Author(s): Tuhkala, Ari; Heilala, Ville; Lämsä, Joni; Helovuo, Arto; Tynkkynen, Ilkka; Lampi, Emilia; Sipiläinen, Katriina; Hämäläinen, Raija; Kärkkäinen, Tommi

Title: The interconnection between evaluated and self-assessed performance in full flight simulator training

Year: 2023

Version: Published version

Copyright: © The Author(s) 2023

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Tuhkala, A., Heilala, V., Lämsä, J., Helovuo, A., Tynkkynen, I., Lampi, E., Sipiläinen, K., Hämäläinen, R., & Kärkkäinen, T. (2023). The interconnection between evaluated and self-assessed performance in full flight simulator training. Vocations and Learning, Early online. https://doi.org/10.1007/s12186-023-09339-6

The interconnection between evaluated and self-assessed performance in full flight simulator training

Ari Tuhkala[1] · Ville Heilala[1] · Joni Lämsä[1] · Arto Helovuo[2] ·
Ilkka Tynkkynen[2] · Emilia Lampi[1] · Katriina Sipiläinen[1] ·
Raija Hämäläinen[1] · Tommi Kärkkäinen[3]

Abstract

This study explores potential disparities between flight instructor evaluations and pilot self-assessments in the context of full flight simulator training. Evaluated performance was based on the Competency-based Training and Assessment framework, a recent development of competency-based education within aviation. Self-assessed performance is derived from survey responses and debriefing interviews. The simulator session involves eight multi-crew pilot training graduates and eight experienced flight captains, encompassing two tasks featuring sudden technical malfucntions during flight. The flight instructor's evaluations reveal no significant differences in pilot performance. However, disparities become apparent when pilots engaged in reflecting their performance. Novice pilots, despite perceiving both tasks as easy, exhibited an overconfidence that led them to underestimate the inherent risks. Conversely, experienced pilots demonstrated greater caution towards the risks and engaged in discussing possible hazards. Furthermore, this study highlights the challenge of designing flight simulator training that incorporates surprise elements. Pilots tend to anticipate anomalies more readily in simulator training than during actual flights. Thus, this study underscores the importance of examining how pilots reflect on their performance, complementing the assessment of observable indicators and predefined competencies.

Introduction

Authentic virtual reality simulators provide a secure environment to practice responding to risky or impossible situations in real life (Chernikova et al., 2020). In aviation, there exist advanced technical training devices that can accurately mimic the operation of real aeroplanes. These devices, known as full flight simulators, combine aeroplane cockpits, computer software, displays, and other hardware to produce an authentic visual view from the aeroplane and force cueing that reacts to pilot's manoeuvres. This

---

Extended author information available on the last page of the article

Springer

enables pilots to immerse themselves in completing routine flight tasks and then face unexpected events that require them to diverge from their original manoeuvres and procedures (Salas et al., 1998). For instance, when a technical malfunction occurs at a critical stage of a flight, the pilot needs to evaluate threats caused by the malfunction and react accordingly (Casner et al., 2013). Thus, it is common to utilise full flight simulators for both pre-service and in-career training, although the high costs and expenses of these devices limit their availability (McLean et al., 2016).

Competency-based education has become the main educational approach in the training of professional pilots. The concept was introduced in the aviation industry in early 2000 as a result of the work made by an expert panel of the International Civil Aviation Organisation (Kearns et al., 2017). The panel proposed that pilots' profession should not be defined by the mere number of flight hours but by the competencies that are developed and deployed throughout the training phases. This creates a systematic approach to educating competent pilots that can operate an aeroplane safely and efficiently through continuous monitoring and evaluation of performance. The most recent development of competency-based education is the Competency-Based Training and Assessment (CBTA) framework. CBTA is composed of nine competencies, each containing several behavioural indicators that illustrate how it should manifest in action (EASA, 2020). The European Union Aviation Safety Agency has mandated that all European flight pilot training organisations incorporate CBTA into their training programs since 2022.

However, competency-based education has been criticised for its potential reductionism, if competencies and behavioural indicators are reduced to mere checklists, potentially limiting pilots' intuition and reflection (Franks et al., 2014; Hattingh et al., 2022; Hodge et al., 2020). Our study considers this issue and explores how the formal evaluation made by the flight instructor relates to the pilot's self-assessment. The formal evaluation is based on systematic grading with the CBTA framework, and the pilot's self-assessment is obtained from surveys and debriefing interviews. To implement this study, we organised a training session utilising the Airbus A320 Full Flight simulator. The session consisted of multiple flight tasks, with our focus being on two specific scenarios: one in which a malfunction occurs during standard take-off and another where landing takes place in a crosswind. The interest of this study is not merely comparing the performance between novice and experienced pilots, as in traditional expertise research. Instead, the interconnective aspect is that the novices had just finished their multi-crew pilot training, a program based on competency-based education and the use of full flight simulators, whereas experienced pilots have gone through more traditional training.

Hence, the aim of this study is to investigate: How does performance evaluated by the flight instructor using the CBTA framework and performance self-assessed by the pilot interconnect?

Background

Pilot training and professional expertise

Persons seeking to pursue a professional career as a pilot without prior aviation experience can apply for ab initio (i.e., from the beginning) training (Marques

et al., 2023). The applicants must have suitable physical characteristics and secondary education and take psychological tests that evaluate qualities such as their spatial perception and pressure resistance. Also, the applicants are interviewed by flight instructors and psychologists, and they undergo medical examinations and security checks. Ab initio training programs generally last for two years and include approximately 750 h of theory lessons and 200 flight training sessions. After the ab initio training, pilots can apply for professional airline pilot licenses, namely, Commercial Pilot License for smaller aircraft and Airline Transport Pilot License to become a captain of a high-capacity aircraft. The license requirements are defined by international standards and regulations by the International Civil Aviation Organisation and based on flying hours.

Ab initio training, as a form of professional education (Eraut, 2009), instils theoretical knowledge, including aerodynamics, principles of flight, air laws, human performance, and meteorology. Flying using only the instruments (i.e., without visibility outside the cockpit) and airplane control represent practical skills and techniques practiced in flight training sessions. Beyond this, necessary generic skills are, for example, communication and leadership, as successful plane operation involves collaboration between pilots, cabin crew, and air traffic control. Furthermore, pilots' general knowledge should cover topics such as the history of aviation and the aviation business.

After completing ab initio training and acquiring the necessary licenses, the pilot can apply for first officer positions. However, before becoming a first officer, they must take a first officer course and secure a rating for the intended plane type. During their career, pilots can complete additional training modules, such as gaining type ratings for other aircraft, taking commander training modules to become a flight captain, or pursuing a course to become a flight instructor. Furthermore, pilots must participate in recurrent training and pass regular tests to maintain their professional qualification to act as a pilot. In addition to the formal pilot training discussed above, much of learning occurs incidentally and informally during the career, for example, through imitating how more experienced pilots have handled different situations (Eraut, 2004).

After the ab initio training, the pilot begins the transition from a novice pilot into the realm of professional expert. In that regard, pilot's expertise is frequently discussed in training context within the concept of deliberate practice, which is defined as "activities that have been specially designed to improve the current level of performance" (Ericsson et al., 1993, p. 368). Within this view, it is not innate qualities, such as talent or gender, nor the mere amount of experience, that underpin better performance, but the amount of high-quality practice (Ericsson, 1998). A similar notion can be found in the works of Bereiter and Scardamalia (1993), where the mere repetition of routine tasks does not contribute to the development of expertise. Instead, professionals need to operate at the boundaries of their existing competency, engaging with novel challenges and problem-solving that push their capabilities to the limit. It is through such endeavors that professionals continually learn and refine their expertise.

Consequently, a long tradition of expertise research has examined the differences between expert and novice pilots' performance (e.g., Bellenkes et al., 1997; Durso

& Dattel, 2012; Jin et al., 2021; Yu et al., 2016; Wiggins & O'hare, 2003; Wiegmann et al., 2002): Expert pilots' attention is less constrained by resource limits than novices' attention because experts can automate domain-relevant tasks and develop efficient resource-management strategies, such as moving between tasks or prioritizing subtasks. Novices concentrate their visual search behaviour (scanning the horizon and the flight instruments) in a smaller area and visit flight instruments less frequently than the experts. Expert pilots can better perceive and interpret informational cues and structural variants in their dynamic environment, leading to better hazard detection and situational awareness. Experts are more aware of where potential hazardous situations may appear and, thus, better able to react to a situation without significant inference from other tasks. Expert pilots develop more efficient mental models that aid in their planning and decision-making. Furthermore, experts seem to have more comprehensive knowledge strategies that allow them greater flexibility in their decision-making and planning. These findings can be found from the efforts of defining and describing pilot's professional competency, as next discussed.

Evaluation and self-assessment in pilot's profession

The evaluation practices in pilot's profession derive from competency-based education. However, competency-based education involves various theoretical traditions, with quite contrasting views (Brockmann et al., 2008; Mulder et al., 2007). One demarcation is to consider whether competencies are understood as generic skills, developed creatively and tacitly in everyday interactions (*competence*), or as specific abilities to do something explicitly defined in advance and evaluated with precise criteria (*competency*) (Antera, 2021). In this sense, aviation industry has clearly adopted the latter, where an ability to participate in pilot's professional practice depends on specific core competencies (Kearns et al., 2017). These competencies include the specific information required to recall facts, identify concepts, apply rules, solve problems and think creatively (knowledge); the ability to perform certain actions or activities (skill); and an internal mental state or disposition that influences on personal choice towards an object, person or event (attitude) (IATA, 2023).

The CBTA is the most recently introduced evaluation framework that consists of nine core competencies (EASA, 2020). Evaluation is meant to be based on performance, compare pilots against pre-defined competencies instead of each other and recognise prior learning gained from earlier training or experience. The core competencies combine the psychomotor and cognitive technical skills to control the aircraft both manually and automatically, as well as the threat and error management skills for minimising risks. The non-technical skills are derived from crew resource management and non-technical skills frameworks, such as the European taxonomy of pilots' non-technical skills (Flin & Martin, 2001; Flin et al., 2003). The core competencies are assessed based on evidence from pilot's professional actions, referred to as behavioural indicators. For example, one behavioural indicator of communication is that the pilot ensures the recipient is ready and able to receive the information before starting the actual communication. As the evaluation is strongly based on the competency descriptions and

behavioural indicators, they play a major role in designing and implementing the actual pilot training (Table 1).

Competency-based education has had positive effects on pilot training, reshaping the emphasis from inputs (such as the number of flying hours) to outputs (what pilot does), acknowledging the diversity in individual learning paces and aptitudes among pilots by allowing training to continue until competency is achieved, irrespective of the accumulated flying hours, and facilitates the provision of training and evaluation tools that are highly pertinent to the demands of the professional pilots (Kearns et al., 2017). However, several scholars have highlighted inherent challenges in competency-based education. For instance, Franks el al. (2014) conducted an examination of competency-based training within Australian industry and advocated for the incorporation of problem-based learning that places greater emphasis on higher-order thinking skills within ab initio training programs. They also suggested the adoption of an assessment framework capable of discerning different levels of expertise. Hattingh et al. (2022) argue that flight instructors may not necessarily possess the capability to interpret competency descriptions accurately, potentially leading to difficulties in implementing training that is strictly aligned with competency requirements. Hodge et al. (2020) raise concerns regarding the assumptions underlying competency documents, particularly whether they can comprehensively capture all facets of a professional's practice and whether they would be consistently interpreted across diverse contexts.

While the CBTA framework does address some of the presented criticism, such as delineating varying levels of competency, it remains notably silent on a crucial factor of learning: the ability of professionals to engage in reflective practice (Hager, 2008; Bontemps-Hommen et al., 2020). In this context, there is emerging body of research considering reflection in pilot's profession (Mavin, 2016; Mavin & Roth, 2014; Mavin & Roth, 2015; Mavin et al., 2018). These inquiries extend beyond the question of how pilots can enhance their professional practice through reflection; it encompasses the notion that reflection itself constitutes a skill that necessitates cultivation (Mavin & Roth, 2014). Furthermore, the enhancement of a pilot's capacity to reflect on their professional practice retrospectively can serve as a pivotal catalyst for fostering greater reflexivity while actively engaged in their practice (Cattaneo & Motta, 2021). Consequently, our considerations regarding a pilot's self-assessment are not confined solely to numerical performance ratings in the form of self-evaluation. Rather, they extend to encompass how pilots engage in reflective discussions about their performance during debriefing sessions following simulator exercises. This multifaceted approach recognizes the nuanced interplay between self-assessment and the broader development of reflective practice within the field of aviation.

Methods

An ethical review was carried out before conducting this study by the Human Sciences Ethics Committee of [anonymised] to ensure that it follows the guidelines for responsible conduct of research. The review conducted an external assessment

Table 1 The competency-based training and assessment framework (EASA, 2020)

Competency	Acronym	Description
Application of knowledge	APK	Demonstrates knowledge and understanding of relevant information, operating instructions, aircraft systems and the operating environment
Application of procedures and compliance with regulations	PRO	Identifies and applies procedures in accordance with published operating instructions and applicable regulations, using the appropriate knowledge.
Communication	COM	Demonstrates effective oral, non-verbal and written communications, in normal and non-normal situations.
Flight path management: manual / automatic control	FPM	Controls the aircraft flight path through manual flight, including appropriate use of flight management system(s) and flight guidance systems.
Leadership and teamwork	LTW	Demonstrates effective leadership and team working.
Problem solving and decision making	PSD	Accurately identifies risks and resolves problems. Uses the appropriate decision-making processes.
Situation awareness and management of information	SAW	Perceives and comprehends all the relevant information available and anticipates what could happen that may affect the operation.
Workload management	WLM	Manages available resources efficiently to prioritize and perform tasks in a timely manner under all circumstances.

of the research plan, the data management plan, the data privacy notification and the research consent for participants.

Participants

All together, we recruited 16 pilots using the intra-corporate communication channels of the stakeholder flight company. All pilots were licensed to operate Airbus A320 as a fully qualified crew member. All pilots had the same nationality, and all were male. Eight of the recruited pilots had just finished their multi-crew pilot license (MPL) training and type-rating courses. When comparing to ab initio training, the MPL aims for qualified multi-crew flight deck instead of single-pilot operations, and the training program is strongly based on the CBTA framework and the use of full flight simulators (Wikander & Dahlström, 2016). Moreover, the training is performed in line-oriented flight training sessions, where pilots practice real-life threats and challenging conditions as part of the crew, instead of practicing in a single-pilot aircraft. In contrast, the other eight pilots were all working as captains and had at least five years of flight experience. To clarify, we will further on refer to the MPL graduates as novice pilots and to the flight captains as experienced pilots.

Simulator sessions

The simulator sessions were performed in the Airbus A320 Full Flight Simulators that is certified in the highest category (D) of regulated flight simulator systems (EASA, 2012). The simulator sessions for the experienced pilots took place in October 2020 and for the novice pilots in November 2020. All participants acted as pilot flying (main responsibility to control the aircraft) and had the same experienced co-pilot as a pilot monitoring. A qualified simulator instructor operated the simulator and acted as an air-traffic controller.

The simulator sessions lasted 47 min on average (SD = 9 min, range = 42 to 61 min) and followed the typical training structure of the training organization. The session begun with an orientation by performing a normal take-off and a short flight along the departure route. The actual training was designed to periodically increase workload and difficulty across the following tasks: normal take-off, approach and landing in light wind, take-off with a flight management system failure, approach and landing in strong crosswind and approach and landing in strong crosswind including an instrument failure. There was a 1–3-minute pause between the tasks, during which the simulator calculated the parameters and the pilots oriented for the next task. All approaches were flown manually based on instrument landing system and raw data, without flight guidance augmentation.

This study focuses on two specific tasks that involved technical malfunctions. In the first task, take-off with a flight management system failure, the environmental conditions were easy, but there was a computer navigation failure at a critical

point after take-off. While the plane was climbing to the assigned altitude and turning toward the assigned heading, there was a sudden loss of both the autopilot and flight director. The failure activated a system fault indicator on the display, which automatically disappeared 30 s after activation. During this half-a-minute, the pilot had to manually fly the plane to keep on the intended flight path while simultaneously assessing the situation and taking the proper actions. In a two-pilot crew, this required active communication about the nature of the problem, prioritisation of actions and coordination of task sharing.

In the second task, crosswind approach and landing with an instrument failure, the pilot needed to fly using a raw data instrument approach and land in a high-velocity crosswind. The task was challenging as the plane needed to be flown with proper wind correction, so it was necessary to point the nose of the plane into the wind to avoid it drifting from the intended path. The technical malfunction occurred when the plane had descended to 2500 feet on its approach path. The monitoring pilot's primary flight display failure made the artificial horizon to drift, and as the plane detected a discrepancy between the altitude indicator on the flying pilot's and that on the monitoring pilot's sides, a failure alert chimed, and an electronic checklist called for pilot action. In this case, the crew needed to compare primary and standby instruments to verify which source was giving a false indication and then change the data source of the faulty instrument to the operating one. The workload of the flying pilot was significantly increased because the normal instrument scanning pattern had to be altered until the abnormal situation was handled.

Data collection

First, the pilot was informed of the study and the rights concerning the experiment and equipped with a microphone. Before entering to the simulator, the pilot filled a pre-survey that included questions about expectations regarding the upcoming session, such as how well does the pilot expect to perform in the simulator session. After the pilot was seated, the simulator session started. The pilot's performance was recorded with one video camera focusing on the cockpit and two GoPro cameras focusing on the flight instruments. When the simulator session was over, the pilot filled out a post-survey that asked the pilot to rate each of the session tasks. The pilot entered a debriefing room, where the interview took place. The interview was recorded with a microphone and one GoPro camera. The interview questions were presented by one of the researchers and moderated by a flight instructor. The debriefing interview started with a general question about the first impressions and thoughts regarding the simulator session. Then, the tasks were discussed in detail, using interview prompts that encouraged the pilot to reflect on the tasks. The same questions were used for both tasks: (a) How familiar the malfunction was? (b) On what basis did you decide on your actions? (c) How do you think you performed in the situations? (d) What do you think was particularly important in the situation? (e) Was something particularly easy or difficult in the situations? (f) Would you do something differently in a similar situation?

Data analysis

For the evaluated performance, the flight instructor who was present in the debriefing, but not in the simulator session, analysed the video recordings. The evaluation followed the guidelines of the CBTA manual (EASA, 2020). The automated flight path management competence was removed because both tasks were flown manually. Additionally, only relevant behavioural indicators for the task were included in the evaluation (Table 2). The applicable indicators were evaluated on a scale of 1 to 5, so that the higher value means more effective and regular demonstration of behaviour with safer outcome (5=significantly enhances safety, 4=enhances safety, 3=safe operation, 2=did not result in an unsafe situation, 1=resulted in an unsafe situation).

Self-assessed performance consisted of survey answers and debriefing interviews. When regarding the survey answers, we applied four items that represented pilot's own view of task performance on a scale of 1 (bad) to 10 (excellent): the expected overall session performance (pre-survey), perceived performance in task 1 (post-survey), perceived performance in task 2 (post-survey) and how performance matches current level of competence (post-survey). We decided to report all answers, instead of mere descriptive statistics, as it was concise enough to be included in a single table.

For the debriefing interviews, we utilised qualitative content analysis following an inductive approach that is recommended when the researchers have insufficient a priori knowledge (Mayring, 2000, 2014). Hence, the interviews were analysed by

Table 2 The evaluated competences and number of applied behavioural indicators

Competency	Original indicators	Task	Applied indicators	Percentage of original indicators
Application of knowledge	7	T1	3	43
		T2	2	29
Application of procedures and compliance with regulations	7	T1	6	86
		T2	6	86
Communication	10	T1	7	70
		T2	7	70
Flight path management: manual	7	T1	6	86
		T2	5	71
Leadership and teamwork	11	T1	4	36
		T2	5	45
Problem solving and decision making	9	T1	5	56
		T2	6	67
Situation awareness and management of information	7	T1	3	43
		T2	4	57
Workload management	9	T1	7	78
		T2	7	78

the researchers who were not pilots, or pilot instructors, and thus not familiar with CBTA or other pilot training frameworks. As such, our goal was to find different ways in which the pilots reflected on the tasks, instead of mapping existing ones.

First, we transcribed the interviews and exported the files to Atlas TI version 22. The first author started the analysis by identifying relevant instances and then developing codes that could group the instances together. For example, an excerpt where a pilot explained the importance of staying calm was coded with 'T1-d act calm', in which the signs marked which task (e.g. T1) and question (e.g. question d) the code refers to. Two other authors analysed the data with the same coding scheme, but only for two experienced (E7, E8) and two novice (N2, N3) pilots. The purpose was to ensure that all relevant instances were identified by the first author and to propose alternative codes if necessary. Eventually, relevant instances were marked as quotations and assigned one or more codes, and each code was briefly described. The authors responsible for the analysis discussed the final coding scheme, seeking to interpret the meanings of and possible relationships between the codes. They combined similar codes into groups that represented the target of reflection. For example, the codes 'act calm', 'control the airplane', 'prioritise actions' and 'analyse before acting' were grouped together and given a description 'The most important action was to [code] in the situation'. Finally, selected quotations were translated into English for use in this paper to demonstrate the research findings.

Results

Evaluated performance

The competence grade is an average calculated from the relevant behavioural indicators. For example, pilot's application of knowledge in the first task is an average of the following three behavioural indicators: (a) identifies the source of the failure, knows how to recover systems and understands the effects on the continuation of the flight, (b) applies standard procedures for take-off sequence correctly, applies correct task sharing, call outs, and ECAM procedure in trouble shooting failure and (c) manages failure effectively by demonstrating knowledge (Table 3).

The average overall performance in the first task was 4.24 for experienced pilots and 4.33 for novice pilots. Despite the rather similar performance as a group, the performance of novice pilots was more stable. When using a cut-off value of four, the performance was lower for three of the experienced pilots (E3, E6 and E8) and for one novice pilot (N1) (Table 4).

The results for the second task were rather similar, although the difference between the two groups was slightly larger. The average overall performance was 4.27 for experienced pilots and 4.41 for novice pilots. The overall performance was below the cut-off value for two experienced and one novice pilot. It can be noticed that the overall performance was at least three (safe operation) for all pilots, but as a group, the novices performed a little bit better than the experienced pilots. Especially when there was one outstanding performance (E1) that raised the

Table 3 Evaluated performance in the task 1 (5 = significantly enhances safety, 4 = enhances safety, 3 = safe operation, 2 = did not result in an unsafe situation, 1 = resulted in an unsafe situation)

Experienced pilots	E1	E2	E3	E4	E5	E6	E7	E8	Average
Application of knowledge	5.00	4.33	3.33	4.00	5.00	3.67	3.33	5.00	4.21
Application of procedures and compliance with regulations	5.00	4.67	3.17	4.50	4.83	4.17	3.50	4.67	4.31
Communication	5.00	5.00	3.83	4.50	4.00	3.86	3.83	4.67	4.34
Flight path management – manual control	5.00	4.83	3.33	3.50	4.67	3.67	4.50	4.33	4.23
Leadership and teamwork	4.50	4.75	3.50	4.50	4.67	4.00	4.25	5.00	4.40
Problem-solving and decision-making	4.80	4.40	3.60	4.00	4.20	4.00	3.80	4.40	4.15
Situation awareness and management of information	5.00	4.67	3.33	3.67	3.67	4.00	3.67	4.00	4.00
Workload management	5.00	5.00	3.71	4.29	4.57	3.71	3.57	4.14	4.25
Overall competence	**4.91**	**4.71**	**3.48**	**4.12**	**4.45**	**3.88**	**3.81**	**4.53**	**4.24**
Novice pilots	N1	N2	N3	N4	N5	N6	N7	N8	Average
Application of knowledge	2.67	4.33	4.33	5.00	4.67	4.33	4.67	4.67	4.33
Application of procedures and compliance with regulations	3.00	4.33	4.33	4.67	4.83	4.83	4.67	4.67	4.42
Communication	3.50	5.00	5.00	5.00	4.50	4.33	4.86	4.86	4.63
Flight path management – manual control	3.67	4.67	3.00	4.50	4.50	5.00	4.33	4.33	4.25
Leadership and teamwork	3.67	4.67	4.67	4.67	4.33	4.33	4.67	4.67	4.46
Problem-solving and decision-making	2.67	4.33	4.33	5.00	4.20	4.50	4.40	4.40	4.23
Situation awareness and management of information	2.67	4.00	4.00	4.67	4.33	4.00	4.67	4.67	4.13
Workload management	3.14	4.00	4.00	4.67	4.67	4.67	4.67	4.67	4.31
Overall competence	**3.12**	**4.42**	**4.21**	**4.77**	**4.50**	**4.50**	**4.62**	**4.62**	**4.34**

Table 4 Evaluated performance in the task 2 (5 = significantly enhances safety, 4 = enhances safety, 3 = safe operation, 2 = did not result in an unsafe situation, 1 = resulted in an unsafe situation)

Experienced pilots	E1	E2	E3	E4	E5	E6	E7	E8	Average
Application of knowledge	5.00	4.00	3.50	4.50	4.00	5.00	4.00	5.00	4.38
Application of procedures and compliance with regulations	5.00	4.50	3.67	4.00	4.33	4.17	3.33	4.83	4.23
Communication	4.50	4.50	3.83	4.50	3.80	3.83	4.33	4.83	4.27
Flight path management – manual control	5.00	5.00	3.60	5.00	4.60	4.25	3.50	4.50	4.43
Leadership and teamwork	4.75	5.00	3.60	4.20	4.00	3.80	4.20	4.80	4.29
Problem-solving and decision-making	4.83	4.83	3.50	4.00	4.20	4.00	4.00	4.33	4.21
Situation awareness and management of information	5.00	5.00	3.75	4.00	4.00	3.60	3.75	4.00	4.14
Workload management	5.00	4.71	3.57	4.29	3.86	4.00	3.57	4.43	4.18
Overall competence	**4.89**	**4.69**	**3.63**	**4.31**	**4.10**	**4.08**	**3.84**	**4.59**	**4.27**
Novice pilots	**N1**	**N2**	**N3**	**N4**	**N5**	**N6**	**N7**	**N8**	**Average**
Application of knowledge	3.00	5.00	5.00	4.50	5.00	3.50	3.50	3.50	4.13
Application of procedures and compliance with regulations	4.00	4.83	4.83	4.83	4.67	4.33	4.33	4.33	4.52
Communication	3.80	5.00	5.00	5.00	4.50	4.33	4.67	4.67	4.62
Flight path management – manual control	4.40	4.60	4.80	4.75	4.75	5.00	4.75	4.75	4.73
Leadership and teamwork	3.50	4.50	4.50	5.00	4.25	4.25	5.00	5.00	4.50
Problem-solving and decision-making	3.33	4.17	4.17	4.80	4.40	3.83	4.17	4.17	4.13
Situation awareness and management of information	3.33	4.67	4.67	4.75	4.75	3.75	4.50	4.50	4.36
Workload management	3.67	4.40	4.40	5.00	4.57	3.83	4.14	4.14	4.27
Overall competence	**3.63**	**4.65**	**4.67**	**4.83**	**4.61**	**4.10**	**4.38**	**4.38**	**4.41**

average performance of the experienced pilots. In addition, there was no considerable differences between different competencies, as the average was over four for all competences.

Self-assessed performance

The pilots' answers to the selected survey items are presented in Table 5. Almost all pilots expected to perform well in the session, except N7, who anticipated a lower performance than other pilots (N5 did not answer the question). N3 reported a worse performance for the first task than the others, as he explained that he had missed the correct altitude. N7 reported that he had failed the second scenario after being too hasty in cancelling the approach. Otherwise, the pilots' self-assessed performance was good for both scenarios, varying between seven and ten. Based on the pilots' own assessment, their performance in the session corresponded to their current level of competence quite accurately, also varying between seven and ten. To conclude, the self-assessed performance was good for all pilots except N3 and N7.

Table 6 presents the codes that resulted from the qualitative content analysis of the debriefing interviews. The code description explains the context where the code manifested. For example, two experienced pilots and four novice pilots demonstrated capacity awareness in the task 1. To assist in interpreting the table, Fig. 1 depicts the target of reflection. For example, capacity awareness, composure, performance and ability to recall the situation codes emerged when the pilot was reflecting himself or his own actions in the simulator.

Capacity awareness demonstrates either their good or limited capacity to handle the occurred malfunction. Of the experienced pilots, one noted that he had not remembered to execute a certain action immediately, a second said that focusing on flying had taken up much of his time and a third brought up the difficulty of analysing the situation from many different perspectives:

> When you are in a hurry, and need to analyse the situation quickly, it is difficult to question your own arguments so that you do not trust only the ones that

Table 5 Survey answers (scale: 1 = bad, 10 = excellent)

Experienced pilots	E1	E2	E3	E4	E5	E6	E7	E8
Expected simulator session performance	10	8	9	8	8	8	8	10
Self-assessed performance in task 1	10	9	8	8	8	8	9	9
Self-assessed performance in task 2	10	8	9	9	8	8	9	9
Performance represents current competence	9	9	8	8	9	8	7	10
Novice pilots	**N1**	**N2**	**N3**	**N4**	**N5**	**N6**	**N7**	**N8**
Expected simulator session performance	7	8	8	8		7	5	8
Self-assessed performance in task 1	8	7	5	8	7	10	7	9
Self-assessed performance in task 2	8	8	7	8	8	10	4	8
Performance represents current competence	9	8	7	9	9	8	7	10

Table 6 Code descriptions and number of pilots who were assigned the code

Code description	Code	Task 1		Task 2	
		Experienced	Novices	Experienced	Novices
Pilot demonstrated [code] in the debriefing	Capacity awareness	2	4	1	5
	Composure	1	4	1	6
	Awareness of good/bad performance	7/0	6/1	5/0	6/1
	Ability to recall the situation	0	6	0	3
Tasks stimulated [code] to improve expertise	Development needs	6	7	1	5
	Risk awareness	4	0	7	2
Pilot referred to [code] when discussing expertise	Experience	5	1	3	3
	Improvisation	2	1	0	3
	Procedures	5	8	2	7
	Training	6	3	5	5
When discussing about the co-pilot, [code] was emphasised	Communication	2	4	6	3
	Responsibilities	2	2	3	5
	Support	1	7	0	4
It was most important to [code] in the situation	Act calm	2	2	1	2
	Analyse before acting	3	1	2	1
	Control the aeroplane	8	7	6	6
	Prioritise actions	6	2	4	3
The task was perceived as [code]	Easy/difficult	5/2	3/1	6/2	4/4
	Familiar/unfamiliar	8/0	8/0	4/4	2/6
	Immersive	3	5	3	5
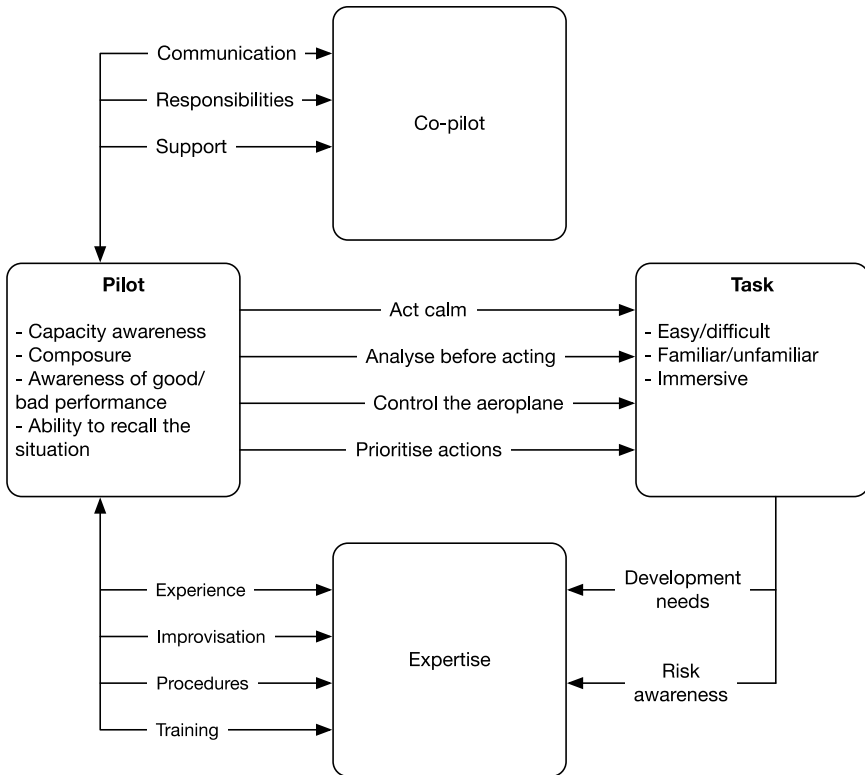
Fig. 1 The targets of reflection

support your own view, and so that you do not just continue, ignoring possibly conflicting indications.

The novice pilots brought up how they can forget some priorities when in hurry. As there is so much to remember even in normal situations (without malfunctions), any additional stimuli in a demanding task causes confusion. They feel compelled to hurry, even if there is no need, and that may add stress to the situation, already hampered by their limited self-esteem due to insufficient experience.

Composure implied that the pilots experienced the task as easy or unworrying. In the first task, the novice pilots remarked that "the situation was quickly under control" (N1), "the situation was stable all the time so there was no need for big changes" (N4), "there was nothing special in the situation" (N6) and "the malfunction was simple" (N8). In the second scenario, novice pilots were unworried by the situation because the failure occurred on the monitoring pilot's display, not their own: "I decided that everything would be alright; even if we could not handle the failure on the monitoring pilot's Primary Flight Display, they could use my display" (N1), and another noted that "the situation was very calm, and in that sense, easy for me, as the failure was only on captain's display, not in my own, so I had no other concerns than concentrating on flying" (N6). In contrast, there was

only one experienced pilot (E1) who described handling the situation with composure. However, his discussion style was humorous and his comments regarding the situation were at times sarcastic. For example, in describing the second task, he remarked, "I trusted myself like I used to trust my neighbours, and knew that I could handle the plane, so I dared to launch some malfunction troubleshooting during the approach."

Awareness of good/bad performance included quotations referring how the pilot performed the task. The only negative quotation in the first scenario was by N3, who said, "it could have been better as we exceeded the correct altitude by 200 feet". In contrast, the pilots described their good performance in how they kept the plane under control, moved it in the correct direction, managed the speed and altitude, controlled the situation, recovered from the malfunction, analysed the situation correctly, prevented the situation from escalating, followed the flight plan and kept the situation safe.

The second task involved only one clearly negative statement, presented by N7:

When the other guy's display started showing false information, we were in the clouds, meaning I could not see anything and did not know which display was working. I was a bit hasty and decided to cancel the approach. So, I put on full thrust and we started to ascend. However, we had not reached altitude when we need to make the cancellation decision, so I could have taken a few seconds to think about the situation.

The experienced pilots brought up the same aspects of good performance as in the first task, but their positive statements were more cautious: "even though my performance was diamond, a more safe option is always to cancel the approach" (E1), "it was okay, but thinking about it afterwards, I could have done something differently" (E2), "not perfect, but okay" (E3), "the situation was under control, but thinking on it now, I could have made other choices" (E5) and "I am quite happy with my performance" (E6). All experienced pilots mentioned that the break caused by COVID-19 negatively affected their performance. In contrast, the novice pilots proposed that their performance was good because they were able to complete the most important task, i.e. keeping the plane under control. As N3 described: "I think the performance was fine as my foremost job in the approach phase is to fly the plane".

Ability to recall the situation implies how easy or difficult it was to recall the malfunctions that occurred in the tasks. Almost all novice pilots hesitated about which situation we were referring to when we asked pilots to describe the malfunction for the first time. A possible reason why some novice pilots had difficulties in recalling the malfunctions was that their focus had been on flying, not on the overall situation. As described by N1: "If there are many approaches, I can remember what I did in the approach phase, and if the malfunction does not affect to this phase, I cannot remember the details of the malfunction." In contrast, the experienced pilots could recall and describe the malfunction itself, as well as their own actions in the situation, rather accurately.

We identified two ways the tasks stimulated possible ways to enhance professional expertise, *development needs* and *risk awareness*. The tasks led both novice and experienced pilots to contemplate their current expertise and possible

development needs. In both scenarios, the main development needs were practicing more precise control of the aeroplane and clarifying the ideal communication with the monitoring pilot. When regarding the other stimuli, only experienced pilots considered the risks that the scenarios involved. In the first scenario, there was a possibility to drift in the wrong direction or enter an erroneous altitude, which creates a risk of collision with other aircraft. Although it can be assumed that all pilots know of this risk, only experienced pilots explicitly pointed out the possibility. When regarding the second task, the risks were discussed by E7:

> If I had thought that I had the fault on my display, well, at that point, the procedure is that your colleague starts to fly. His display showed that the aeroplane was descending and curving left, and in reality, we would have gone too far left and down into a very steep slide. Then, at that altitude, of course, the GPS warnings start appearing fast. After that, well, the situation could have escalated very quickly.

While all experienced pilots brought this danger up at some stage of the discussion, only N7 and N8 among the novice pilots explicitly considered this risk.

When the pilots discussed their professional expertise, they brought up *experience, improvisation, procedures* and *training*. Not surprisingly, the experienced pilots often mentioned that their experience aids in handling a malfunction situation, as put by E5:

> For me, through this background experience, plane control itself is not really challenging. My experience also provides me with additional capacity. For sure, plane control, for inexperienced pilots, and automatic shut-off and other such tasks, or the disabling of all warnings, must be stressful since it is not routine.

The novice pilots also referred to their experience, but not experience from flight hours in an aeroplane; instead, they meant experience gained from training. They described having recently practiced different malfunction situations for several months, making them accustomed to them, whereas more experienced pilots had been flying routine paths without any technical surprises. Procedures were constantly brought up when the pilots were asked what they based their actions on in a malfunction scenario, especially the golden rule of Airbus: fly, navigate and communicate. Surprisingly, improvisation was mentioned more frequently by novice pilots than experienced ones.

Communication with the co-pilot was referred to as crucial by both experienced and novice pilots. The pilots pointed out that the whole crew needs to know what the others are doing, with decisions made together. Despite this, two pilots (N2, N4) mentioned that they had difficulties in agreeing with the monitoring pilot in the first scenario. The pilots emphasised that the monitoring pilot has important *responsibilities*, such as to observe the correct path, analyse the malfunction and communicate with the air-traffic controller. As such, the pilots emphasised that the monitoring pilot's responsibility is to take care of matters besides flying so that the flying pilot can focus on that alone. The third code here,

support, reflects how all novice pilots brought up at some stage of the interview how the captain (i.e. monitoring pilot) supported their own actions by staying calm, explaining what was going on, giving instructions, leading the situation and acting as a backup in case the pilot lost control of the plane. The captain's important supporting role was well-described by N1:

> In my opinion, what has a great effect is teamwork. Even in a more challenging situation, what the captain does in the neighbouring chair affects the outcome a lot. The captain does not let you fail. Instead, they give you time early in the process to seek some advice, and if the advice is not sufficient, then the captain fixes the situation a little. It helps that you do not get the feeling that everything depends purely on you.

The most important actions in the malfunction scenarios were to *control the aeroplane, prioritise actions, analyse before acting* and *act calmly*. These four codes overlapped, and the importance of taking and keeping control of the plane was brought up by all pilots. They explained that in any situation, the correct speed, altitude and direction need to be reached before any other actions are taken, to prevent any additional dangers from arising, such as drifting into another plane's path. This was evidently important in the first scenario where the malfunction caused the autopilot to turn off, meaning the pilot needed to start flying manually. In the second scenario, this step was then intentionally challenged: before the pilot could control the plane correctly, situation analysis was necessary to identify which one of the two Primary Flight Displays was showing the correct information. Still, the pilots emphasised the importance of keeping the focus on flying, as it was possible to cancel the approach to gain more time to analyse the situation. As such, acting calmly did not only mean the pilot's calm behaviour but also moving the plane into a safe position to buy more time.

The tasks themselves were described either *difficult/easy, familiar/unfamiliar* and *immersive*. As brought up in numerical self-assessment, almost all pilots rated their performance as good in both tasks. This could be observed in the interviews. There were only three quotations that described the first scenario as difficult: the situation was surprising, the malfunction occurred when the aeroplane was in the middle of turning and the correct actions needed to be prioritised. In turn, taking control of the aeroplane when the malfunction occurred was unanimously perceived as easy, something that they are trained to do. The second task was perceived as difficult because the situation was something that the pilots had not encountered before, requiring that the pilot intensively focus on keeping the plane in the correct position, analyse the situation and communicate with the co-pilot at the same time. Then, the pilot needed to decide whether to continue or cancel the approach, and the malfunction occurred in bad weather conditions. The novice pilots seemed to perceive the task as easier than those who were experienced; they saw that their responsibility was to control the aeroplane and let the monitoring pilot focus on troubleshooting the malfunction.

As pointed out by the pilots, the difficulty of the task was strongly connected to how familiar it was. The first task was familiar for all pilots, either from simulator training or real flights. In contrast, the second task was familiar to four experienced

and two novice pilots. Finally, the pilots' quotations considered how immersive the task was. Although the pilots are advised to act in a simulator as they would on a real flight, we identified occasions where the pilots thought that they might have acted differently in real life. Two experienced pilots (E4, E6) commented that they may have been more eager to cancel the approach in the second task in a real situation, for additional safety. As put by E4: "I thought that this might call for cancelling the approach, but I did not know if it was the purpose of this practice, so I decided to continue the approach." The novice pilots pointed out that simulator practice is mentally different because the pilot is prepared for surprising anomalies in the scenario, and the pilot's career, or life, is not in danger in the situation. It was also brought up that while the cockpit itself is an accurate replica of a real aeroplane, the window screen graphics do not equal the real visual environment.

Discussion

Flight training organisations can derive valuable insights from an examination of their assessment and feedback procedures, with potential to reduce the number of student dropouts, for example (Wulle et al., 2020). Drawing upon our findings, certain disparities between evaluated and self-assessed performance came to light. Initially, despite the deliberative design of tasks to challenge even experienced pilots, all participated pilots garnered commendable evaluations from the flight instructor, with novice pilots even outperforming their experienced counterparts. Secondly, the most striking contrast between evaluated and self-assessed performance manifested in the case of one novice pilot, who received the highest evaluation among novice pilots, and still, he self-assessed his performance as poor, describing his experience as a mere attempt of surviving the tasks.

These findings raise a pertinent question regarding the CBTA framework: To what extent can the framework accurately discern developmental needs in pilots' performance? In essence, does the framework provide precise feedback that can effectively support pilot's professional development, as a sound evaluation should? These considerations merit attention when the framework is put into use, extending its validation beyond sole evaluative purposes (Franks et al., 2014; Hattingh et al., 2022; Hodge et al., 2020). In our study, discernable differences between novice and experienced pilots were not apparent in the flight instructor's evaluation. However, disparities between groups became more evident when pilots engaged in reflective debriefing interviews that followed the simulation session. Consequently, there is a compelling need for additional research into flight instructor evaluation methods, and evaluation frameworks, that takes into account the pilot's own view of their performance.

This study contributes also to the limited body of contextual research concerning flight pilots' ability to reflect professional practice (Mavin & Roth, 2014; Mavin et al., 2018). Despite the fact that most of the novice pilots perceived both of the tasks as rather straightforward, it seems that they were overly confident, whereas the experienced pilots were more adept at discerning potential risks inherent in the situation. This disparity can be elucidated by the fact that the experienced pilots, all captains, focused on the comprehensive situational context

over their individual performance, even though responsibility for troubleshooting malfunctions typically falls upon the monitoring pilot.

Moreover, it becomes evident that the novice pilots, to varying degrees, tended to downplay the significance of malfunctions, whether consciously or subconsciously, as they had difficulties in accurately recalling the sequence of events and their specific details during the tasks. This presents a formidable challenge when designing simulator tasks: if the objective is to introduce malfunctions suddenly and unexpectedly, how can the attention of aviation professionals be effectively captured, allowing for subsequent active reflection? This is a crucial question because while technical competencies may be enhanced through iterative practice and correction, the improvement of non-technical competencies necessitates an examination of the situational dynamics, such as exploring what, how and why certain actions were executed, along with critical thinking of whether any future changes should be considered (Cattaneo & Motta, 2021; Mavin & Roth, 2014; Mavin et al., 2018).

Our study also raises a question concerning the design of pilot training: how can simulator sessions be effectively instructed to ensure that pilots comprehend the learning purpose and objectives of the task while still incorporating the element of unpredictability? Pilots are consistently advised to act in a simulator exactly as they would in an actual flight. However, it is noteworthy that some pilots articulated the possibility of responding differently, if they encountered a similar malfunction during a real flight rather than in a simulator. Such disparities may arise if pilots are not entirely comprehending the purpose of the practice. For instance, a pilot might make decisions based on the assumption that task aims to facilitate a challenging landing, even though in an actual flight he would cancel the approach to ensure maximum security. Furthermore, the surprise factor in simulator sessions may be diminished if pilots anticipate that flight instructors deliberately introduce malfunctions and other unforeseen events, given that these occurances often constitute the main purpose of simulator training.

When considering the limitations of this study, it is apparent that this was just a single simulator training occasion, involving only limited cohorts of pilots. However, it is important to acknowledge that flight simulator time incurs substantial costs, and opportunities to conduct research in this environment are rare. The evaluation employing the CBTA was also conducted by just one flight instructor. However, the CBTA framework itself involves the premise that the evaluation should be independent of the evaluator, as it relies on detailed competency descriptions and behavioural indicators. To conclude, our findings should not be extrapolated to imply universal applicability across all pilot training. Rather, they underscore a pertinent concern, indicating that reliance solely on the CBTA framework may not constitute the most judicious approach for addressing potential development needs and knowledge gaps in pilots' professional development, underscoring the continued importance of human debriefing, feedback and reflection within training and evaluation.

Declarations

Ethics approval and consent to participate The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions. The dataset is anonymous and stored in secure cloud services whose server rooms are located at the University of Jyväskylä. The metadata has been stored at the research portal of the University of Jyväskylä (Converis). When we conducted our study, we followed the guidelines of the Finnish National Board on Research Integrity. The Human Sciences Ethics Committee of the University of Jyväskylä made a positive statement about our study (755/13.00.04.00/2020) before we recruited the participants and collected the data.

Competing interests No potential conflict of interest is reported by the authors.

References

Antera, S. (2021). Professional competence of Vocational Teachers: A conceptual review. *Vocations and Learning*, *14*(3), 459–479. https://doi.org/10.1007/s12186-021-09271-7.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. Aerospace Medical Association.

Bereiter, C., & Scardamalia, M. (1993). Surpassing ourselves: An inquiry into the nature and implications of expertise. Open Court Publishing Company.

Bontemps-Hommen, M. C. M. M. L., Baart, A. J., & Vosman, F. J. H. (2020). Professional workplace learning: Can practical wisdom be learned? *Vocations and Learning*, *13*(3), 479–501. https://doi.org/10.1007/s12186-020-09249-x.

Brockmann, M., Clarke, L., Méhaut, P., & Winch, C. (2008). Competence-based Vocational Education and Training (VET): The cases of England and France in a european perspective. *Vocations and Learning*, *1*(3), 227–244. https://doi.org/10.1007/s12186-008-9013-2.

Casner, S. M., Geven, R. W., & Williams, K. T. (2013). The effectiveness of Airline Pilot training for abnormal events. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(3), 477–485. https://doi.org/10.1177/0018720812466893.

Cattaneo, A. A. P., & Motta, E. (2021). I reflect, therefore I Am… a good Professional. On the relationship between reflection-on-Action, reflection-in-action and professional performance in Vocational Education. *Vocations and Learning*, *14*(2), 185–204. https://doi.org/10.1007/s12186-020-09259-9.

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based learning in higher education: A Meta-analysis. *Review of Educational Research*, *90*(4), 499–541. https://doi.org/10.3102/0034654320933544.

Durso, F. T., & Dattel, A. R. (2012). Expertise and transportation. In *T*he Cambridge Handbook of Expertise and Expert Performance, edited by K. A. Ericsson, N. Charness, P. J. Feltovich and R. R. Hoffman, 355–72. Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.020.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, *26*(2), 247–273. https://doi.org/10.1080/158037042000225245.

Eraut, M. (2009). Transfer of Knowledge between Education and Workplace Settings. In Knowledge, Values and Educational Policy: A Critical Perspective, edited by Harry Daniels, Hugh Lauder, Jill Porter, and Sarah Harthorn, 1st ed. New York: Routledge.

Ericsson, K. A. (1998). The scientific study of expert levels of performance: General implications for optimal learning and creativity. *High Ability Studies*, *9*(1), 75–100. https://doi.org/10.1080/1359813980090106.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295x.100.3.363.

European Union Aviation Safety Agency (2020). Appendix to opinion No 08/2019 (A) (RMT.0599). https://www.easa.europa.eu/sites/default/files/dfu/Appendix%20to%20Opinion%20No%2008-2019%20%28A%29%20%28RMT.0599%29.pdf.

European Union Aviation Safety Agency (2012). Certification specifications for aeroplane flight simulation training devices. https://www.easa.europa.eu/en/downloads/1735/en.

Flin, R., & Martin, L. (2001). Behavioral markers for Crew Resource Management: A review of current practice. *The International Journal of Aviation Psychology*, *11*(1), 95–118. https://doi.org/10.1207/S15327108IJAP1101_6.

Flin, R., Martin, L., Goeters, K. M., Hörmann, H. J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, *3*(2), 95–117.

Franks, P., Hay, S., & Mavin, T. (2014). Can competency-based training fly? An overview of key issues for *ab initio* pilot training. *International Journal of Training Research*, *12*(2), 132–147. https://doi.org/10.1080/14480220.2014.11082036.

Hager, P. (2008). Current Theories of Workplace Learning: A Critical Assessment. In N. Bascia, A. Cumming, A. Datnow, K. Leithwood, & D. Livingstone (Eds.), International Handbook of Educational Policy (pp. 829–846). Springer. https://doi.org/10.1007/1-4020-3201-3.

Hattingh, A., Hodge, S., & Mavin, T. (2022). Flight instructor perspectives on competency-based education: Insights into educator practice within an aviation context. *International Journal of Training Research*, *20*(3), 264–282. https://doi.org/10.1080/14480220.2022.2063155.

Hodge, S., Mavin, T., & Kearns, S. (2020). Hermeneutic dimensions of competency-based education and training. *Vocations and Learning*, *13*(1), 27–46. https://doi.org/10.1007/s12186-019-09227-y.

International Air Transport Association (2023). Competency assessment and evaluation for pilots, instructors and evaluators. Guidance material. Second edition. https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/competency-assessment-and-evaluation-for-pilots-instructors-and-evaluators-gm.pdf.

Jin, H., Hu, Z., Li, K., Chu, M., Zou, G., Yu, G., & Zhang, J. (2021). Study on how expert and novice pilots can distribute their visual attention to improve flight performance. *Ieee Access : Practical Innovations, Open Solutions*, *9*, 44757–44769. https://doi.org/10.1109/ACCESS.2021.3066880.

Kearns, S. K., Mavin, T. J., & Hodge, S. (2017). *Competency-based education in aviation: Exploring alternate training pathways*. Routledge.

Marques, E., Carim, G., Campbell, C., & Lohmann, G. (2023). Ab Initio Flight training: A systematic literature review. *The International Journal of Aerospace Psychology*, *33*(2), 99–119.

Mavin, T. J. (2016). Models for and Practice of Continuous Professional Development for Airline Pilots: What We Can Learn from One Regional Airline. In S. Billett, D. Dymock, & S. Choy (Eds.), *Supporting learning across working life* (Vol. 16). Springer International Publishing. https://doi.org/10.1007/978-3-319-29019-5.

Mavin, T. J., & Roth, M. W. (2014). Between reflection on practice and the practice of reflection: A case study from aviation. *Reflective Practice*, *15*(5), 651–665. https://doi.org/10.1080/14623943.2014.944125.

Mavin, T. J., & Roth, W. M. (2015). Optimizing a workplace learning pattern: A case study from aviation. *Journal of Workplace Learning*, *27*(2), 112–127. https://doi.org/10.1108/JWL-07-2014-0055.

Mavin, T. J., Kikkawa, Y., & Billett, S. (2018). Key contributing factors to learning through debriefings: Commercial aviation pilots' perspectives. *International Journal of Training Research*, *16*(2), 122–144. https://doi.org/10.1080/14480220.2018.1501906.

Mayring, P. (2000). Qualitative content analysis. Forum: Qualitative Social Research, 1(2).

Mayring, P. (2014). Qualitative content analysis: Theoretical foundation, basic procedures and software solution. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173.

McLean, G. M., Lambeth, S., & Mavin, T. (2016). The use of simulation in ab initio pilot training. *The International Journal of Aviation Psychology*, *26*(1–2), 36–45.

Mulder, M., Weigel, T., & Collins, K. (2007). The concept of competence in the development of vocational education and training in selected EU member states: A critical analysis. *Journal of Vocational Education & Training*, *59*(1), 67–88. https://doi.org/10.1080/13636820601145630.

Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training. *The International Journal of Aviation Psychology*, *8*(3), 197–208.

Wiegmann, D. A., Goh, J., & O'Hare, D. (2002). The role of situation assessment and flight experience in pilots' decisions to continue visual flight rules flight into adverse weather. *Human Error in Aviation*, *44*(2), 465–474. https://doi.org/10.4324/9781315092898-23.

Wiggins, M. W., & O'Hare, D. (2003). Expert and novice pilot perceptions of Static In-Flight images of Weather. *The International Journal of Aviation Psychology*, *13*(2), 173–187. https://doi.org/10.1207/S15327108IJAP1302_05.

Wikander, R., & Dahlström, N. (2016). The Multi Crew Pilot Licence-Revolution, Evolution or not even a Solution? A review and analysis of the emergence, current situation and future of the multi-crew pilot licence (MPL). Lund University.

Wulle, B. W., Whitford, D. K., & Keller, J. C. (2020). Learning theory and differentiation in Flight instruction: Perceptions from Certified Flight Instructors. *Journal of Aviation/Aerospace Education & Research*, *29*(2), https://doi.org/10.15394/jaaer.2020.1814.

Yu, C. S., Wang, E. M. Y., Li, W. C., Braithwaite, G., & Greaves, M. (2016). Pilots' visual scan patterns and attention distribution during the pursuit of a dynamic target. *Aerospace Medicine and Human Performance*, *87*(1), 40–47. https://doi.org/10.3357/AMHP.4209.2016.

Ari Tuhkala works as a senior lecturer in the Department of Education at the University of Jyväskylä. His research considers learning in technology-enhanced environments and semi-automatic natural language learning tools.

Ville Heilala works as a doctoral researcher in the Faculty of Information Technology at the University of Jyväskylä. He holds two Master's degrees, one in education and one in computer science. His research is about learning analytics and student agency.

Joni Lämsä works as a postdoctoral researcher in the Department of Education at University of Jyväskylä, Finland. In his research, he has adopted the temporal perspective on learning, teaching, and interaction in technology-enhanced contexts by applying multimodal data sources.

Arto Helovuo works as Head of Compliance and Business Development, flight instructor and examiner (TRI/TRE) and A330/350 captain at Finnair. He holds a Master's degree in Human Factors and System Safety.

Ilkka Tynkkynen works as a Chief Theoretical Knowledge Instructor (CTKI), a synthetic flight instructor (SFI) and a co-pilot A330/A350 at Finnair. He holds a Master's degree in business law.

Emilia Lampi works as a project researcher at the Finnish Institute for Educational Research. Her research interests are related to the use of extended reality in competence building and professional development.

Katriina Sipiläinen works as a research assistant in the Faculty of Education and Psychology at the University of Jyväskylä. She holds a Bachelor of Science in Statistics.

Raija Hämäläinen works in the field of technology-enhanced learning at the Center for Research for Learning and Teaching at the University of Jyväskylä, Finland. Hämäläinen seeks to understand learning and professional development by investigating how learning and interaction processes occur and unfold over time with novel methods, such as eye-tracking, heartrate variability and prosodic analysis of voice.

Tommi Kärkkäinen serves as a professor at the Faculty of Information Technology, University of Jyväskylä, leading currently the Learning and Cognitive Sciences division and research group on human and machine-based intelligence. His main research fields include computational sciences and computing education research with special emphasis on data mining, machine learning, and learning analytics. He has published over 180 research papers, led dozens of R&D projects, and supervised over 30 PhD theses.

Authors and Affiliations

Ari Tuhkala[1] · **Ville Heilala[1]** · **Joni Lämsä[1]** · **Arto Helovuo[2]** · **Ilkka Tynkkynen[2]** · **Emilia Lampi[1]** · **Katriina Sipiläinen[1]** · **Raija Hämäläinen[1]** · **Tommi Kärkkäinen[3]**

✉ Ari Tuhkala
ari.s.tuhkala@jyu.fi

Ville Heilala
ville.s.heilala@jyu.fi

Joni Lämsä
joni.lamsa@jyu.fi

Arto Helovuo
arto.helovuo@finnair.com

Ilkka Tynkkynen
ilkka.tynkkynen@finnair.com

Emilia Lampi
emilia.k.lampi@jyu.fi

Katriina Sipiläinen
katriina.m.sipilainen@jyu.fi

Raija Hämäläinen
raija.h.hamalainen@jyu.fi

Tommi Kärkkäinen
tommi.p.karkkainen@jyu.fi

1   Department of Education, University of Jyväskylä, P.O. Box 35, FI-40014 Jyväskylä, Finland

2   Finnair, Helsinki, Finland

3   Faculty of Information Technology, University of Jyväskylä, University of Jyväskylä, Jyväskylä, Finland