ORIGINAL ARTICLE

Competency-based assessment of pilots' manual flight performance during instrument flight training

Hong Sun[1] · Xin Zhou[1] · Peiwen Zhang[2] · Xin Liu[1] · Yunsong Lu[3] · Hong Huang[3] · Wuyang Song[3]

Abstract

A pilot's control behavior has a direct impact on aviation safety. To adapt to new flight training reforms, a competency-based assessment method is proposed in this paper. First, assessment scenarios, key competencies and observable behaviours were determined. Then, observable behavior assessment indexes were developed based on standard operating procedures and practical examination standards. Flight training data were analyzed based on curve similarity theory and a time series fast retrieval algorithm, and the assessment indexes were scored. The competency assessment method was established by quantifying "HOW MANY" and "HOW OFTEN" observable behaviors according to the assessment indexes score results. Finally, the fuzzy C-means clustering algorithm was used to classify the assessment values, and assessment results were obtained. A teardrop pattern procedure was used to test the reliability of the assessment method. The results indicate that the correlation coefficient between the model evaluation and the expert evaluation is 0.947 with a significance level of 0.01, reliably and objectively quantifying manual flight performance and improving the scientific nature of manual flight performance assessment methods. Flight training institutions can use this method to establish diverse assessment scenarios and formulate corresponding assessment indexes based on the characteristics of each pilot to reasonably evaluate the manual flight performance of pilot. Targeted training methods can then be adopted for pilot with different performance levels to improve their competency.

Keywords Flight training · Competency · Assessment method · Control behaviour

1 Introduction

Approximately one-fifth of commercial air transport large aeroplane accident and serious incident reports identify human factors or human performance issues as key causes (EASA 2022). Loss of Control In-flight (LOC-I) remains one of the most significant contributors to fatal accidents worldwide. Due to the development of automation and system integration in modern civil aircraft cockpits, previous training methods focused on repeatedly practicing specific situations are no longer sufficient to handle unexpected events. Competency-Based Training and Assessment (CBTA) is expanding across the regulations and the industry (ICAO 2013, 2020). Competency is an indication of human

✉ Xin Zhou
  xinxinzhou219@163.com

  Hong Sun
  2375812292@qq.com

  Peiwen Zhang
  zhangpeiwen@cafuc.edu.cn

  Xin Liu
  1052530247@qq.com

  Yunsong Lu
  30498737@qq.com

  Hong Huang
  578975610@qq.com

  Wuyang Song
  25540512@qq.com

[1] Present Address: Civil Aviation Flight Technology and Flight Safety Research Base, Civil Aviation Flight University of China, No. 46, Section 4, Nanchang Road, Luocheng Street, Guanghan, Deyang 618307, Sichuan, China

[2] College of Economic and Management, Civil Aviation Flight University of China, Guanghan 618307, China

[3] Guanghan Branch College, Civil Aviation Flight University of China, Guanghan 618307, China

performance that can be used to reliably predict successful job performance. Since its proposal in 1973 (McClelland 1974), the method has been widely applied in personnel selection, training assessment, and training course design within the medical field (Song et al. 2022; Chen et al. 2022; Alidrisi and Mohamed 2022; Vaughn et al. 2021; Zhang et al. 2022; Hashemian et al. 2021).The Governments, the airlines, and the training organizations have realized that CBTA is a far more efficient way to develop a competent pilot workforce when compared to the traditional task—or hour-based training and checking. The CBTA concept identifies nine core competencies that pilots should possess, namely application of knowledge, application of procedures and compliance with regulations, management of aircraft flight path (automation), management of aircraft flight path (manual control), communication, leadership and teamwork, workload management, problem-solving and decision-making skills, situation awareness and information management. International Air Transport Association (IATA) provides a general framework for recognizing the observable behaviors (OBs) of various abilities (IATA 2021). Application by the world's advanced airlines proves that a flight training system based on competency can break through the separation between operation and training, realizing the purpose of training for operational service and operation for training inspection. The transformation of the civil aviation flight training mode has become an inevitable trend. Evaluating the quality of flight training is an important link in pilot training, so proposing improvements to competency-based flight training assessment systems is of great significance.

Traditional flight performance assessment methods only rely on the subjective experience of the instructor, and one disadvantage of methods is that the assessment results are easily affected by the instructor's physical or psychological state at the time, leading to inconsistent assessment criteria. Scholars have proposed various assessment methods based on traditional flight training systems (Wu et al. 2018; Zhang et al. 2015; Jeong et al. 2018; Yao and Xu 2017; Chen and Jin 2021). Flight training evaluation indices are selected according to flight maneuvering actions, the index weights are determined using a comprehensive weighting method, and flight training evaluations are conducted based on fuzzy sets and grey relational analysis. This approach improved the efficiency of flight training evaluations (Liu et al. 2021). In a study by Jirgl et al. (2020), statistics based on pilot behaviour model parameters were associated with military flight training data, pilots' abilities to adapt to controlled dynamic systems were characterized, and objective evaluations at different flight training levels were realized. Wang et al. (2019) selected seven typical flight parameters as evaluation indices, organically combined an Analytic Hierarchy Process (AHP) and the Technique for Order Preference by

Similarity to an Ideal Solution (TOPSIS), and established an optimal ranking model so that the assessment results accurately reflected the subjective intention of experts to the greatest extent to realize flight quality assessment. Zhu et al. (2021) selected five flight parameters, such as the flight path angle, altitude and airspeed, as assessment indices and proposed a manual flight performance assessment model based on time series similarity and obtained assessment results that were consistent with expert scores.

To some extent, the above studies can improve traditional manual flight performance assessment methods, which are influenced by subjective judgments and inefficient; however, most of these studies evaluate only the aircraft control accuracy. According to the CBTA assessment concept advocated by the ICAO, manual pilot control ability is fully described from seven OBs, including aircraft attitude control, trajectory safety management, trajectory deviation monitoring and correction. However, existing research does not conform to this concept and cannot be applied to the development trend of the competency-based flight training assessment industry after the reform. From a competency perspective, there are still gaps in the existing research on pilots' manual flight performance.

In view of the above deficiencies, based on the initial flight training phase, in this paper, a teardrop pattern procedure is considered as an assessment scenario, and a competency-based manual flight performance assessment method is proposed. For each competency OB, an observation item based on flight data is designed, and an assessment model is constructed based on two OB dimensions, "HOW MANY" and "HOW OFTEN". The fuzzy C-means clustering method is used to classify the assessment values, and the assessment results are obtained.

2 Methodology

2.1 Participants

The participants in this study included male pilot trainees and flight experts from the same flight training schools. The average age of the trainees was 20 years, the trainees had an average flight time of 87 h, and all trainees had obtained their Private Pilot licenses. Furthermore, the trainees had the proficiency to take the instrument rating practice exam. The trainees participated in the same instrument rating practice exam. For this study, 20 students were selected based on their excellent evaluation results, and another 30 students were randomly chosen. Three flight experts with an average age of 35 years and an average flight time exceeding 10,000 h, as well as extensive experience in flight training teaching, also participated in the research.
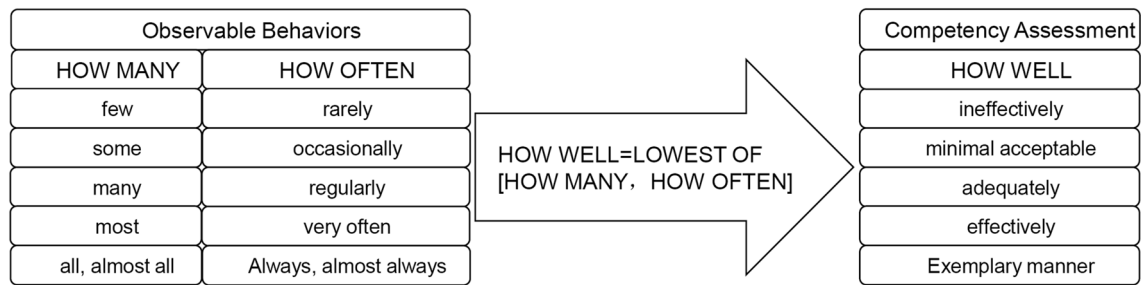
Fig. 1 Competency assessment method

2.2 Experiment design

2.2.1 Assessment scenario

Degraded pilot performance can be explained not by a lack of competence but by availability or ability to act or interact (Frédéric 2021). Thus, competence assessment methods need to be contextualized. A teardrop pattern procedure was set as the assessment scenario in this work, in which a pilot needs to strictly maintain the specified procedure parameters according to the instrument instructions to maintain the obstacle requirement for each segment and establish the required state for landing. Therefore, a teardrop pattern procedure can fully reflect a pilot's manual control ability. This study takes it as an assessment scenario to verify the reliability of the assessment method. A teardrop pattern procedure has an angle between the initial approach segment and the final approach segment (ϕ). Its segment composition is shown in Fig. 1, where r is the turning radius, IAF is the initial approach fix, IF is the intermediate approach fix, FAF is the final approach fix, and $\theta = 180 + \phi$.

After passing the IAF at the specified height, the aircraft will join the teardrop pattern procedure, fly along the back of the outbound leg, reach the anchor point of the turn, control the start time of the turn, make a baseline turn, and cut into the inbound leg, creating favourable conditions for the intermediate and final approach segments. In the intermediate and final approach segments, it is necessary to further adjust the landing configuration and speed of the aircraft, strictly maintain the flight path and altitude to establish appropriate create good landing conditions, and switch to visual landing when obtaining visual reference; otherwise, it is necessary to perform missed approach procedure. The teardrop pattern procedure in this study is a precision approach procedure, which can provide localizer and glide path information for the aircraft in the final approach segment and guide the aircraft to descend and land along predetermined glide path. During the final approach, the aircraft first intercepts the heading path by levelling off, aligns with the runway, and then intercepts the glide path and begins its descent.

2.2.2 Key competency

During flight training assessment, the key competencies that pilots need to improve is evaluated, which can be any one or several of the nine core competencies. In the initial flight training stage, the main purpose is to enable pilots to master basic flight skills and then manually control the flight path, so the key competency to be evaluated is flight path management, manual control. This competency consists of seven OBs, as shown in Table 1.

Among them, OB3 and OB5 are advanced OBs. After pilots master basic flight skills, they need to be evaluated in combination with more nontechnical factors, and the initial flight training stage is not considered. At present, the main training aircraft in use include single-engine aircraft such as the Cessna 172 and DA20 and multi-engine aircraft such as the PA44 and DA42, which basically do not involve automatic pilot systems, so OB6 and OB7 are not considered. Therefore, the assessment focuses on OB1 (aircraft attitude control), OB2 (trajectory deviation monitoring and correction) and OB4 (trajectory safety management) in the initial flight training stage, namely, basic OBs. Specific quantitative assessment standards can be obtained from flight training data.

Table 1 OBs and specific requirements

OB	Specific requirement
OB1	Control the aircraft manually with accuracy and smoothness as appropriate to the situation
OB2	Monitor and detect deviations from the intended flight path and takes appropriate action
OB3	Manually control the aircraft using the relationship between the attitude, speed and thrust, and navigation signals or visual information
OB4	Manage the flight path safely to achieve optimum operational performance
OB5	Maintain the intended flight path during manual flight while managing other tasks and distractions
OB6	Use appropriate flight management and guidance systems, as installed and applicable to the conditions
OB7	Effectively monitor flight guidance systems including engagement and automatic mode transitions

2.2.3 Competency assessment method

CBTA evolved from later developments in mastery learning and criterion-referenced testing, whereby knowledge and skills had to be demonstrated at levels that met the entry-level occupational requirements, and assessments had to be based on observable behaviours or outcomes. In this study, the assessment method was established by referencing observable behavioral dimensions in the VENN model (IATA 2021). This facilitates the consistency and objectivity of the evaluation results to the greatest extent.

The relevant OB for each capability is assessed based on the following dimensions:

1. "HOW MANY" OBs the pilot trainee demonstrates when required, providing evidence related to having acquired competency;
2. "HOW OFTEN" the pilot trainee demonstrates the OB(s) when required, providing evidence related to competency robustness.

The competency assessment (HOW WELL) is the combination of the number of OBs demonstrated and their frequency of demonstration relating specifically to the competency being assessed. The assessment process is shown in Fig. 2.

2.2.4 Design of OB assessment indexes

ICAO's description of OB is very general, with a large degree of freedom. If the assessors directly reference the OBs in their evaluation of the pilot's manual flight performance, the assessment outcomes may be subjective, potentially leading to divergent conclusions among different assessors. To address this issue, in this study, OBs were combined with standard operating procedures and practical examination standards, and specific evaluation indices were proposed and scored. Taking the outbound flight phase of the teardrop pattern procedure as an example, three evaluation indices were determined for aircraft attitude control: pitch control, roll control, and track control. Two evaluation indices were identified for trajectory deviation monitoring and correction: the timeliness of the trajectory deviation correction and altitude deviation correction. For trajectory safety management, four evaluation indices were identified: the pitch attitude change rate within a safe range, the roll angle within a safe range, the descent rate within a safe range, and the indicated airspeed within a safe range.

When a pilot trainee meets the requirements based on the evaluation indices, they are considered to have demonstrated the corresponding OBs. This method can not only enhance the measurability of specific competencies and facilitates the fairness and objectivity of the results but also increase the credibility of the evaluation results so that pilot cadets can reach the required competency level.

Considering the differences in the importance of different assessment indexes, an actual flight training evaluation is on a four-point scale, so two scoring criteria are defined, as shown in Table 2.

Furthermore, the scoring vector (A) of all assessment indexes is obtained:

$$\mathbf{A} = \left(a_i\right)_{1 \times I} = (a_1, \dots, a_i, \dots, a_I), \tag{1}$$

where a_i is the score of the ith assessment index, $1 \le i \le I$, and a_i^{max} is the full score of this assessment index. When all assessment indexes receive the highest possible score, the full score of the scoring vector ($\mathbf{A}^{max}$) can be obtained:

Table 2 Assessment index classification and scoring criteria

Classification	Scoring criteria
Check item	1: can meet the requirements of the observation item
	0: cannot meet the requirements of the observation item
Key item	4: effectively for OB evaluation
	3: acceptably for OB evaluation
	2: marginally for OB evaluation
	1: ineffectively for OB evaluation

Fig. 2 Flight segment composition of a teardrop pattern procedure
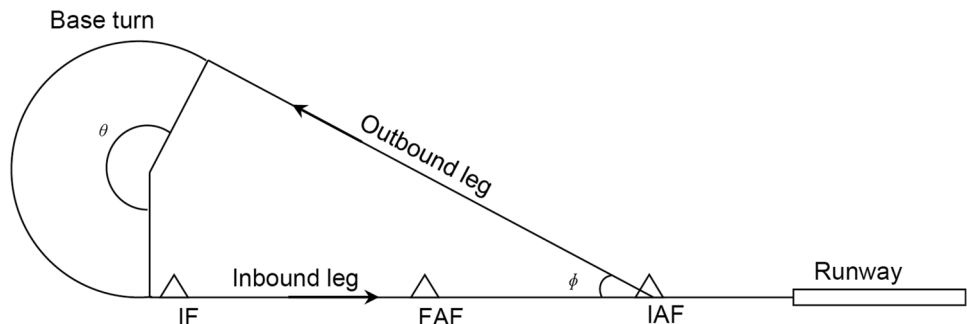
$$\mathbf{A}^{\max} = \left(a_i\right)_{1 \times I}^{\max} = \left(a_1^{\max}, \dots, a_i^{\max}, \dots, a_I^{\max}\right). \tag{2}$$

2.3 Data analysis

2.3.1 Data acquisition

In this study, live flight data generated by a cohort of flight students who participated in the instrument rating practice exam, all of whom flew a Cessna 172 aircraft, were utilized. The data were stored on a secure digital memory card located onboard the aircraft, and various parameters, including the time, altitude, wind speed, and descent rate, were recorded at a sampling frequency of one sample per second.

2.3.2 Data pre-processing

First, the data were filtered to include only parameters that were relevant to the research content, such as the longitude, latitude, pitch, roll and indicated airspeed. Second, based on the initial approach fix (IAF), intermediate fix (IF) and final approach fix (FAF) latitude and longitude positions, as well as the characteristics of the teardrop pattern procedure, the flight phase data required for the research purposes were retained. Finally, parameter acquisition errors or other issues may lead to incorrect or missing information in certain data fields; thus, these fields need to be filtered, or completion operations need to be performed to address missing data. All the aforementioned preprocessing and subsequent data analysis processes were completed in a Python environment using Visual Studio Code.

2.3.3 Quantification of OB assessment indexes

Flight training is a dynamic process, and there is no standard curve. However, based on the large sample flight data statistics, some key flight parameters, such as the aircraft roll and pitch angles, the time change curve trends have essentially the same trend. Therefore, under the restrictions of several conditions, such as airport conditions, aircraft types, and wind speed, a target curve can be obtained through data screening and fitting. Therefore, curve similarity theory can be used to quantify the assessment indexes that best reflect the accuracy of a pilot student's aircraft maneuvering ability. The Fréchet distance is a spatial path similarity description method proposed by French mathematician Maurice Rene Fréchet in 1906. The feature of this method is that it focuses on considering the similarity of the curve space distance and has high efficiency for evaluating curve similarity with time series (Chan and Rahmati 2018; Guo et al. 2017; Agarwal et al. 2014).

Let U and V be two continuous curves in S, i.e., U: $[0,1] \rightarrow S$, V: $[0,1] \rightarrow S$ and alpha and beta be two reparametrized functions of the unit interval, i.e., α: $[0, 1] \rightarrow [0, 1]$, β: $[0, 1] \rightarrow [0, 1]$; then, the Fréchet distance $F(U, V)$ for curves U and V is defined as:

$$F(U, V) = \inf_{\alpha, \beta} \max_{t \in [0,1]} \{d(U(\alpha(t)), V(\beta(t)))\}, \tag{3}$$

where d is the metric function on S.

Based on the above basic Fréchet distance idea, this study adopts a discrete Fréchet distance algorithm suitable for computers to describe the distance between the actual flight parameter curve and the target curve. The specific implementation process is as follows:

1. The actual flight parameter curve L_1 can be expressed as:

$$P = \{P(1), P(2), \dots, P(n), \dots, P(N)\}, \tag{4}$$

where $P(n) = \left(x_n, y_n\right)$ and n is the sequence number of flight parameter data points on curve L_1. $n = 1$ is the initial sampling point, and $n = N$ is the final sampling point. x_n is the abscissa of the nth sampling point, and y_n is the ordinate of the nth sampling point.

2. The target curve L_2 can be expressed as:

$$P' = \left\{P'(1), P'(2), \dots, P'(m), \dots, P'(M)\right\}, \tag{5}$$

where $P(m) = \left(x'_m, y'_m\right)$ and m is the flight parameter data point sequence number on curve L_2. $m = 1$ is the initial sampling point, and $m = M$ is the final sampling point. x'_m is the abscissa of the mth sampling point, and y'_m is the ordinate of the mth sampling point.

3. The distance between flight parameter points on L_1 and flight parameter points on L_2 is calculated, and the distance matrix (D) is obtained as follows:

$$\mathbf{D} = \begin{bmatrix} d_{21} & \cdots & d_{1N} \\ \vdots & & \vdots \\ d_{M1} & \cdots & d_{MN} \end{bmatrix}, \tag{6}$$

where $d_{mn} = \sqrt{\left(x'_m - x_n\right)^2 + \left(y'_m - y_n\right)^2}$ represents the distance from the mth flight parameter point on L_2 of the target curve to the nth sampling point on L_1 of the actual flight parameter curve, $1 \leq m \leq M, 1 \leq n \leq N$.

4. The maximum distance $d_{\max} = \max(\mathbf{D})$ and the minimum distance $d_{\min} = \min(\mathbf{D})$ in the distance matrix D are found, the target distance $f = d_{\min}$ is initialized, and the loop distance (d_{loop}) is set:

$$d_{\text{loop}} = \frac{d_{\max} - d_{\min}}{100}, \tag{7}$$

5. The elements less than or equal to f in the distance matrix ($\mathbf{D}$) are set to 1, and those greater than f are set to 0, thus obtaining the binary matrix ($\mathbf{D}'$) as follows:

$$\mathbf{D}' = \begin{bmatrix} d'_{11} & \cdots & d'_{1N} \\ \vdots & & \vdots \\ d'_{M1} & \cdots & d'_{MN} \end{bmatrix} \tag{8}$$

where $d'_{mn} = \begin{cases} 1, & d'_{mn} \leq f \\ 0, & d'_{mn} > f \end{cases}$, $1 \leq m \leq M, 1 \leq n \leq N$.

6. In the binary matrix ($\mathbf{D}'$), the path Q that satisfies the following conditions is found: the initial point of Q is d'_{11}, the final point of Q is d'_{MN}, after a path passes through a point d'_{mn}, its next passing point can only be one of $d'_{(m+1)n}, d'_{m(n+1)}$ or $d'_{(m+1)(n+1)}$, and all points on the path Q must have a value of 1. As a mathematical expression, $\forall Q = \{d'_{11}, \ldots, d'_{mn}, \ldots, d'_{MN}\}$ satisfies:

$$d'_{11} \times \cdots \times d'_{mn} \times d'_{(m+k)(n+k\prime)} \times \cdots \times d'_{MN} = 1, \tag{9}$$

$$1 \leq m \leq M, 1 \leq n \leq N, 1 \leq m+k \leq M, 1 \leq n+k\prime \leq N, k$$

where $= \{0,1\}, k\prime = \{0,1\}$.

7. If no path satisfying the conditions is found in step (6), the target distance is set to $f = f + d_{\text{loop}}$ and steps (5) and (6) are repeated. If the path or target distance $f = d_{\text{max}}$ satisfying the conditions is found in step (6), proceed to the next step.

8. The Fréchet distance between the actual curve of the flight parameters and the target curve is $F = f$.

2.3.4 Representation of the assessment model

Each observation item corresponds to an OB. By consulting flight experts, the mapping relationship between each observation item and OB is obtained, and a correlation matrix ($\mathbf{B}$) is constructed.

$$\mathbf{B} = \left[\mathbf{B_1}, \ldots, \mathbf{B}_j, \ldots, \mathbf{B}_J\right] = \begin{bmatrix} b_{11} & \cdots & b_{1J} \\ \vdots & & \vdots \\ b_{I1} & \cdots & b_{IJ} \end{bmatrix}, \tag{10}$$

where b_{ij} represents the correlation attribute between the ith observation item and the jth OB, $1 \leq i \leq I, 1 \leq j \leq J$. When $b_{ij} = 1$, there is a mapping relationship between the ith observation item and the jth OB; otherwise, $b_{ij} = 0$.

According to the VENN model, a pilot trainee's competency level can be measured by the number and frequency of OBs shown in the assessment, the scoring vector ($\mathbf{A}$) and correlation matrix ($\mathbf{B}$) are used, and the key competency assessment matrix ($\mathbf{Y}$) is:

$$\mathbf{Y} = \left[\mathbf{Y_1}, \ldots, \mathbf{Y}_j, \ldots, \mathbf{Y}_J\right] = \begin{bmatrix} a_1 b_{11} & \cdots & a_1 b_{1J} \\ \vdots & & \vdots \\ a_I b_{I1} & \cdots & a_I b_{IJ} \end{bmatrix}, \tag{11}$$

where $a_i b_{ij}$ represents the contribution level of the ith observation item to the jth OB.

Using the norm of the vector/matrix to measure the length of the vector (or matrix) space, the number and frequency of OB presentation can be represented by the norm of the $\mathbf{Y}$ matrix. Each OB corresponds to multiple observations. Whether an OB is displayed is determined based on the scores of the associated assessment indexes. Thus, based on the opinions of flight experts, we believe that in existing flight training evaluation systems, if the trainee's performance level for an OB is less than 25% of the total score, the OB was not performed. If more than 25% of the assessment indexes are valid, the OB is considered to have been performed. The frequency is the number of valid assessment indexes. The number (f_{mny}) and frequency (f_{ofn}) of OB presentations based on the competency assessment matrix can be obtained by calculating the norm of the assessment matrix:

$$f_{\text{mny}} = \text{count}\left\{Y_j, \left\|Y_j\right\|_1 \geq 25\% \times \left\|Y_j^{\max}\right\|_1, \forall j = 1, 2, \ldots, n\right\}, \tag{12}$$

$$f_{\text{ofn}} = \|Y\|_1 = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_{ij}, \tag{13}$$

$$\mathbf{Y}_j^{\max} = \mathbf{A}^{\max} \times \mathbf{B}_j. \tag{14}$$

To ensure that the number and frequency have the same measurement scale, the concept of "max–min" (Saheed et al. 2022; Ahlrichs and Rockstuhl 2022; Panda and Jana 2015) is adopted to normalize the data:

$$f'_{\text{mny}} = \frac{f_{\text{mny}} - \min\left(f_{\text{mny}}\right)}{\max\left(f_{\text{mny}}\right) - \min\left(f_{\text{mny}}\right)}, \tag{15}$$

$$f'_{\text{ofn}} = \frac{f_{\text{ofn}} - \min\left(f_{\text{ofn}}\right)}{\max\left(f_{\text{ofn}}\right) - \min\left(f_{\text{ofn}}\right)}, \tag{16}$$

where $f'_{\text{mny}}, f'_{\text{ofn}} \in [0, 1]$.

According to the VENN model, the assessment value is:

$$Z = \min\left\{f'_{\text{mny}}, f'_{\text{ofn}}\right\}. \tag{17}$$

2.3.5 Classification of pilot manual flight performance

The range of the model assessment values is (0, 1). If the assessment value is close to 1, the pilot showed good manual performance. In contrast, an assessment value close to zero indicates poor performance by the pilot. It is difficult to judge the flight performance of a pilot with an assessment score in the middle of this range. To assess the pilot performance more intuitively according to the assessment values, the assessment values are divided into four levels according to the expert evaluation values. However, the ranges of the evaluation values cannot simply be divided into four equal segments. Therefore, the clustering method is used to classify the assessment values, and the threshold ranges of different evaluation values are obtained.

The fuzzy C-means (FCM) clustering algorithm is an unsupervised learning algorithm that has been widely used in the field of data analysis (Migkos et al. 2022; Shi 2022). By optimizing the objective function, the membership degree of each sample point to all clustering centres can be obtained, which determines the category of sample points and achieves automatic classification of sample data.

First, the data set ($X = \{x_1, \dots x_g, \dots x_G\}$) is defined. x_g represents the gth pilot trainee's manual flight performance assessment sample, which has two-dimensional characteristics: the model assessment and expert assessment values. C denotes the number of categories, and $E_c (1 \leq c \leq C)$ is the clustering centre of the cth category. The FCM algorithm implementation process is described as follows:

1. Input the initial values of each parameter and randomly select a group of initial clustering centres;
2. Calculate the membership matrix ($\mathbf{H} = \left[h_{cg} \right]$);

$$h_{cg} = \frac{1}{\sum_{s=1}^{c} \left(\frac{\varepsilon_{cg}}{\varepsilon_{sg}} \right)^{\frac{2}{v-1}}}, \tag{18}$$

$$\sum_{c=1}^{C} h_{cg} = 1, \forall g = 1, 2, \dots, G, \tag{19}$$

where $h_{cg} (1 \leq c \leq C, 1 \leq g \leq G)$ represents the membership degree of the gth sample x_g belonging to the cth group and $\varepsilon_{cg} = \left\| E_c - x_g \right\|$ for the cth clustering centre and the Euclidean distance between the gth data points.
3. Update the cluster centre E_c:

$$E_c = \frac{\sum_{g=1}^{G} h_{cg}^v x_g}{\sum_{g=1}^{G} h_{cg}^v}, \tag{20}$$

where $v \in [1, +\infty)$ is an adjustable weighted index, usually $v = 2$.

4. Calculate the objective function (W). If it is below a certain threshold or if the change in the last objective function value is less than a certain threshold, the algorithm will stop and output both the clustering centre and membership matrix. Otherwise, return to Step (2).

$$W(H, E_1, \dots, E_C) = \sum_{c=1}^{C} W_C = \sum_{c=1}^{C} \sum_{g=1}^{G} h_{cg}^b \varepsilon_{cg}^2. \tag{21}$$

5. The cluster centre in the model assessment values was divided at equal distances to obtain the classification range.

3 Results

3.1 Flight phase division

According to the task requirements, the teardrop pattern procedure is divided into five flight phases: outbound level flight, outbound descent, base turn, inbound level flight, and inbound descent. The flight characteristics are shown in Table 3.

The descent rate range of each phase depends on the characteristics of the teardrop pattern procedure and the approach speed characteristics of a small trainer aircraft. The approach speed range for these mainstream aircraft (such as the C172 or DA40/42) is 60–100 knots. In theory, the descent rate during level flight should be 0 ft/min; however, in practice, there may be slight fluctuations due to wind influence. Therefore, the descent rate was set at 0 ± 50 ft/min during this phase. The primary objective of the outbound descent phase is to decrease altitude with a minimum descent rate of 400 ft/min. During the base turn section, the main task is to adjust the aircraft heading and intercept the heading path, resulting in a generally lower descent rate of less than 300 ft/min. During the inbound descent phase, the aircraft must steadily descend along a 3° glide path with an estimated descent rate of 300–400 ft/min based on the approach velocity. The partition results are shown in Fig. 3.

3.2 Pilot manual flight performance assessment

The outbound level flight phase in the teardrop pattern procedure was considered as an example, and Table 4 shows the assessment indexes corresponding to the OBs and their assessment criteria. The assessment criteria are based on statistical data, relevant regulations, and communication with flight experts.

Since weather conditions, altitude, and temperature can affect aircraft performance parameters during the approach stage, a flight expert's psychological quality and flight

Table 3 Teardrop pattern procedure approach flight phase and flight characteristics

Instrument approach section	Flight phase	Flight characteristic
Initial approach	Outbound level flight	① Maintain track ② Maintain speed ③ Maintain altitude
	Outbound descent	① Maintain track ② Maintain speed ③ Decrease altitude
	Base turn	① Decrease altitude ② Adjust flight path ③ Make turns according to the specified slope ④ Maintain speed
Intermediate approach	Inbound level flight	① Adjust the flight path to the designated flight path and correct the runway ② Adjust the configuration of the aircraft ③ Decrease speed to the target approach speed ④ Maintain altitude and intercept glide path
Final approach	Inbound descent	① Maintain track ② Maintain speed ③ Decrease altitude according to the standard descent rate
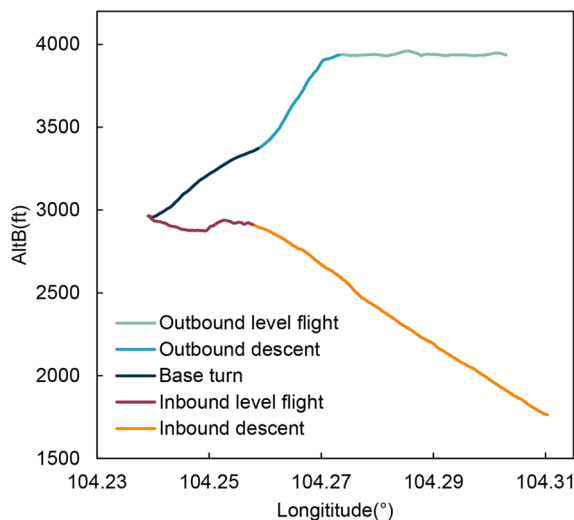
Fig. 3 Teardrop pattern procedure vertical flight path phase division

skills are obviously better than those of pilot trainees in the initial training stage. Therefore, we selected teardrop pattern procedure data from 20 pilot trainees rated as excellent by experts in the same batch of practice exams to fit the parameters and images of the target curve. Figure 4 shows the corresponding track, roll, and pitch curves.

According to the aforementioned Fréchet distance calculation process, Table 5 presents the quantitative values and corresponding scores for each aircraft attitude control observation item.

Trajectory deviation monitoring and correction and trajectory safety management assessment indexes are the tracked metrics. It is only necessary to determine whether there are data satisfying the observation conditions in the data set. To quantify the data, the time fast retrieval algorithm (Yin and Ye 2008; Luo et al. 2021; Pailot et al. 2020) is adopted to traverse the data set and obtain the observation item scores.

In the complete teardrop pattern procedure, 45 assessment indexes are related to the aforementioned three operational behaviours. Given the substantial amount of data involved, a pilot trainee's outbound level flight data during the teardrop pattern procedure are utilized as an illustrative example to demonstrate the assessment process.

1. Observation item score vector (A):

$$A = (3, 3, 3, 4, 4, 1, 1, 1, 1). \tag{22}$$

2. Correlation matrix (B):

$$B = [B_1, B_2, B_3] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}. \tag{23}$$

3. Assessment matrix (Y):

Table 4 Assessment indexes and assessment criteria in the outbound level flight phase

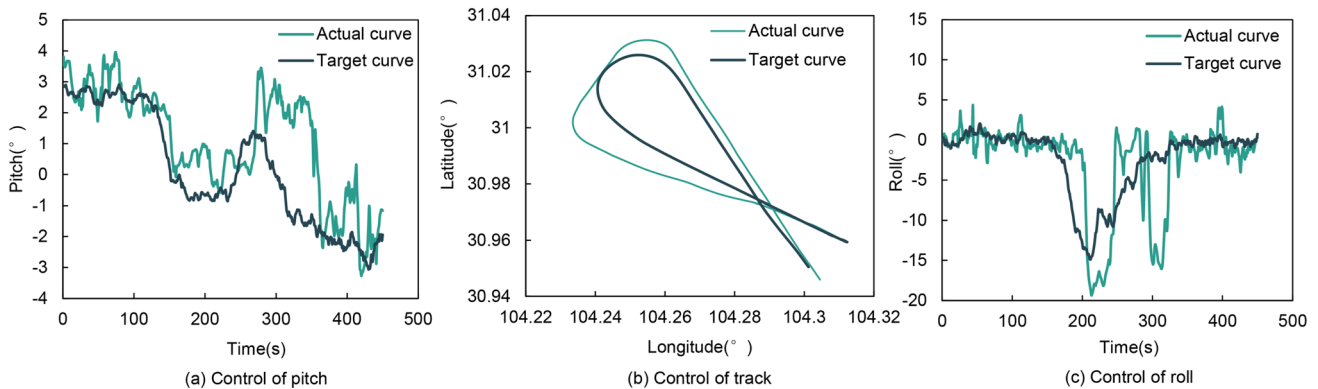| OB	Assessment index	Assessment criteria
Aircraft attitude control	(1) Control of pitch	4 score: $0 \leq$ Fréchet distance(pitch) ≤ 4.3 3 score: $4.3 <$ Fréchet distance(pitch) ≤ 8.7 2 score: $8.7 <$ Fréchet distance(pitch) ≤ 13.1 1 score: $13.1 <$ Fréchet distance(pitch)
	(2) Control of track	4 score: $0 \leq$ Fréchet distance(track) ≤ 17.2 3 score: $17.2 <$ Fréchet distance(track) ≤ 34.4 2 score: $34.4 <$ Fréchet distance(track) ≤ 51.7 1 score: $51.7 <$ Fréchet distance(track)
	(3) Control of roll	4 score: $0 \leq$ Fréchet distance (roll) ≤ 10 3 score: $10 <$ Fréchet distance (roll) ≤ 20 2 score: $20 <$ Fréchet distance (roll) ≤ 32 1 score: $32 <$ Fréchet distance (roll)
Trajectory deviation monitoring and correction	(4) Timeliness of trajectory deviation correction	4 score: trajectory deviation $\leq 10°$ 3 score: trajectory deviation $> 10°$, correct within 3 s 2 score: trajectory deviation $> 10°$, correct within 5 s 1 score: Trajectory deviation $> 10°$, correct after 5 s
	(5) Timeliness of altitude deviation correction	4 score: Altitude deviation ≤ 100 ft 3 score: Altitude deviation > 100 ft, correct within 3 s 2 score: altitude deviation > 100 ft, correct within 5 s 1 score: altitude deviation > 100 ft, correct after 5 s
Trajectory safety management	(6) Pitch attitude change rate within safe range	0 score: pitch attitude change rate $> 18°$/s 1 score: pitch attitude change rate $\leq 18°$/s
	(7) Roll angle within safe range	0 score: roll angle $> 55°$ 1 score: roll angle $\leq 55°$
	(8) Descent rate within safe range	0 score: descent rate > 1850 ft/min 1 score: descent rate ≤ 1850 ft/min
	(9) Indicated airspeed within safe range	0 score: indicated airspeed > 129 kn 1 score: indicated airspeed ≤ 129 kn

Fig. 4 Actual curve and target curve

Table 5 Assessment indexes and assessment criteria for Aircraft attitude control in the outbound level flight phase

	Control of pitch	Control of track	Control of roll
Fréchet distance	5.6	11.5	20.7
Score	3	3	3

$$\mathbf{Y} = [Y_1, Y_2, Y_3] = \begin{bmatrix} 3 & 0 & 0 \\ 3 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}.$$

4. "HOW MANY" (f_{mny}) and "HOW OFTEN" (f_{ofn}):

$$f_{mny} = \text{count}\left\{ Y_j, \left\| Y_j \right\|_1 \geq 25\% \times \left\| Y_j^{max} \right\|_1, \forall j = 1, 2, 3 \right\} = 3, \quad (25)$$

$$f_{ofn} = \| Y \|_1 = \sum_{i=1}^{9} \sum_{j=1}^{3} a_i b_{ij} = 21. \quad (26)$$

After the normalization treatment, $f'_{mny} = 1, f'_{ofn} = 0.8$

5. Assessment of outbound level flight:

$$Z = \min\left\{ f'_{mny}, f'_{ofn} \right\} = 0.8. \quad (27)$$

Fig. 5 Model and expert evaluations

The aforementioned steps are repeated for all assessment indexes in the teardrop pattern procedure to derive the final assessment value.

3.3 Verification of consistency

To validate the efficacy of the model, a consistency test was conducted on 30 pilot practice test data points, and both model evaluation results and expert evaluation results for the same student were obtained, as depicted in Fig. 5. It is evident that there exists a similar trend between the model evaluation results and expert evaluation results. Notably, it should be emphasized that while the expert evaluations are based on a 4-point scale, the model evaluations range from 0 to 1. Due to the influence of nonobjective factors on subjective expert evaluation, a certain degree of deviation between the expert evaluation and model results is acceptable.

To objectively measure the correlation between the expert and model evaluations, we introduced the Spearman correlation coefficient (Rosa et al. 2022; Rymarz et al. 2022; Liu et al. 2018). This nonparametric index can be used to measure correlations between different variables, regardless of their distribution pattern or sample size, as long as they appear in pairs. The larger the correlation coefficient is, the stronger the relationship. After calculation, the correlation coefficient between the model evaluation and the expert evaluation is 0.947 with a significance level of 0.01, indicating that the proposed evaluation model has achieved a significant and positive effect.

3.4 Assessment level classification threshold

Based on the assessment method developed in this study, the teardrop pattern procedure flight training quality of 30 pilots was evaluated, which corresponded with the expert evaluation value one by one, and the results are shown in Table 6. Under the expert evaluation system, a score of 4 is "excellent," a score of 3 is "good," a score of 2 is "medium," and a score of 1 is "unqualified".

Table 6 30 pilot trainees' model and expert evaluations

Number	1	2	3	4	5	6	7	8	9	10
Model evaluation	0.83	0.76	0.81	0.77	0.74	0.73	0.94	0.88	0.65	0.69
Expert evaluation	4	3	4	3	3	3	4	4	2	2
Number	11	12	13	14	15	16	17	18	19	20
Model evaluation	0.70	0.61	0.51	0.55	0.62	0.58	0.76	0.77	0.71	0.80
Expert evaluation	3	2	1	1	2	2	3	3	2	4
Number	21	22	23	24	25	26	27	28	29	30
Model evaluation	0.78	0.86	0.54	0.78	0.73	0.72	0.64	0.59	0.58	0.83
Expert evaluation	3	4	1	3	3	3	2	2	2	4

The fuzzy C-means clustering method was employed to determine the thresholds for each grade. Specifically, the range for "excellent" was (0.81, 1], that of "good" was (0.69, 0.81], the range for "medium" was (0.58, 0.69], and the range for "unqualified" was [0, 0.58].

4 Discussions

The present study proposes an evaluation method for assessing the manual flight performance of pilots during instrument flight training, focusing on competency. To validate the performance of the proposed method, the flight parameter data from a teardrop pattern procedure were utilized. Since the main purpose was to establish and analyse the proposed evaluation method, fewer participants were selected to reduce the amount of analysis. However, more experimental data lead to more reliable results; thus, the study had several limitations, and these limitations provide directions for future research.

First, although multiple flight parameters were used to evaluate the pilot's manual flight performance in this study, the interdependencies among the different flying parameters were not considered. Using additional models to evaluate the deviations between the target and actual curves for different parameters may be helpful in evaluating the aircraft attitude control performance.

Second, in this study, the teardrop pattern procedure in instrument flight training was used as the evaluation scenario, and the designed OB evaluation index is only applicable to this scenario. The competency performance of the pilot will vary according to the level of experience and expertise. When the method is extended to different evaluation scenarios, the actual situation must be considered, and the corresponding OB evaluation index must be constructed.

Third, in this study, only skill-related OBs based on flight parameters were selected to evaluate the pilots' manual flight performance. According to the CBTA concept, pilots are expected to demonstrate observable behaviours and meet performance standards in different situations based on knowledge, skills and attitudes (ICAO 2020). Therefore, in future studies, knowledge and attitude should be integrated to evaluate the manual flight performance on the basis of the evaluation method proposed in this study.

5 Conclusions

In this study, a novel method for evaluating the manual flight performance of pilots based on competency was developed, and the following findings were obtained.

1. To verify the validity of the evaluation method proposed in this study, 30 pilot trainees were selected for evaluation. The correlation coefficient between the evaluation results of flight experts and the evaluation results obtained by the method proposed in this study was 0.947, which shows the high consistency between the results. The results show that the method can be used to objectively evaluate pilots' manual flight performance based on flight data.

2. The FCM clustering method adopted in this study effectively reduces the subjectivity of the manual classification threshold, improves the evaluation results, and can be applied to quickly determine the pilot's manual flight performance according to the obtained evaluation value.

3. In future work, the proposed competency-based manual flight performance evaluation method will be applied to other evaluation scenarios as a reference for the evaluation schemes used in flight training institutions and by airlines.

4. It is worth noting that the evaluation indices selected in the study are mainly based on the teaching experience of flight experts. Therefore, in future studies, the correlations among different flight parameters should be analyzed to determine more objective evaluation indices.

Declarations

References

Agarwal PK, Avraham RB, Kaplan H, Sharir M (2014) Computing the discrete Fréchet distance in subquadratic time. SIAM J Comput 43(2):429–449

Ahlrichs J, Rockstuhl S (2022) Estimating fair rent increases after building retrofits: a max-min fairness approach. Energy Policy. https://doi.org/10.1016/j.enpol.2022.112923

Alidrisi HM, Mohamed S (2022) Developing a personal leadership competency model for safety managers: a systems thinking approach. Int J Environ Res Public Health 19(4):2197–2197

Chan TM, Rahmati Z (2018) An improved approximation algorithm for the discrete Fréchet distance. Inf Process Lett 138:72–74

Chen R, Jin J (2021) Detection of influential features of hard landing based on QAR data. In: 2021 IEEE 3rd international conference on civil aviation safety and information technology. https://ieeexplore.ieee.org/document/9633658

Chen CC, Wei CC, Chen SH, Sun LM, Lin HH (2022) AI predicted competency model to maximize job performance. Cybernet Syst 53(3):298–317

EASA (2022) Annual safety review 2022. https://www.easa.europa.eu/en/document-library/general-publications/annual-safety-review-2022. Accessed 12 May 2023

Frédéric V (2021) Heuristic-based method for conflict discovery of shared control between humans and autonomous systems—a driving automation case study. Robot Auton Syst. https://doi.org/10.1016/j.robot.2021.103867

Guo N, Ma M, Xiong W, Chen L, Jing N (2017) An efficient query algorithm for trajectory similarity based on Fréchet distance threshold. ISPRS Int J Geo Inf. https://doi.org/10.3390/ijgi6110326

Hashemian M, ZareFarashbandi F, Yamani N, Rahimi A, Adibi P (2021) A core competency model for clinical informationists. J Med Lib Assoc JMLA 109(1):33–43

IATA (2021) Competency assessment and evaluation for pilots, instructors and evaluators. https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/competency-assessment-and-evaluation-for-pilots-instructors-and-evaluators-gm.pdf. Accessed 12 May 2023

ICAO (2013) Manual of evidence-based training (Doc 9995). https://store.icao.int/en/manual-of-evidence-based-training-doc-9995. Accessed 12 May 2023

ICAO (2020) Procedures for air navigation services-training (Doc 9868). https://store.icao.int/en/procedures-for-air-navigation-services-training-doc-9868. Accessed 12 May 2023

Jeong GM, Lee Y, Lee CH, Kim MK, Lee JW (2018) Assessment of pilot training effectiveness of VR HMD based flight training device. J Korean Soc Aviat Aeronaut 26(4):129–141

Jirgl M, Boril J, Jalovecky R (2020) Statistical evaluation of Pilot's behavior models parameters connected to military flight training. Energies 13(17):4452–4452

Liu Q, Li C, Wanga V, Shepherd BE (2018) Covariate-adjusted Spearman's rank correlation with probability-scale residuals. Biometrics 74(2):595–605

Liu H, Wang H, Meng G, Hao Wu, Zhou M (2021) Flight training evaluation based on dynamic Bayesian network and fuzzy gray theory. Acta Aeronaut Astronaut Sin 42(8):250–261

Luo W, Li Y, Yao F, Wang S, Li Z, Zhan P, Li X (2021) Multi-resolution representation for streaming time series retrieval. Int J Pattern Recognit Artif Intell. https://doi.org/10.1142/S0218001421500191

McClelland DC (1974) "Testing for competence rather than for 'intelligence'": reply. Am Psychol 29(1):59–59

Migkos SP, Sakas DP, Giannakopoulos NT, Georgios K, Anastasia M (2022) Analyzing Greece 2010 memorandum's impact on macroeconomic and financial figures through FCM. Economies 10(8):178–178

Oh CG (2020) Pros and cons of a VR-based flight training simulator; empirical evaluations by student and instructor pilots. Proc Hum Fact Ergon Soc Annu Meet 64(1):193–197

Pailot-Bonnétat S, Harris AJ, Calvari S, De Michele M, Gurioli L (2020) Plume height time-series retrieval using shadow in single spatial resolution satellite images. Remote Sens 12(23):3951. https://doi.org/10.3390/rs12233951

Panda SK, Jana PK (2015) Efficient task scheduling algorithms for heterogeneous multi-cloud environment. J Supercomput. https://doi.org/10.1016/j.enpol.2022.112923

Rosa JC, Aleman JO, Mohabir J, Liang Y, Breslow JL, Holt PR (2022) The application of Spearman partial correlation for screening predictors of weight loss in a multiomics dataset. OMICS J Integr Biol 26(12):660–670

Rymarz J, Niewczas A, Hołyszko P, Dębicka E (2022) Application of Spearman's method for the analysis of the causes of long-term post-failure downtime of city buses. Appl Sci 12(6):2921–2921

Saheed YK, Abiodun AI, Misra S, Holone MK, Colomo-Palacios R (2022) A machine learning-based intrusion detection for detecting internet of things network attacks. Alex Eng J 42(12):429–449

Shi Y (2022) Application of FCM clustering algorithm in digital library management system. Electronics 11(23):3916–3916. https://doi.org/10.3390/electronics11233916

Song Y, Chun D, Xiong P, Wang X (2022) Construction of talent competency model for senior care professionals in intelligent institutions. Healthcare 10(5):914–914

Vaughn S, Rye J, Allen A, Bok A, Mauk K, Park L, Pierce L, Holler WW (2022) Updated competency model for professional rehabilitation nursing: practice applications. Rehabil Nurs J 47(1):12–23

Wang B, Jun Du, Ding C, Zhentao Wu, Zhang S (2019) Flight quality evaluation of aircraft take-off stage based on AHP-TOPSIS method. Flight Dyn 37(1):80–84

Wu Q, Wang P, Liu E et al (2018b) Design and implementation of learning management platform for aviation flight training based on SCORM/AICC standard—a case study of k airline company flight training learning platform. Adv Sci Lett 24(7):5194–5198

Yao Y, Xu K (2017) Quality assessment of flight training based on BP neural network. Acta Aeronaut Astronaut Sin 38(S1):24–32

Yazgan E (2018) Development taxonomy of human risk factors for corporate sustainability in aviation sector. Aircr Eng Aerosp Technol 90(6):1012–1022

Yin C, Ye Sh (2008) CBR algorithm supporting time series data. Appl Res Comput 2:395–397+400

Zhang L, Zhang W, Zhu Z, Ding L (2015) Probing into the evaluation of flight training in special environment based on EMGs. Procedia Manuf 3(C):4493–4500

Zhang Z, Ning M, Wu Z, Liu C, Li J (2022) Construction of a competition practice system for robotics engineering based on a competency model. Comput Appl Eng Educ 30(3):876–891

Zhu Z, Qian Y, Wang J, Yi X (2021) Evaluation of flight training quality based on time series similarity. Flight Dyn 39(6):75–81