AVIATION
ISSN: 1648-7788 / eISSN: 1822-4180
2022 Volume 26 Issue 4: 217–223
https://doi.org/10.3846/aviation.2022.18037

MODELLING A QUANTITATIVE ASSESSMENT METHOD OF PILOTS' PERFORMANCE IN EVIDENCE-BASED TRAINING

Zarif ZABIROV *, Anvar ZABIROV , Vladimir SHESTAKOV

Institute of Aeronautics, Faculty of Mechanical Engineering, Transport and Aeronautics, Riga Technical University, Riga, Latvia

Received 19 January 2022; accepted 20 April 2022

Abstract. With the improvement in reliability of modern aircraft systems, an increase in their complexity led to the fact that failures became non-standard and problematic to predict. As a result, IATA guided the development of Evidence-Based Training (EBT), which is an innovative approach to the training and assessment of pilots' competencies based on the evidence obtained from their performance. Yet, operators planning to use this framework should develop line-oriented scenarios based on EBT methodology and a grading system for a clear assessment. In this paper, a method for modelling a quantitative assessment and grading of pilots' performance is proposed. The method may be used as part of an EBT assessment programme due to its adaptability to different grading strategies according to quantifiable error criteria, the number of occurrences during the evaluation, and prior data on the distribution of pilots by their competency levels.

Keywords: evidence-based training, pilot assessment, core competencies, grading system, error criteria, flight occurrences.

Introduction

As more sophisticated technologies emerge in aviation and the complexity of aviation systems continue to increase, it is essential for aviation personnel to possess the required competencies, and for the training methods to be revised accordingly to cope with these challenges (European Aviation Safety Agency [EASA], 2016a, 2016b, 2019).

Evidence-based training is a competency-based training programme that was designed to improve the quality of pilot training and hence increase operational safety (International Air Transport Association [IATA], 2013). It is based on the principle that competencies are transferable, meaning that mastering the competencies in a scenario or the management of one malfunction would produce the same proficiency with another malfunction or scenario of the same group, therefore addressing unpredictable situations which can occur in flight (International Civil Aviation Organization [ICAO], 2013).

Pilots taking an EBT module go through three phases: the evaluation phase, the manoeuvres training phase, and the scenario-based training phase. The evaluation phase may include a single line-oriented scenario related to the air operator's needs, based on the methodology (IATA, 2013) involving a number of occurrences for the instructor or examiner to be able to rate the required core competencies (ICAO, 2013).

The instructor evaluates pilots using the respective behavioural indicators for each of the core competencies by observing the pilots' actions, interaction with other crew members, and the instrument readings of the flight simulation training device. In case of uncertainty, the instructor can also use a video recording of the session for its benefits (Flight Safety Foundation [FAF], 1997). At the end of the evaluation phase, the instructor conducts a debriefing with the pilots and grades their competencies. The more accurate the evaluation, the more efficient the training becomes, as the subsequent phases depend on the results of the evaluation phase.

The instructor assesses the performance of the flight crew based on the behavioural indicators and the compliance of the controlled parameters (operational limitations, stabilised approach criteria, etc.). Then the instructor decides on the grade for each of the core competencies of the flight crew members; in other words, the instructor categorises the pilots by their competency levels. Furthermore, it is known that the conduct of any categorisation task with sampling, confined by the scope of the assessment, is accompanied by type I and II errors, as in statistical

*Corresponding author. E-mail: *zarif.zabirov@outlook.com*

hypothesis testing. As a result, the instructor may overrate or underrate the performance of flight crew members.

The method for modelling pilots' performance assessment presented in this paper is based on the approach that was adopted in studies for modelling the decision-making of a pilot in flight and aeronautical knowledge tests by Karaush (1991) and the decision-making of an ATC controller by Hazzauri and Shestakov (2000). It is necessary to mention that a significant number of studies was conducted on the subject of pilot performance assessment and evaluation criteria; one of the early technical reports of such done by Mixon and Moroney (1981). However, the singularity of the proposed method is in its application within the EBT framework. Modelling the quantitative assessment is conducted through calculations of confusion matrices of pilots' performance levels, selection of decision boundaries justified by error criteria and grading of the core competencies.

1. Implementation of EBT and core competencies

There are two options for airlines considering implementation of EBT principles in their training: the baseline EBT and an enhanced EBT. The baseline EBT programme does not require a detailed programme design by the operator or the aviation training organization (ATO), as opposed to an enhanced EBT programme, which makes use of extensive data collection and analysis that acts as the source for adjustments to the training programme, improving its efficiency and operational safety. In addition to the operator's internal data, the enhanced EBT programme may include data from aircraft and equipment manufacturers (OEM), other aircraft operators, and training organizations. A short summary of the two EBT options is presented in Figure 1.

However, prior to any implementation of EBT, the minimum requirements considered necessary are the following:

- development of a set of competencies, and of an assessment and grading system;
- training of instructors in accordance with principles of the EBT programme;
- availability of information to pilots regarding principles of the EBT programme;
- available method for measuring training system performance (ICAO, 2013).

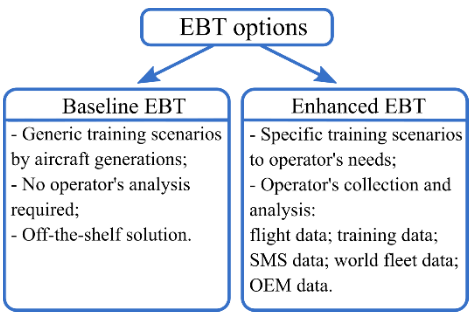


Figure 1. Differences in implementing baseline and enhanced EBT

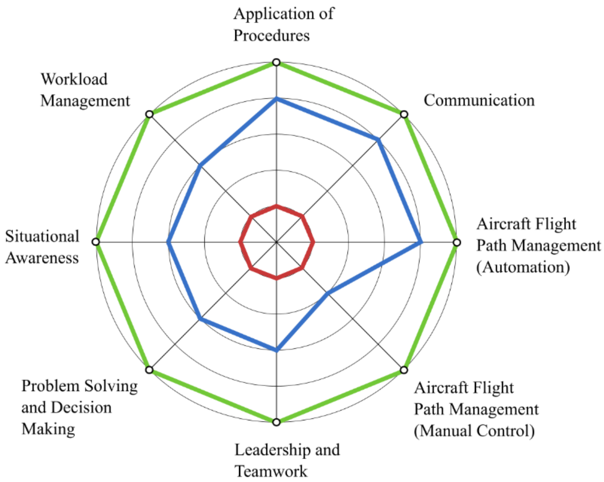


Figure 2. Graphic representation of grades by core competencies

The need for an accurate assessment of the core competencies of flight crew members is directly related to the effectiveness of subsequent EBT phases after the evaluation phase, which are based on the assessment of strengths and weaknesses, and are customised specifically to focus on shortcomings. The root cause of any deficiency should be identified rather than the symptoms.

According to an analysis (IATA, 2013), a five-point grading system, where 1 is unsatisfactory performance, is more suitable for the purposes of the evaluation framework. Each of the core competencies is graded during the evaluation phase and at the end of every training session.

The core competencies shown in Figure 2 are considered the foundation for any competent pilot. These eight core competencies may be amended by other competencies, as industry practice considers additional behavioural indicators that form an extra competence as "knowledge", and therefore they may be included (EASA, 2015). In Figure 2, the green regular octagon represents the maximum performance across all of the competencies. The irregular polygon represents a hypothetical pilot with various grades among the competencies, having a pronounced degradation of manual flight skills, which is one of the negative trends in modern civil aviation and which shall be addressed by operators (EASA, 2013; Federal Aviation Administration [FAA], 2013).

2. Quantitative assessment method of flight crew performance

Flight instructors evaluate pilots with a variable degree of accuracy, since it is not possible to observe flight crew in all possible emergency situations, in-flight occurrences, threats, and errors. Due to the multitude of these events, for explanatory purposes, these events are referred to further as occurrences. As a result, the assessment is carried out based on a sample of occurrences M from the universal set of occurrences N.

During a session, the flight crew encounters some occurrences ($M \ll N$) that serve as a foundation for the assessment. Ideally, it would be the most precise to consider the generation of the aircraft and assess how the flight crew handles all occurrences specific to this factor. Depending on the number of successfully tackled occurrences n, the final grade would be allocated to one of the intervals shown in Table 1.

Table 1. Intervals of performance

Interval	Performance
$n \in [N_5, N]$	exemplary performance
$n \in [N_4, N_5)$	effective
$n \in [N_3, N_4)$	adequate
$n \in [N_2, N_3)$	minimum
$n \in [0, N_1)$	unsatisfactory

Under these conditions, the assessment would be full-scale; otherwise type I and type II errors, where the grade is overrated or underrated, are inevitable. Hereafter, when it is stated that a pilot is able to counteract or handle an occurrence, this implies that the occurrence is addressed using the applicable indicators, also meaning that appropriate corrective crew actions are taken to ensure continued safe flight and landing (FAA, 1988).

Pilots in an airline can be expressed as elements x of the set X ($x \in X$). The set of pilots can be subdivided into subsets X_g, by their proficiency levels expressed through grades $g \in G = \{1 \ldots 5\}$ which may vary depending on the ability to effectively demonstrate a competency in n number of occurrences. In the case when an instructor rates a pilot as $x \in X_j$ but the pilot's true grade corresponds to $x \in X_i$, ($j \in i$), it indicates the i-th and j-th hypotheses are confused, and the probability of this event is denoted $P_{i,j}$. If $i > j$, then it indicates a type 2 error (the competency is underrated), and $i < j$ corresponds to a type 1 error (overrated); if $i = j$, the assigned grade corresponds to the true grade. Based on the conditions, a confusion matrix of probabilities $\mathbf{P}$ is obtained for a five-point grading system, where i is the true grade, and j is the assigned grade for a competency:

$$\mathbf{P} = \begin{bmatrix} P_{55} & P_{54} & P_{53} & P_{52} & P_{51} \\ P_{45} & P_{44} & P_{43} & P_{42} & P_{41} \\ P_{35} & P_{34} & P_{33} & P_{32} & P_{31} \\ P_{25} & P_{24} & P_{23} & P_{22} & P_{21} \\ P_{15} & P_{14} & P_{13} & P_{12} & P_{11} \end{bmatrix}. \quad (1)$$

The main diagonal corresponds to the probabilities when the assigned grade equals the true grade ($P_{55}, P_{44}, \ldots P_{11}$); below the main diagonal are the probabilities related to type I errors, and above to type II errors.

Certain error criteria are used to establish the minimum probability values of type I and II errors (Equation (2)) and (Equation (3)), and the maximum probability value of assigning true grades (Equation (4)).

$$P = \sum_{i=1}^{g} \sum_{j=1}^{g} P_{i,j}[i < j] \rightarrow \min ; \quad (2)$$

$$P = \sum_{i=1}^{g} \sum_{j=1}^{g} P_{i,j}[i > j] \rightarrow \min ; \quad (3)$$

$$P = \sum_{i=1}^{g} \sum_{j=1}^{g} P_{i,j}[i = j] \rightarrow \max . \quad (4)$$

3. Quantitative assessment modelling of pilots' performance

Having a number of occurrences M during the evaluation phase and pertaining to one of the core competencies, the objective is to effectively set decision boundaries to group the occurrences corresponding to the appropriate grades, in order to rate the pilot using the criteria (Equations (2–4)). A probability matrix is established, having g rows corresponding to the grade system and M columns, corresponding to the number of occurrences in the assessment. Each element $r_{g,m}$ of the matrix $\mathbf{R}$ (Equation (6)) indicates probabilities that a pilot undergoing evaluation, whose competence is of true grade g, will handle m occurrences ($0 \leq m \leq M$). Each element of this probability matrix is found using the formula:

$$r_{g,m} = \frac{C_{n_g}^{m} \cdot C_{N-n_g}^{M-m}}{C_N^M}, \quad (5)$$

where n_g – the number of occurrences the pilot is able to counteract (within one of intervals shown in Table 1); N – the total number of occurrences.

$$\mathbf{R} = \begin{bmatrix} r_{g,M} & r_{g,M-1} & \cdots & r_{g,0} \\ r_{g-1,M} & r_{g-1,M-1} & \cdots & r_{g-1,0} \\ r_{g-2,M} & r_{g-2,M-1} & \cdots & r_{g-2,0} \\ \vdots & \vdots & \vdots & \vdots \\ r_{1,M} & r_{1,M-1} & \cdots & r_{10} \end{bmatrix}. \quad (6)$$

The matrix $\mathbf{P}$ for a sample of occurrences M can be obtained from the matrix $\mathbf{R}$ (Equation (6)), however in order to create a confusion matrix for the five-point grading system, decision boundaries are established, and $r_{g,m}$ elements are summed accordingly. In addition, prior information about the distribution of pilots by their competency levels allows to adjust the confusion matrix by multiplying it with the diagonal matrix $\mathbf{Q}$ that has the main diagonal bearing the distribution data:

$$\mathbf{P}' = \mathbf{P} \cdot \mathbf{Q}. \quad (7)$$

An unfolded form of Equation (7) is the following:

$$\mathbf{P}' = \left[\begin{array}{c:c:c} r_{g,M}+r_{g,M-1} & r_{g,M-2}+r_{g,M-3}+r_{g,M-4} & \cdots \\ r_{g-1,M}+r_{g-1,M-1} & r_{g-1,M-2}+r_{g-1,M-3}+r_{g-1,M-4} & \cdots \\ r_{g-2,M}+r_{g-2,M-1} & r_{g-2,M-2}+r_{g-2,M-3}+r_{g-2,M-4} & \cdots \\ \vdots & \vdots & \vdots \\ r_{1,M}+r_{1,M-1} & r_{1,M-2}+r_{1,M-3}+r_{1,M-4} & \cdots \end{array}\right. \left.\begin{array}{c} r_{g,1}+r_{g,0} \\ r_{g-1,1}+r_{g-1,0} \\ r_{g-2,1}+r_{g-2,0} \\ \vdots \\ r_{11}+r_{10} \end{array}\right] \cdot \begin{bmatrix} q_g & 0 & 0 & .. & 0 \\ 0 & q_{g-1} & 0 & .. & 0 \\ 0 & 0 & q_{g-2} & .. & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & .. & q_1 \end{bmatrix}. \quad (8)$$

The dashed lines represent elements of $a_g, a_{g-1} \ldots a_1$, shown in Equation (8), that establish decision boundaries for the grades according to the performance levels. This division allows to build several probability matrices, depending on the position of the decision boundaries, and analyse them with respect to the error criteria.

The Equation (5) is based on the assumption that all occurrences during the assessment are homogeneous, however, this does not reflect actual flight operations. To address the issue, the occurrences are subdivided into two categories: average K and advanced N. In this partition, average occurrences are considered to correspond to the definition of minor failure conditions, while advanced occurrences are taken to correspond to the major and hazardous failure conditions (FAA, 1988). The difference in magnitude between these two categories is addressed using the weight coefficient α, and the weighted sum of advanced and average occurrences is expressed through ζ which corresponds to the true score:

$$\zeta = \alpha n + r, \quad (9)$$

where n – the number of advanced occurrences that the pilot is able to handle ($0 \le n \le N$); r – the number of average occurrences that the pilot is able to counteract ($0 \le r \le K$). Since the values N, K represent the number of all occurrences, the Equation (5) must be adapted to fit an actual assessment, which includes only a sampling of these occurrences; the achieved score η is expressed as follows:

$$\eta = \alpha m + l, \quad (10)$$

where m – the number of advanced occurrences that the pilot will solve ($0 \le m \le M$; $M \ll N$); l – the number of average occurrences that the pilot will solve as well ($0 \le l \le L; L \ll K$).

Accounting for the difference in difficulty between the occurrences, the probability that the true score corresponds to the obtained score is calculated using the following Equation:

$$P(\eta|\zeta) = \frac{C_n^m \cdot C_{N-n}^{M-m}}{C_N^M} \cdot \frac{C_{\zeta-\alpha n}^{\eta-\alpha m} \cdot C_{K-(\zeta-\alpha n)}^{L-(\eta-\alpha m)}}{C_K^L}. \quad (11)$$

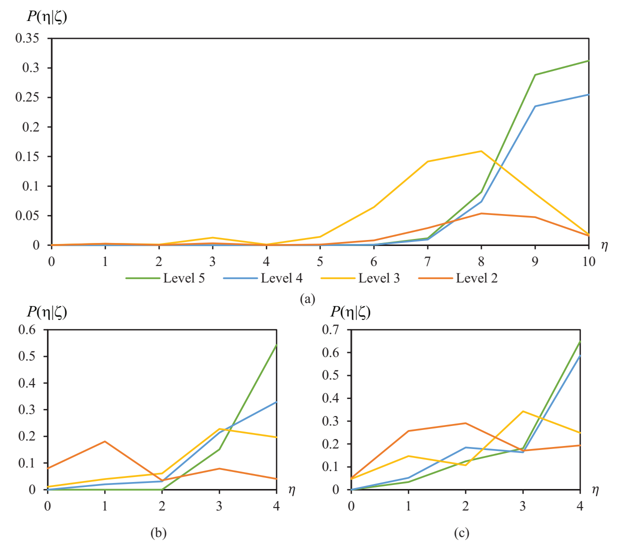
In order to obtain the probability function $P(\eta|\zeta)$ for four pilots of different competency levels, certain assumptions were made regarding their knowledge levels that are presented in Table 2. Various possibilities in which an evaluation may happen are shown in Figure 3.

Figure 3a has the highest number of occurrences, and the distinction in the probability of pilot scores, according to their level, is more pronounced, which decreases confusion in the evaluation process. In contrast, Figure 3b and 3c involve fewer occurrences during the evaluation. The former figure has a lower confusion between the level of pilots, however, it does not account for various difficulty levels of occurrences. Conversely, Figure 3c has the highest degree of confusion with a limited number of occurrences and difficulty levels, which reflects the conditions of a realistic simulator session. Therefore, the data used to obtain Figure 3c is analysed. The numbers of advanced and average occurrences $M = 1$, $L = 2$ and weight coefficient of $\alpha = 2$ are taken.

Table 2. Minimum numbers of occurrences that pilots can competently handle sorted by their level

Parameter	Pilot level			
	X_5	X_4	X_3	X_2
$n(N = 50)$	42	38	35	20
$r(K = 50)$	44	38	32	26
ζ	128	120	100	75

As can be seen from the graphs in Figure 3c, the probability of receiving a score corresponding to the true grade is the highest. However, there is a higher chance for a level-4 pilot to obtain 2 points than 3; the same tendency is observed for other pilots of different levels.



Note: a – for maximum score $\eta = 10(\alpha = 2; M = 2; L = 6)$; b – for maximum score $\eta = 4(\alpha = 1; M = 2; L = 2)$; c – for maximum score $\eta = 4(\alpha = 2; M = 1; L = 2)$.

Figure 3. The probability function $P(\eta|\zeta)$ for four pilots of different competency levels

Table 3. Required total score for grades and error criteria

Criteria	Criteria value:	Required total score for grades:			
		Grade 5	Grade 4	Grade 3	Grade 2
	q'	$q'_5 = 0.25$	$q'_4 = 0.25$	$q'_3 = 0.25$	$q'_2 = 0.25$
Max. $P_{ij}(i=j)$	0.307	4	3	2	1
Min. $P_{ij}(i<j)$	0.459	4	3	2	1
Min. $P_{ij}(i>j)$	0.279	4	3	2 or 1	0
Criteria	q''	$q''_5 = 0.15$	$q''_4 = 0.45$	$q''_3 = 0.35$	$q''_2 = 0.05$
Max. $P_{ij}(i=j)$	0.308	4	3 or 2	1	0
Min. $P_{ij}(i<j)$	0.488	4	3	2	1
Min. $P_{ij}(i>j)$	0.224	4	3	2	1

Furthermore, the error criteria are calculated with the obtained data using the probability function (Equation (11)), decision boundaries arrangement (Equation (8)), and are adjusted with prior data on the distribution of pilots by their competency levels. Specifically, q' corresponds to equal distribution among all proficiency levels, i.e. uncertainty, and q'' corresponds to exact figures shown in Table 3. For instance, the maximum probability of assigning the true grade (0.308) with available prior data on the distribution of pilots by their proficiency levels, a grade of 4 must be assigned for 3 or 2 points.

Although, the numerical data utilized, shown in Table 2, does not reflect real data; nonetheless, it was used to illustrate the principles of obtaining grading strategies by calculating error criteria and analysing the effect of variables such as the number of occurrences, their difficulty, and prior data on the distribution of pilots by their proficiency levels in an airline.

According to the chosen grading strategy based on the data from Table 3, the final grading of the core competencies is conducted, which is described in the following section.

4. Modelling grading process of core competencies

At the end of the evaluation phase, the instructor conducts a debriefing with the flight crew and grades their core competencies. To ensure the debriefing meets the required performance standards, the instructor's observations must be classified according to the competencies and respective behavioural indicators (IATA, 2013). In addition, the instructor is required to find the root cause of any pilots' deficiencies that happen during the assessment. This task is complicated by the fact that the final competency grades must be based on all occurrences that happened during the assessment.

In order to reduce the workload and improve the efficiency and accuracy of the grading process, it is proposed that the behavioural indicators of the core competencies must be assigned with weight coefficients for every occurrence that may be introduced during the evaluation phase. Therefore, a vector of weight coefficients $\boldsymbol{w}_{c_b}$ is assigned for each competence (c) in an occurrence (u), and the elements of which are denoted as $w_{c_{bu}}$, where c – the number of the competence; b – behavioural indicator; u – the number of the occurrence.

The fact that some of the behavioural indicators were demonstrated and their frequency, as observed by the instructor, is expressed through a vector of demonstrated behavioural indicators $\boldsymbol{d}_{c_b}$ for each competence, where elements $d_{c_{bu}}$ of the vector have values of 0 if an indicator is not demonstrated, 0.5 if it is observed intermittently, and 1.0 if demonstrated continually.

By the element-wise multiplication (Hadamard product) of $\boldsymbol{w}_{c_b}$ and $\boldsymbol{d}_{c_b}$, a weighted vector of demonstrated indicators $\boldsymbol{d'}_{c_b}$ in an occurrence (u) is obtained:

$$\boldsymbol{d}'_{c_b} = \boldsymbol{w}_{c_b} \circ \boldsymbol{d}_{c_b} = \begin{Bmatrix} w_{c_{1u}} \cdot d_{c_{1u}} \\ w_{c_{2u}} \cdot d_{c_{2u}} \\ \vdots \\ w_{c_{bu}} \cdot d_{c_{bu}} \end{Bmatrix}. \quad (12)$$

Using the sum of weighted vectors $\boldsymbol{d'}_{c_b}$ for all occurrences of the evaluation session, a column vector of the scores of behavioural indicators $\boldsymbol{s}_c$ is calculated as follows:

$$c = \begin{Bmatrix} d'_{c_{11}} + d'_{c_{12}} + \ldots + d'_{c_{1u}} \\ d'_{c_{21}} + d'_{c_{22}} + \ldots + d'_{c_{2u}} \\ \vdots \\ d'_{c_{b1}} + d'_{c_{b2}} + \ldots + d'_{c_{bu}} \end{Bmatrix}. \quad (13)$$

Once a vector of the scores $\boldsymbol{s}_c$ is obtained, a vector of grades $\boldsymbol{g}_c$ (14) for the relevant behavioural indicators must be derived using the criteria selected from Table 3.

$$\boldsymbol{g}_c = \begin{Bmatrix} s_{c_1} \to g_{c_1} \\ s_{c_2} \to g_{c_2} \\ \vdots \\ s_{c_b} \to g_{c_b} \end{Bmatrix}. \quad (14)$$

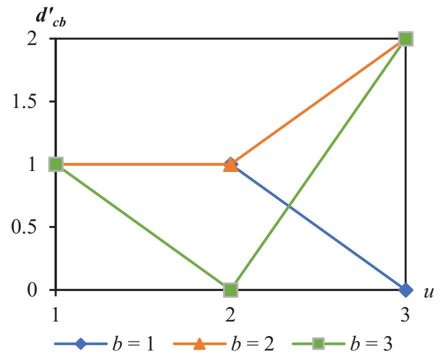


Figure 4. Graphical representation of core competence grading

Subsequently, the final grade G_c for a competency is the average of $\boldsymbol{g}_c$.

In order to demonstrate a calculation of G_c, a graph shown in Figure 4 was produced, using an imaginary competence with three behavioural indicators.

The highest scores for behavioural indicators ($b = 2$, $b = 3$) are assigned for the third occurrence ($u = 3$) due to the complexity of the occurrence which is expressed through weight coefficients ($\boldsymbol{w}_c$; α). Specific values obtained for Figure 4 are the following:

Table 4. Example of grading a competency throughout occurrences

Occurrences	$u = 1$		$u = 2$		$u = 3$ ($\alpha = 2$)		Total Score	Grade
Parameters	w_c	d_c	w_c	d_c	w_c	d_c	s_c	g_c
Indicator $b = 1$	1	1	2	0.5	2	0	2	3
Indicator $b = 2$	2	0.5	1	1	2	1	4	5
Indicator $b = 3$	1	1	1	0	2	1	3	4
							Final average grade G_c:	4

The use of weight coefficients $\boldsymbol{w}_{c_b}$ in the estimated calculation of G_c (Table 4) should ensure uniformity and standardization of the grading process, thus providing reliable reference information for the instructor conducting the evaluation. However, this should be used as a reference and a guide for grading, not as the grade itself, due to the individual qualities of pilots and the variety of possible deviations that may happen in a simulator session.

5. Further development of the method

Certain steps must be taken for the presented method to be fully prepared as an assessment and grading system for an enhanced EBT programme. Figure 5 presents a chart with steps in further development of the method.

It is imperative to develop a database and malfunction clustering system of occurrences, then classify them. In addition, a firm correlation between the occurrences and core competencies should be identified to adjust the

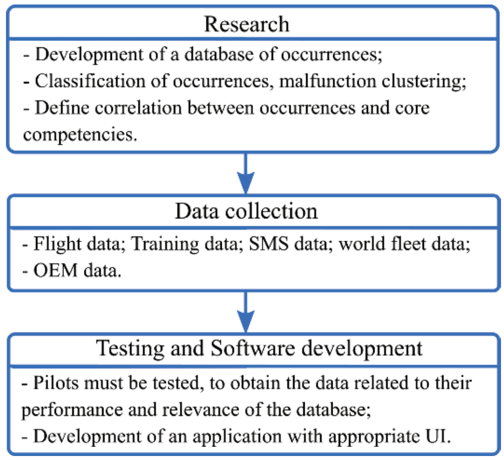


Figure 5. Further development chart

weight coefficients. Data collection and its use should introduce some corrections and improvements to the method. Finally, application of the method in the field will reveal the inaccuracies present and its subsequent refinement should allow for the development of software with the required features for instructors in EBT.

Conclusions

The utility of the proposed method is justified by the fact that the content and effectiveness of the subsequent training phases in Evidence-Based Training depend on the accuracy of the evaluation phase. This method employs a variation of statistical hypothesis testing to improve that accuracy. It is important to clarify that a multitude of studies was conducted on the subject of pilot performance assessment and evaluation criteria. Yet, the distinction of the proposed method is its application in the evaluation phase of the EBT framework. As is demonstrated by the results, the method has the following features:

1) adapted to various grading strategies using the quantifiable error criteria;
2) optimised to the number of occurrences during the evaluation phase;
3) adjusted to prior data on the distribution of pilots by their competency levels.

However, for an EBT programme to be developed using the proposed method, certain areas must be addressed:

a) development of an occurrence database categorised by difficulty in relevant flight operations including malfunction clustering;
b) training, validation and test sets of pilots of different competency levels must be tested to obtain the data related to their performance throughout the database with additional corrective inputs to the model and database of occurrences.

Disclosure statement

Authors declare that they do not have any competing financial, professional, or personal interests from other parties.

References

European Aviation Safety Agency. (2013). *EASA Safety Information Bulletin SIB 2013-05* (pp. 1–3). https://ad.easa.europa.eu/blob/SIB_201305_Manual_Flight_Training_and_Operations.pdf/SIB_2013-05_1

European Aviation Safety Agency. (2015). *ED Decision 2015-027-R. Implementation of evidence-based training (EBT) within the European regulatory framework* (pp. 1–11). https://www.easa.europa.eu/sites/default/files/dfu/ED Decision 2015-027-R - Explanatory Note.pdf

European Aviation Safety Agency. (2016a). *EASA Part-FCL*. https://www.easa.europa.eu/sites/default/files/dfu/Part-FCL.pdf

European Aviation Safety Agency. (2016b). *European Aviation Safety Agency Acceptable Means of Compliance (AMC) and Guidance Material (GM) to Annex III-Part-ORO* (Issue 2, p. 206). https://www.easa.europa.eu/sites/default/files/dfu/Consolidated unofficial AMC%26GM_Annex III Part-ORO.pdf

European Aviation Safety Agency. (2019). *European Plan for Aviation Safety (EPAS) 2020–2024*. https://www.easa.europa.eu/sites/default/files/dfu/EPAS_2020-2024.pdf

Federal Aviation Administration. (1988). *Advisory Circular 25.1309–1A System Design and Analysis*. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_25.1309-1A.pdf

Federal Aviation Administration. (2013). SAFO 13002 manual flight operations. In *Safety alert for operators*. http://www.ict-21.ch/com-ict/IMG/pdf/Binder3-FAA-SAFO.pdf

Flight Safety Foundation. (1997). *Facilitating LOS debriefings: A Training Manual* (Issue November). FSF.

Hazzauri, M., & Shestakov, V. (2000). Statistical modelling of operative decision making by the ATC controller in conditions of insufficient information. In *Nauchnyi vestnik MGTU GA* (Issue September, pp. 67–76). MGTUGA.

International Air Transport Association. (2013). *Evidence-based training implementation guide* (1st ed., Issue July). IATA. https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/ebt-implementation-guide.pdf

International Civil Aviation Organization. (2013). *Manual of Evidence-based Training DOC 9995* (1st ed.). ICAO.

Karaush, V. L. (1991). *Statisticheskoe modelirovanie processa prinyatia reshenia operatorom (pilotom) v usloviah ogranichennoi informacii* [Doctoral Dissertation, Saint Petersburg State University of Civil Aviation]. Saint Petersburg, Russia.

Mixon, T. R., & Moroney, W. F. (1981). *An annotated bibliography of objective pilot performance measures (Technical report: IH-330)*. Naval Training Equipment Center. Orlando, FL. https://apps.dtic.mil/sti/pdfs/ADA113170.pdf