-19 on aviation safety based on occurrences captured through Flight Data Monitoring. *Ergonomics* **2022**, *2022*, 1–39. [CrossRef] [PubMed]
8. Gavrilovski, A.; Jimenez, H.; Mavris, D.N.; Rao, A.H.; Shin, S.; Hwang, I.; Marais, K. Challenges and opportunities in flight data mining: A review of the state of the art. In Proceedings of the AIAA SciTech Forum, San Diego, CA, USA, 4–8 January 2016.
9. Wang, L.; Wu, C.; Sun, R.; Cui, Z. An analysis of hard landing incidents based on flight QAR data. In Proceedings of the Engineering Psychology and Cognitive Ergonomics: 11th International Conference, EPCE 2014, Heraklion, Crete, Greece, 22–27 June 2014; pp. 398–406.
10. Wang, L.; Wu, C.; Sun, R. Pilot operating characteristics analysis of long landing based on flight QAR data. In Proceedings of the Engineering Psychology and Cognitive Ergonomics. Applications and Services: 10th International Conference, EPCE 2013, Las Vegas, NV, USA, 21–26 July 2013; pp. 157–166.
11. Wang, L.; Wu, C.; Sun, R. An analysis of flight Quick Access Recorder (QAR) data and its applications in preventing landing incidents. *Reliab. Eng. Syst. Safety* **2014**, *127*, 86–96. [CrossRef]
12. Thomas, L.C.; Gast, C.; Grube, R.; Craig, K. Fatigue detection in commercial flight operations: Results using physiological measures. *Procedia Manuf.* **2015**, *3*, 2357–2364. [CrossRef]
13. Shao, S.; Zhou, Q.; Liu, Z. A new assessment method of the pilot stress using ECG signals during complex special flight operation. *IEEE Access* **2019**, *7*, 185360–185368. [CrossRef]
14. Jun, C.; Lei, X.; Jia, R.; Xudong, G. Real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors. *Chin. J. Aeronaut.* **2022**, *35*, 450–463.
15. Mansikka, H.; Harris, D.; Virtanen, K. An input–process–output model of pilot core competencies. *Aviat. Psychol. Appl. Human Factors* **2017**, *7*, 78–85. [CrossRef]
16. Liu, S.; Zhang, Y.; Chen, J. A system for evaluating pilot performance based on flight data. In Proceedings of the Engineering Psychology and Cognitive Ergonomics: 15th International Conference, EPCE 2018, Las Vegas, NV, USA, 15–20 July 2018; pp. 605–614.
17. Sembiring, J.; Drees, L.; Holzapfel, F. Extracting unmeasured parameters based on quick access recorder data using parameter-estimation method. In Proceedings of the AIAA Atmospheric Flight Mechanics (AFM) Conference, Boston, MA, USA, 19–22 August 2013; p. 4848.
18. Wang, L.; Ren, Y.; Sun, H.; Dong, C. A landing operation performance evaluation system based on flight data. In Proceedings of the Engineering Psychology and Cognitive Ergonomics: Cognition and Design: 14th International Conference, EPCE 2017, Vancouver, BC, Canada, 9–14 July 2017; pp. 297–305.

19. Lahtinen, T.M.; Koskelo, J.P.; Laitinen, T.; Leino, T.K. Heart rate and performance during combat missions in a flight simulator. *Aviat. Space Environ. Med.* **2007**, *78*, 387–391. [PubMed]
20. Maciejewska, M.; Galant-Gołębiewska, M. Case study of pilot's Heart Rate Variability (HRV) during flight operation. *Transp. Res. Procedia* **2021**, *59*, 244–252. [CrossRef]
21. Jirgl, M.; Jalovecky, R.; Bradac, Z. Models of pilot behavior and their use to evaluate the state of pilot training. *J. Electr. Eng.* **2016**, *67*, 267–272. [CrossRef]
22. Sarkar, S. Competency based training need assessment–approach in Indian companies. *Organizacija* **2013**, *46*, 253–263. [CrossRef]
23. Chen, J.; Xue, L.; Liu, Z. A pilot workload evaluation method based on EEG data and physiological data. In Proceedings of the 2020 IEEE international conference on signal processing, communications and computing (ICSPCC), Macau, China, 21–24 August 2020; pp. 1–6.
24. Dehais, F.; Causse, M.; Pastor, J. Embedded eye tracker in a real aircraft: New perspectives on pilot/aircraft interaction monitoring. In Proceedings of the 3rd International Conference on Research in Air Transportation, Fairfax, VA, USA, 1–4 June 2008.
25. Arana, R. Horizontalização na Observação de Habilidades Não-Técnicas nos Treinamentos em Simulador. Available online: https://repositorio.animaeducacao.com.br/handle/ANIMA/13557 (accessed on 16 May 2023).