

RESEARCH ARTICLE

Assessing evidence-based training in a collaborative virtual reality flight simulator

J.I. Cross  and T. Ryley 

Griffith University Griffith Sciences, Brisbane City, Australia

Corresponding author: J.I. Cross; Email: jamie.cross@griffith.edu.au

Received: 11 March 2024; **Revised:** 20 September 2024; **Accepted:** 25 September 2024

Keywords: collaboration; evidence-based training; multi-crew cooperation; pilot training; virtual reality flight simulator

Abstract

In recent times, there has been increased focus on the utilisation of virtual reality flight simulators in flight training, driven by their advantages compared to conventional methods. However, a paucity of empirical evidence has prevented their widespread introduction and regulatory approval. Existing research focuses on single-user simulators, leaving a gap in studies of collaborative training within virtual environments. Consequently, this paper investigates evidence-based simulator training within a collaborative virtual environment.

A mixed methods approach was adopted, where behaviours related to industry-standard competencies were observed in a virtual reality complex aircraft and thematic analysis applied to a post-experiment participant debrief. The findings showcase the feasibility of utilising a collaborative virtual environment for evidence-based training purposes in scenarios aligned to typical initial First Officer airline training programmes, which is a precursor to supplementing traditional professional pilot training techniques. In addition, the study found that the visual barriers imposed by head-mounted displays were overcome through the adoption of refined communication strategies, thus laying the groundwork for physically separated multi-crew pilot training.

Nomenclature

AI	Artificial Intelligence
APU	auxiliary power unit
B738	Boeing 737-800
CASA	Civil Aviation Safety Authority (Australia)
CBT	competency-based training
EBT	evidence-based training
FCOM	Flight Crew Operating Manual
HMD	head-mounted display
ICAO	International Civil Aviation Organization
MCC	multi-crew cooperation
SET	simulation for experiential training
SQLR	systematic quantitative literature review
VE	virtual environment
VR	virtual reality
VRFS	virtual reality flight simulator

1.0 Introduction

Since the advent of aviation, flight simulators have been an integral and essential part of flight crew training [1]. Virtual reality (VR) is increasingly being used for gaming and has provided multiple

opportunities in industries such as education, fashion, healthcare and tourism [2, 3]. With recent developments of immersive technologies, virtual reality flight simulators (VRFSs) are increasingly attracting attention (e.g. Ref. [4]) since they offer significant advantages over traditional training methods including their versatility, mobility, increased throughput and reduced size and cost [5]. In addition, the effectiveness of immersive training supports improved psychomotor performance, knowledge acquisition, engagement and spatial ability when individuals or teams are trained with VR [6].

Although the feasibility of VRFSs for flight training has been demonstrated (e.g. Refs. [2, 7–9]), it is limited to single-user simulators. Most commercial pilot operations employ at least two pilots, and their collaboration is paramount in any multi-crew operation [10–12]. Since evidence-based training (EBT) is a global safety improvement initiative that is being introduced for multi-crew professional pilot training [13], the efficacy of collaborative VRFS employing EBT must be researched.

The overarching aim of this paper is therefore to explore the feasibility of incorporating EBT in a multi-crewed VRFS using the application of professional pilot training competencies. While competencies will be observed against standards prescribed by the International Civil Aviation Organization (ICAO) [13], this study's focus lies in assessing the viability of a VRFS for EBT while also offering a comparative analysis against a non-immersive environment, namely a desktop simulator. The study does not aim to gauge proficiency in specific competencies; rather, it employs ICAO competencies as a benchmark to ascertain the feasibility of employing a VRFS for EBT purposes.

To assess competencies such as teamwork, it is important to ensure collaboration between two pilots in a VRFS, and that the system provides adequate fidelity and usability to complete the required tasks. To further align the research to commercial pilot operations, the study simulated a multi-crewed Boeing 737-800 (B738) aircraft with exercises aligned to that of a typical initial First Officer airline training programme. For comparison, competencies were measured in a virtual environment (VE) and a desktop simulator.

2.0 Literature review

2.1 Collaboration

Multi-crew cooperation (MCC) is the collaboration, or teamwork, required by pilots to operate in a multi-crewed aircraft. Operationalisations of the terms collaboration and teamwork are often blurred, especially when describing the interactions of pilots on a flight deck. Although the difference is subtle, this paper will predominantly use the term collaboration, since this research is focused on investigating the competencies of participants working together in a VE to achieve a shared goal.

This ambiguity has led to many interpretations of collaboration, but a generic definition put forward by Ref. [14] that has withstood time is that collaboration is 'a process in which two or more agents work together to achieve a shared goal' (p. 67). Therefore, the flight crew can be considered as two or more pilots who are responsible for individual tasks, yet they must collaborate to operate an aircraft efficiently. Since aviation is safety-critical, the shared goal exceeds the constituent parts, analogically to that of Gestalt theory which emphasises that the whole of anything is greater than its parts [15]. When working within a team, collaborators bring their individual expertise to achieve a shared goal [16]. Without the individuals willing to share their expertise through collaboration, achieving the shared goal may not be possible. Therefore, the collaborative team is not merely a summation of individuals, but a 'distinguishable set of two or more people who interact dynamically, interdependently and adaptively toward a common valued goal/object/mission, who have each been assigned specific roles or functions to perform and who have a limited life span of membership' (Ref. [17], p. 4). More recent research has focused on teamwork and the importance of achieving the shared goal. Teamwork is subsequently defined as a set of interrelated thoughts, actions, and feelings of each team member that are needed to function as a team and that combine to facilitate coordinated, adaptive performance and task objectives resulting in value-added outcomes (e.g. Refs [16, 18]).

2.2 Event driven training

It is essential to measure the suitability and transferability of the skills gained in simulators for real-world flight operations. Traditional training sessions have focused on the application of manoeuvres and procedures through repeated exposure to known emergencies based on the evidence of hull losses from early generation jets [13]. The assumption was that to mitigate a risk, simply repeating an event in a training programme was sufficient. Over time, many new events occurred and the subsequent addition of these events to the training requirements saturated recurrent pilot training programs and created an inventory or 'tick box' approach to training [13]. Regulations also dictate that pilot training be standardised and scenarios are therefore relatively predictable. Not surprisingly, trainees became aware of this list of events, and subsequently rehearsed them, which led to an inability to recognise and subsequently react in a suitable manner to variations of those events when they unexpectedly occurred in real life. Skills taught in this manner are 'brittle' as opposed to adaptive and transfer well to predictable situations like tests, but they may not be sufficient in emergency situations, which are typically novel and unexpected [19].

There are many examples where simulator-taught skills were inadequate and led to fatal accidents, including the notorious case of Air France 447, an Airbus A330 with 228 occupants that fatally crashed into the Atlantic Ocean on 1st June 2009. The cause was attributed to the crew, who failed to recognise that the aircraft had stalled and subsequently made inappropriate control inputs that destabilised the flight path even though stall recovery was possible [20]. Although the crew had practiced stall recovery on numerous occasions in the airline's fixed-base simulator, one explanation for the inappropriate control input is that the training and testing for these situations has become a highly predictable routine with pilots often aware of what to expect [21–23].

2.3 Evidence-based training and pilot competencies

Event-driven simulator training is currently being superseded by the concept of competency-based learning called evidence-based training (EBT) [13]. Although EBT and assessment is still based on operational data, it is characterised by developing and assessing the overall capability of a trainee across a range of core competencies, rather than by measuring their performance in individual events or manoeuvres. Competencies are a combination of knowledge, skills and attitudes required to perform a task to the prescribed standard. EBT is designed to ensure trainees possess the required competencies rather than demonstrate the minimum skills required for an event. An industry-wide consensus led to the implementation of EBT in order to reduce aircraft accident rates after reviewing existing training for airline pilots [13]. Therefore, the goal of an EBT program is to identify, develop and assess the competencies required by pilots to operate safely, effectively and efficiently in a commercial air transport environment by managing the most relevant threats and errors based on evidence collected in operations and training.

Modern aviation systems are highly reliable, but also complicated. It is therefore impossible to foresee all plausible accident scenarios. EBT addresses this by moving from pure event-based training to prioritising the development and assessment of key competencies [13]. A recent enhancement of EBT is the introduction of simulation for experiential training (SET) which provides a more focused framework than operational and simulator data [24]. Mastering a finite number of competencies should allow a pilot to manage unforeseen situations in flight and for which the pilot has not been specifically trained. EBT and SET therefore avoid rote-memorised skills which may not enable pilots to identify and respond in an appropriate manner during a real flight [21], potentially avoiding accidents such as Air France 447. However, there has been criticism of competency-based training (CBT) and assessment methods, which is a key component of modern EBT. For example, Ref. [25] propose that while CBT may be used appropriately for initial development of physical flying skills for ab initio level pilots, its application is limited in areas of training which require complex decision-making and critical judgement because it fails to comprehensively address the full range of training requirements.

Since the publication of a systematic quantitative literature review (SQLR) by Ref. [5], there has been little empirical evidence collected on the use of any type of VRFSs. Evidence of multi-crewed VRFSs is limited to Ref. [26] who demonstrated the fidelity and useability of a collaborative VRFS, as well as reduced participant workload (effort) and heightened team situational awareness in the VE compared to a desktop flight simulator. Therefore, a similar methodology will be utilised in this study. There is no known literature on VRFS employing EBT, although, for example [27], provide a proof of concept for gamification of flight instructor learning in EBT scenarios.

2.4 The importance of presence in a VRFS

A large proportion of the immersive experiments examined in the SQLR were only partially successful due to issues with the technology that led to a lack of presence, which in turn created an unrealistic user experience. Therefore presence, the adopted definition of which is a feeling of ‘being there’, is essential for successful simulation training in a VE. Composed of immersion and involvement factors, presence was successfully demonstrated in a single-pilot VRFS [2], and most critical to this study is that presence can be demonstrated in a multi-crewed VRFS.

3.0 Method

3.1 Measuring collaboration

Since there is no consensus on what constitutes collaboration, various survey instruments are used to assess collaborative dimensions. To determine appropriate measurement tools, a conceptual model must be applied that is relevant to the team task. A ‘generic’ teamwork model was suggested by Ref. [16] that, regardless of the team task that is being examined, has components that are found in almost all teamwork taxonomies. Their ‘Big Five’ model is composed of the components team leadership, mutual performance monitoring, backup behaviour, adaptability and team orientation. However, Ref. [16] also concur with Ref. [28] in that teamwork is dynamic and its manifestation can vary based on many variables, including the team environment, type of task, individual differences and perceived workload. Therefore, to fully understand team performance, academics typically stipulate that it is insufficient to take a single snapshot of team performance such as a post-experiment questionnaire (e.g. Refs [29, 30]). Instead, performance should be sampled during a variety of conditions and situations [16].

Measuring collaboration requires a conceptual consistency between the method employed and the theory explored in the research context [31]. Furthermore, methods for assessing teamwork skills generally involve the use of observer-based rating scales, which often are specific to the task at hand [32]. Rating scales have the distinct advantage of continuous assessment throughout the task, although they are subject to rater bias (discussed later). Specifically for aviation, behavioural markers are commonly used in training and crew resource management [33, 34]. The behavioural markers are descriptions of observable, non-technical behaviours that are present in teams or individuals. Therefore, this study adopted the various recommendations suggested in the literature by using live teamwork observation coupled with post-experiment analysis to measure collaboration as follows:

- Observation of the pertinent behavioural indicators such as leadership and teamwork during the experiment
- Post-experiment participant feedback on how well they worked together
- Successful completion of a scenario, which necessitated collaboration

3.2 Evidence-based training behavioural competencies

EBT was assessed against the core-competency framework proposed by ICAO [35]. The eight competencies and their associated descriptions are shown in Table 1. The table also includes a sample

Table 1. *The eight pilot competencies observed in this study*

	ICAO competency	ICAO competency description	Sample behavioural indicators/observables
1	Communication	Demonstrates effective oral, non-verbal and written communications, in normal and non-normal situations.	Selects appropriately what, when, how and with whom to communicate.
2	Aircraft flight path management – Manual control	Controls the aircraft flight path through manual flight, including appropriate use of flight management system(s) and flight guidance systems.	Controls the aircraft manually with accuracy and smoothness as appropriate to the situation.
3	Aircraft flight path management – Automation	Controls the aircraft flight path through automation, including appropriate use of flight management system(s) and guidance.	Controls the aircraft using automation with accuracy and smoothness as appropriate to the situation.
4	Leadership and teamwork	Demonstrates effective leadership and team working.	Engages others in planning and allocates activities fairly and appropriately according to abilities.
5	Problem solving and decision making	Accurately identifies risks and resolves problems. Uses the appropriate decision-making processes.	Employs proper problem-solving strategies.
6	Application of procedures	Identifies and applies procedures in accordance with published operating instructions and applicable regulations, using the appropriate knowledge.	Correctly operates aircraft systems and associated equipment.
7	Workload management	Manages available resources efficiently to prioritise and perform tasks in a timely manner under all circumstances.	Manages time efficiently when carrying out tasks.
8	Situational awareness	Perceives and comprehends all the relevant information available and anticipates what could happen that may affect the operation.	Identifies and assesses accurately the state of the aircraft and its systems.

of the behavioural indicators/observables that were observed, a full list of which can be found in Ref. [13, p. II-App 1-1].

3.3 Measuring participant feedback

It is important to understand from the participant's perspective how user-friendly and lifelike the experiment interfaces were. It is also valuable to obtain participant feedback on how well they feel they worked together and addressed the competencies. For this purpose, a post-experiment debrief was utilised. The

qualitative experiential aspect of this study therefore focused on participant interpretations of their experiences, as discussed by Ref. [36]. This was achieved using semi-structured reflective debriefs to elicit detailed descriptions of participant experiences during the study.

The debriefs were led by the researcher and undertaken with both participants simultaneously upon completing the experiment and lasted between 10 and 20 mins with each participant pair. All debriefs were audio-recorded and then subjected to thematic analysis. In addition, the researcher took notes during the debriefs. The structure of the debrief composed of asking participants if they knew each other, discussing previous flight, simulator and VRFS experience, their experience during the experiment in the VE or real-world, and other pertinent questions.

As part of a qualitative research paradigm, thematic analysis of the reflective debriefs was used to generate the key descriptive themes in the participant's narrative using the six-phase guidelines and checklist described by Ref. [36]. The first phase involved transcribing each reflective debrief, which included initial familiarisation and examining the breadth and depth of the data. This was performed by reading all transcriptions in a single pass, and non-relevant narrative was removed at this point. For example, in one instance, a participant excused himself momentarily from the debrief. In his absence, the conversation with the remaining participant was not related to the study and was deleted from the transcript. When the participant returned, the reflective debrief resumed, and the transcript continued.

The second phase involved creating initial thematic codes. These codes were based on meanings and patterns observed in the transcript. During this phase, a semantic approach was employed, in that the codes were identified within the explicit or surface meanings of the transcript, and the analysis did not look for anything beyond what had been transcribed. Codes were independently verified by a second researcher. As discussed by Ref. [37], all codes were emergent since the reflective debrief structure composed of non-leading and unbiased prompts. Each individual transcript was analysed in its entirety before moving to the next one.

The third phase involved selecting which excerpts to code, still employing a semantic approach, underpinned by an essentialist paradigm [36]. This was achieved by reading through the transcripts, identifying interesting excerpts and applying appropriate codes to them. Excerpts with the same meaning had the same code applied to them. New codes were added as necessary. The codes were generated inductively, allowing them to be driven from the transcript rather than trying to fit it into a pre-existing coding frame or the researcher's analytic preconceptions. The fourth phase was to collate all excerpts associated with a particular code. This allowed all the excerpts for a given code to be analysed and compared against each other. Some adjustment and revision of the codes was made during this phase. During the fifth phase, all the excerpts associated with a particular code were grouped together into a theme. The sixth phase revealed the themes discovered by the thematic analysis and is discussed in the Results section.

In terms of presence, Ref. [2] employed a quantitative approach to measure the feeling of 'being there'. However, to reduce the number of measures imposed on participants, this study adopted a qualitative approach in which participants were invited to discuss their feeling of 'being there' in the semi-structured reflective debriefs (as discussed by Ref. [2]).

3.4 Scenario development and competency observation

Scenarios were developed to ensure that high levels of interdependence were incorporated so that the 'Big Five' teamwork processes [16] would apply. It was also necessary to create scenarios sufficiently challenging to demonstrate the competencies prescribed by ICAO, yet not overly complicated given the limited flying experience and aviation knowledge of participants. The concepts of pilot flying (PF) and pilot not flying (PNF) roles were utilised in this study and implemented as follows:

- The PF occupied the left seat. The PF operated the flight controls of the aircraft and was responsible for all the activities which directly affected flight path management (i.e. taxiing and flying the aircraft). The PF was also responsible for confirming the actions of the PNF.

- The PNF occupied the right seat. The PNF monitored the course of the flight and was responsible for reading and actioning the checklists, navigation (on the ground and in the air), all radio communication, landing gear and flap operation and generally assisting the PF as necessary.

The scenarios in this study were based on a typical initial First Officer type rating and the ICAO Evidence-Based Training Implementation Guide [38]. A type rating assumes a level of flying proficiency and aviation knowledge. This study was not concerned with a participant's flying proficiency or aviation knowledge, but rather the exploration of EBT in a multi-crewed VRFS. Therefore, this study did not obligate the assumptions of the type rating. In addition, the lack of professional pilots in this study did not hinder the concepts that were measured, which are independent of operational experience.

Participants only received sufficient instruction to adequately complete each phase of their allotted scenario. For example, aircraft systems (electrical, pneumatic, hydraulic, fuel) and the flight management system were not explained, but basic operational use of the autopilot and autoland was provided. This, along with the supporting materials, provided sufficient information to enable all participants to complete their designated scenario. In addition, some allowance was made for the lack of professional pilot experience among participants during measurement of the ICAO behavioural indicators/observables. For example, participants were not expected to adhere to standard radiotelephone phraseology or manage the flight path with any degree of accuracy. In addition, although participants were observed against each ICAO competency and the same level of scrutiny consistently applied by the same researcher across the whole experiment, they were not experienced commercial pilots, so it was not expected that they could complete the ICAO competency to the same standard as commercial pilots.

3.4.1 The type rating

Preparation for a type rating comprises theoretical and simulator training components that cover areas such as aircraft systems, flight procedures, aircraft handling and crew cooperation for a particular 'type' (make and model) of aircraft [39, 40]. Since these programmes can be several weeks in duration, for realism and potential transferability, this study duplicated two typical simulator sessions, one an exercise involving aircraft handling and normal procedures, and the other an exercise involving non-normal and emergency procedures. These two simulator sessions will be referred to as scenarios.

3.4.2 Scenario competencies

Both scenarios involved the operation of a B738, which is categorised as a third-generation jet by ICAO [38]. Therefore, the scenarios were designed with reference to the training modules for third-generation jets in this manual and the ICAO Manual of Evidence-Based Training [13]. A scenario typically involves all eight ICAO competencies to a greater or lesser extent. However, the two scenarios in this study were designed to focus on the specific competencies shown in Table 2. Participants were required to perform all actions specified in the scenario. A summary of the scenarios is as follows:

- Scenario 1 (aircraft handling and normal procedures): based at Heathrow Airport during daylight, in a heavy rainstorm, and involved pushback, taxi, takeoff, a circuit, landing and parking.
- Scenario 2 (non-normal and emergency procedures): based at Sydney Airport during daylight in CAVOK (i.e. fair weather), and involved starting the Auxiliary Power Unit (APU), pushback, starting the main engines and APU shutdown, taxi and managing an aborted takeoff due to an engine fire.

To increase realism, various static aircraft were present in the scenario, and also eight operational aircraft controlled by artificial intelligence (AI).

Table 2. *Scenario competencies*

ICAO competency	Scenario 1	Scenario 2
	Typical initial first officer airline training program structure	
	'Sim-A' (exercises involving aircraft handling and normal procedures)	'Sim-B' (exercises involving non-normal and emergency procedures)
1 Communication	✓	✓
2 Aircraft flight path management – Manual control	✓	
3 Aircraft flight path management – Automation	✓	
4 Leadership and teamwork	✓	✓
5 Problem solving and decision making		✓
6 Application of procedures	✓	✓
7 Workload management	✓	
8 Situational awareness	✓	✓

3.5 Description of the system

Two computers were set up alongside each other, both running the X-Plane Flight Simulator. The left computer was for the PF (as viewed from behind the PF, or from the flight deck door) and the right computer for the PNF. The two computers were networked and set up as 'master' (for the PF) and 'slave' (for the PNF) within X-Plane to allow independent views of the same simulation; that is, the PF controlled their own view, and the PNF controlled their own view. See Fig. 1.

The PF controlled the aircraft using a yoke, throttle quadrant and rudder pedals connected to the master computer. The PNF operated the cockpit switches, buttons and levers using a mouse, also connected to the master computer. Any simulation flight data (e.g. location, speed, altitude) or movement of switches was fed from the master to the slave computer. This enabled both PF and PNF to be in the same simulation and cockpit environment. A researcher sat behind both the PF and the PNF and could monitor their screens, and also follow the simulation on a tablet.

Three configurations of the real-world and the VE were used in the study, as shown in Table 3, and Figs 2, 3 and 4. Participants immersed in the 'Mixed VR/Desktop' configuration can be seen in Fig. 5.

3.6 Materials developed for the study

Substantial supporting materials were developed for the study, summarised below:

- Participant briefing PowerPoint slides used pre-experiment
- Flight crew operating manual (FCOM), based on a typical commercial airline's B738 FCOM, although significantly simplified
- Scenario guides (objectives and instructions), one for each for scenario
- Apron and aerodrome ground movement charts, obtained for both Heathrow and Sydney airports
- Checklists, based on a real B738 checklist

In any configuration that involved the real-world, participants had access to hardcopy documentation including the checklists. In any configuration that involved the VE, participants were restricted to virtual checklists and reliant on the researcher to provide additional guidance where necessary (discussed later).

Table 3. Virtual environment and real-world configurations

Configuration	Description
Dual desktop	Both the PF and PNF operated in the real-world (i.e. both using a desktop simulator).
Mixed VR/Desktop	The PF operated in the VE and PNF operated in the real-world.
Dual VR	Both the PF and PNF operated in the VE (i.e. both using a VRFS).

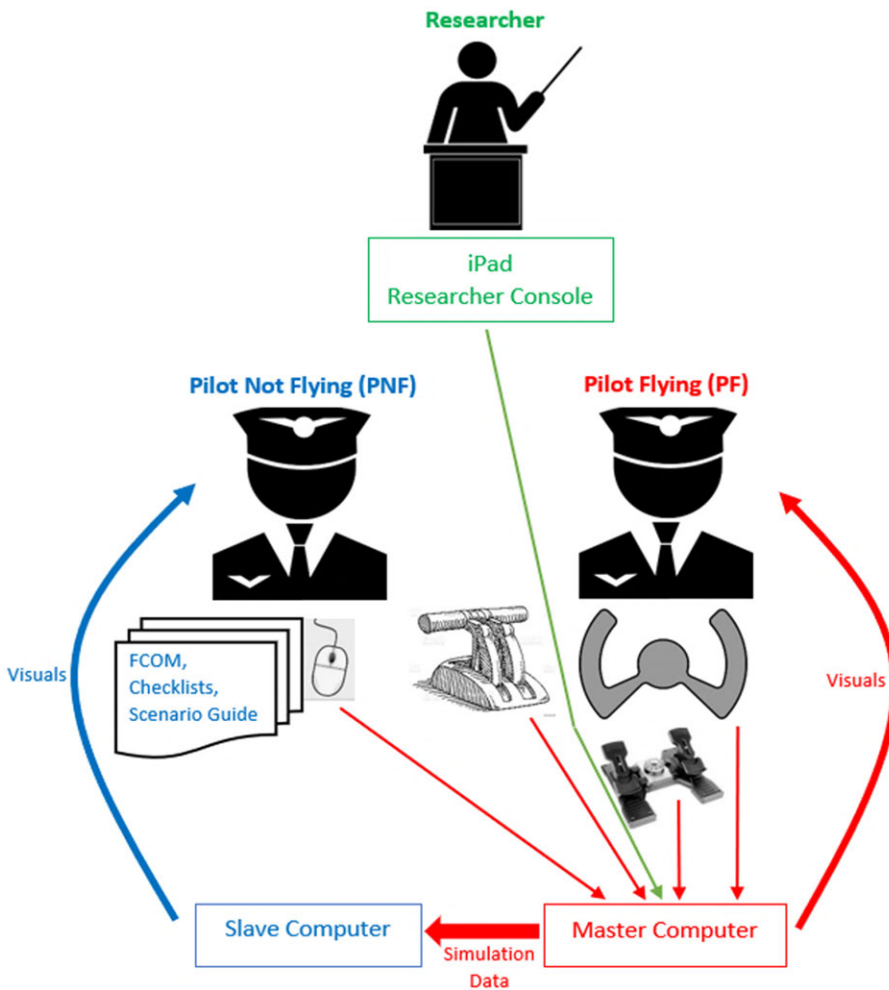


Figure 1. Hardware configuration and data flow.

Furthermore, Excel-based electronic competency observation sheets were generated for both scenarios, intended for the researcher’s utilisation. While the primary focus of this study wasn’t on assessing participants’ competency levels in specific areas, but rather on evaluating the feasibility of a VRFS for EBT, these sheets offer insights into the effectiveness of various competencies within the system. The sheets were based on those presented by ICAO (Ref. [38], pp. 40–45 & p. 87). The spreadsheets were aligned to the scenario guides, and broken down into phase of flight (e.g. Taxi, Takeoff) since the breakdown is seen as advantageous [11]. Within each phase of flight, a list of the ICAO competencies being observed was presented. An example is given in Fig. 6. Note that it was not appropriate to complete

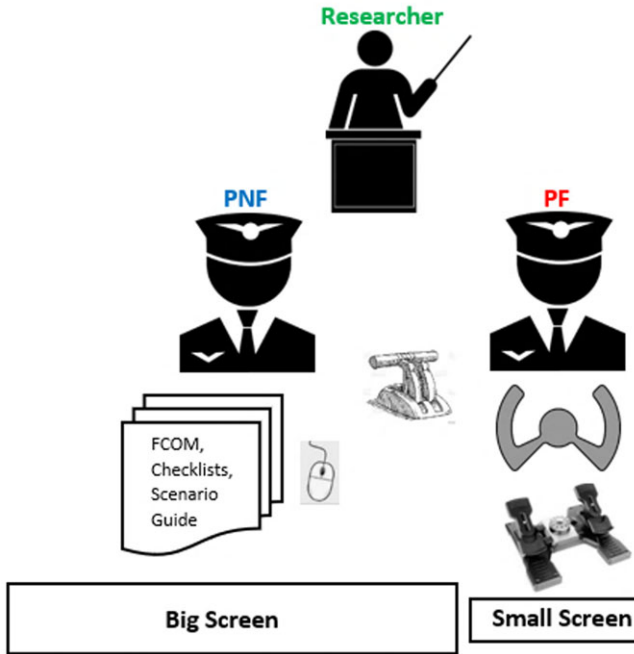


Figure 2. 'Dual Desktop' configuration.

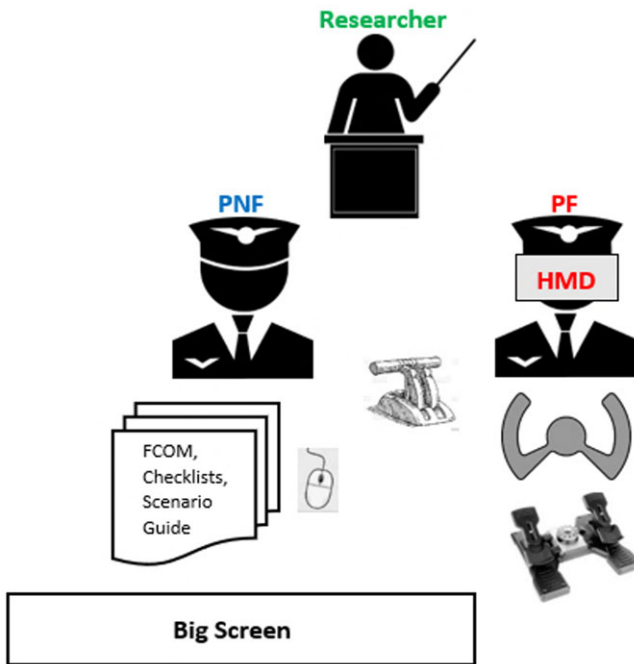


Figure 3. 'Mixed VR/Desktop' configuration.

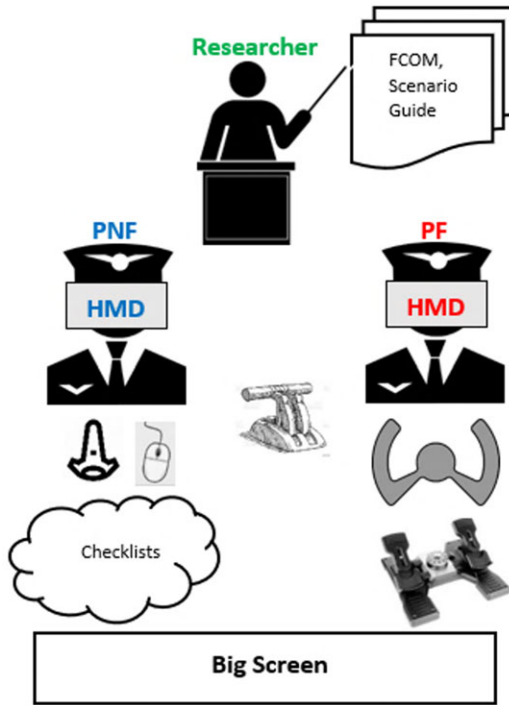


Figure 4. 'Dual VR' configuration.

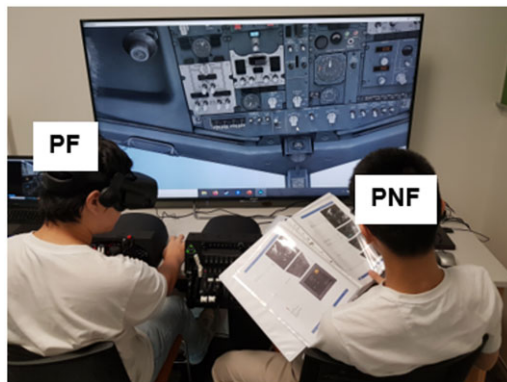


Figure 5. Participants in the 'Mixed VR/Desktop' configuration.

every ICAO competency for both the PF and the PNF for every phase of flight. For example, the PF's responsibility was to taxi the aircraft and was subsequently observed against the ICAO competency Aircraft Flight Path Management – Manual Control (AFPM – Man) during taxi. Since the PNF was not required to taxi the aircraft, they were not assessed against this ICAO competency.

3.7 Experimental procedure

A within-participants design was utilised which involved four stages: pairing and scenario assignment, pre-simulation, flight simulation and post-simulation, as follows:

Phase	Competency	1	2	3	4	5	1	2	3	4	5
		PF					PNF				
Taxi out	Commun										
	AFPM - Man										
	Lead/Team										
	App Procs										
	Sit Aware										

Figure 6. An extract of the competency observation spreadsheet.

3.7.1 i. Pairing and scenario assignment

In the first stage, participants were randomly paired together, and within the pair, randomly assigned to act as the PF or the PNF. Participant-pairs were assigned to either the first or the second scenario and allocated to one of the three configurations in a cyclic fashion.

3.7.2 ii. Pre-simulation

In the second stage, participant-pairs were asked to read an information sheet and provide their consent. They then completed a pre-simulation demographic survey, after which they were given a short presentation which was completed at a spare workstation to allow practice with the flight controls. Participants indicated familiarity with the simulator and head-mounted display (HMD) (if used) by demonstrating fluid and proficient actions i.e. participants demonstrated confidence with the use of the controls (yoke, throttles, rudders, mouse, etc.) and the flight simulator environment. There were no measures taken at this stage, but adequate familiarity was deemed to have occurred after about 10 mins, which established a baseline of experience. This reduced confounds or unrelated performance errors due to the effects of familiarity, apprehension or novelty bias.

3.7.3 iii. Flight simulation

In the third stage, participant pairs were taken to the two flight simulator computers where use of the flight controls was demonstrated once again, and the scenario objectives were restated. If the PNF was operating in any configuration that involved the real-world (see Figs 2 and 3), that participant was provided with the necessary materials (e.g. scenario guide, checklists, charts and paper/pen). If either or both of the PF and/or the PNF were assigned to operate in any configuration that involved the VE and use VR (see Figs 3 and 4), they donned the HMD and were shown how to adjust the strap fittings and the interpupillary distance. Participants were advised to stop the simulation if they felt any simulator sickness and then asked to complete their designated scenario.

During the experiment, the researcher acted as air traffic control and provided some cues, although unnecessary interference was avoided, and conformed to ICAO instructor training guidelines (Ref. [13], p. I-7-2, para. 7.4.3). The researcher also took observational notes and completed the appropriate competency observation sheet (see Fig. 6) by recording grades against each ICAO competency for all phases of the scenario's flight. Marking was performed consistent with the competency grades (range 1 – low to 5 – high) as described by ICAO (Ref. [13], p. 150). Where any low or high scores were recorded, these were annotated, as were general observations, for use in the reflective debrief.

A concern for ensuring the quality of any system for rating pilots' behaviour is the reliability of the raters' judgements [34]. A degree of bias or systematic error can be expected in any performance rating task, arising from personal interpretation, scale use and biases due to motivation. However, such errors are largely mitigated by using the same rater (in the case of this study, the researcher) across the whole experiment who consistently applied the same level of marking as described by ICAO (Ref. [13], p. 150). The researcher was previously employed as a commercial check pilot and is proficient in this process.

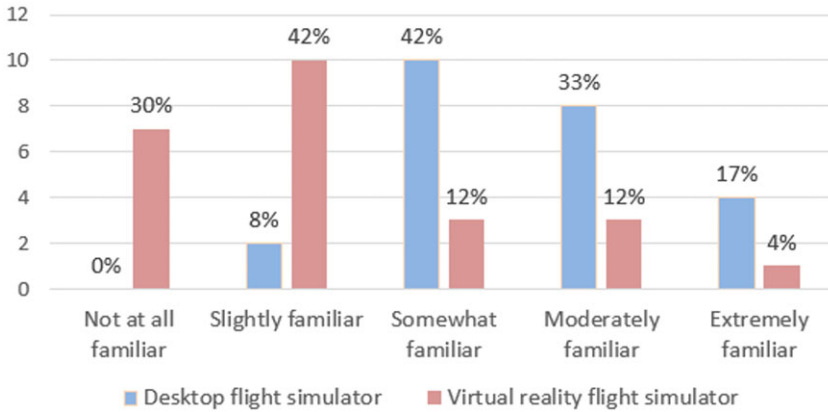


Figure 7. Number of participants familiar with desktop simulators and number of participants familiar with virtual reality flight simulators.

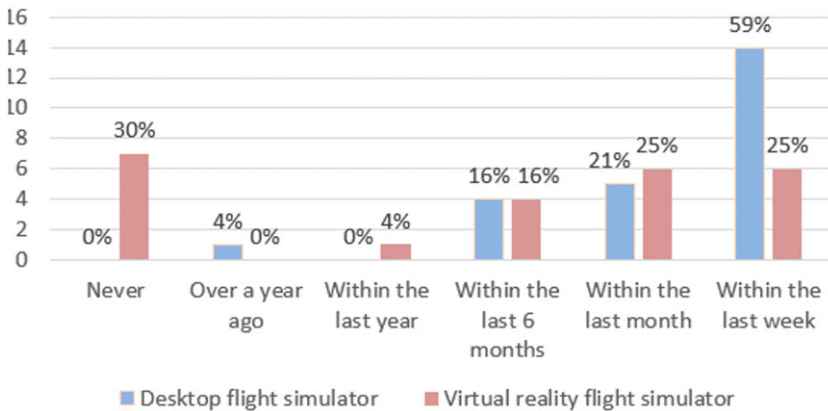


Figure 8. The last time participants used desktop simulators and the last time participants used virtual reality flight simulators.

3.7.4 iv. Post-simulation

In the fourth stage immediately after completion of the scenario, participant pairs were asked to take part in the reflective debrief, driven by the predefined structure and researcher annotations compiled during the experiment.

4.0 Results

4.1 Participants

After ethical clearance was obtained from Griffith University, data was collected from 24 participants sourced from the Griffith University aviation programmes and the local flight school utilised by the university for flight training. One participant already had some experience as a commercial pilot. The participants composed of males ($n = 21, 88\%$) and females ($n = 3, 12\%$) between the age of 18 and 42 ($M = 21.1, SD = 4.8$). Post-experiment debriefs revealed that most participant-pairs did not know each other before the study (8 pairs; $n = 16, 67\%$).

Familiarity with desktop simulators and VRFSs is shown in Fig. 7, and the most recent time desktop simulators and VRFSs were used by each participant is shown in Fig. 8. Most participants had some

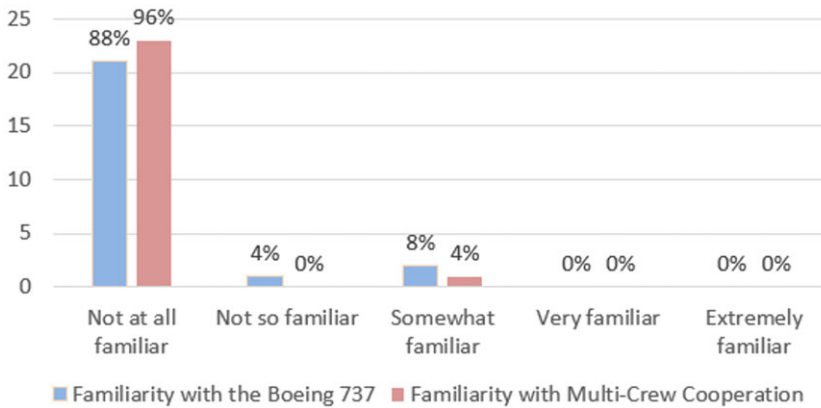


Figure 9. The number of participants familiar with the B738 and the number of participants familiar multi-crew cooperation.

actual flying experience ($n = 21$, 88%) in light, single-engine aircraft, and only one participant had multi-engine turbine experience. Familiarity with a B738 and MCC is shown in Fig. 9. The B738 familiarity was revealed in the post-experiment debrief to be recreational simulator usage, that is, no participants had any actual B738 experience.

As determined by the participant pairing, there was an even distribution of participants across the two scenarios and the three configurations. In total, 12 participants acted as PF and 12 acted as PNF. Also, in total, 12 participants operated in any configuration that involved the real-world, and 12 operated in any configuration that involved the VE. Although this study did not measure simulator sickness, no participants reported any sickness symptoms. Even though the sample size was small, which could lead to more type II errors (i.e. an increase in false negatives), it was deemed appropriate for the aims of the study.

4.2 Pilot competencies

Given that this study aimed to evaluate the feasibility of a VRFS for EBT rather than assessing participants' specific competencies, statistical analysis on individual competency observation data was not conducted. Moreover, participants generally lacked professional experience in these competencies, and the small sample size would hinder meaningful interpretation (i.e. four participant-pairs in each configuration). Instead, competency observation served to determine which aspects of scenarios worked well and steer post-experiment debriefs.

In order to complete a scenario, participants had to utilise all the competencies outlined in Table 2 to some degree. Given that all participant pairs successfully completed their designated scenarios, the study illustrates the efficacy of a VRFS in eliciting professional pilot training competencies, although they may not be to industry standards. As anticipated, certain competencies were more effectively demonstrated in one scenario compared to the other. For instance, Scenario 2 (Sydney) provided optimal conditions for observing competencies such as Application of Procedures, Communication and Leadership and Teamwork, given the involvement of relatively intricate checklists during engine and APU start procedures. Similar to how EBT highlights common problem areas to airline management during pilot training and assessment, incorporating pilot competencies in this study helped identify which aspects of the scenarios worked well for participants and which did not.

4.3 Reflective debriefs and observations

Scenario 1 (Heathrow) required more researcher guidance than scenario 2 (Sydney) because it involved some flying and a landing, and participants had very little B738 operating experience. Similarly, design

limitations imposed some restrictions in the VE, including restrictive checklists (items had to be presented one at a time), inability to take notes and the unavailability of scenario guides and charts. Although this meant that more researcher guidance was given to participants in the VE, there was no evidence in the audio recordings to suggest that this affected their ability to complete the scenario. In addition, the guidance provided was limited in nature (i.e. to specific B738 operational issues and navigation), and as such did not interfere with the researcher's objective observations.

The themes discovered by the thematic analysis and researcher observations are discussed in the following sections.

4.3.1 Leadership, teamwork and collaboration

In both of the 'Dual Desktop' and 'Dual VR' configurations, participants reported, and were observed, to engage in a constant flow of communication with one another, worked well together as a cohesive team, and checklist actions flowed from one item to the next. However, in the 'Mixed VR/Desktop' configuration, the synergy was less apparent. There appeared to be a lack of unison in this mixed environment, and participants – the majority of whom had not met before the experiment – appeared to bond less and there was greater miscommunication.

The following two narratives demonstrate the differences between participant's self-observed synergy in a 'dual' environment compared to a 'mixed' environment:

Scenario:	2 – Sydney
Configuration:	'Dual Desktop'
PNF (Real-World):	<i>I think we worked well together. We got better. It started to flow better as we got the hang of it.</i>
PF (Real-World):	<i>I agree. Even though I've never done this before.</i>
Scenario:	1 – Heathrow
Configuration:	'Mixed VR/Desktop'
PNF (Real-World):	<i>...he seemed to be having more fun in there [referring to the PF operating in the VE]. I felt outside. Not really part of things.</i>

On two occasions, when participant pairs were using the 'Mixed VR/Desktop' configuration, it was observed that the VE participant tended to take the lead in the scenario (e.g. took control of the taxiing navigation or dealing with the engine fire), irrespective of their PF or PNF status and even if their role did not warrant this. A possible explanation for the VE participants taking the lead is that they felt more involved in the scenario and experienced more situational awareness than their real-world co-pilot [26].

4.3.2 Presence, immersion, involvement and situational awareness

The debriefs revealed positive feedback on presence (i.e. 'being there') for participants immersed in the VE, especially for the 'Dual VR' configuration where both participants utilised VR. Participants using VR reported being immersed in the VE, whereas those operating in the real-world reported a feeling of not being involved in the flight.

Participants in the VE stated that they were very aware of the surroundings in the simulation environment. For example, in the first scenario (Heathrow, in a heavy rainstorm) some participants ($n = 3$, 13%) stated that they could almost 'feel' the rain around them, which was made more intense with

the noise of the rain ‘drumming’ against the aircraft windshield. Real-world participants had an opposite experience, with participants recalling an unidentified background noise, and observing the rain on the monitor/screen in front of them which tended to diminish their experience due to blurring. It can therefore be concluded that the soundscape provided by the VE contributed to a sense of presence.

In the following narrative, participants have just settled themselves at the beginning of the scenario:

Scenario: *1 – Heathrow*

Configuration: *‘Mixed VR/Desktop’*

PF (VE): *...it was fantastic. Very real. When I first went into it [the VE], it was kinda weird as though I’d been transported somewhere. There was this background hum in the cockpit [engine noise], and the noise of the rain outside. Just a feeling of really being in the ‘plane.*

Researcher: *What about you? [Referring to the PNF]*

PNF (Real-World): *Well, yeah, I could see the rain and hear something. But I don’t think it was the same for me just looking at a screen. I always knew I was in the lab [the flight lab, where the experiment took place]. Like I was...detached.*

Participants reported physiological responses to the critical event in Scenario 2 (engine fire) and their reactions were observed to be more intense when immersed in the VE as opposed to the desktop simulator. These findings are synonymous to those found by Ref. [2]. The same conclusion can be made from the observations in this study, which is that VEs give rise to a sense of presence due to involvement and immersion factors, and instigate stress during critical events.

In the following narrative, the participants are discussing the critical event. The fire alarm for the engine fire – which is a bell and a red flashing ‘master caution’ light – immediately gets the attention of the participant immersed in the VE, whereas the real-world participant doesn’t initially associate the bell with the flight simulator:

Scenario: *2 – Sydney*

Configuration: *‘Mixed VR/Desktop’*

PF (VE): *That bell really scared me. It got my attention, although I wasn’t sure what to do about it at first.*

PNF (Real-World): *It was loud. I wondered if it was the building fire alarm.*

4.3.3 Communication

Generally, communication improved as participant pairs progressed through the scenario, becoming more familiar with both the procedures and each other. However, the ‘Mixed VR/Desktop’ configuration generated the least desirable outcomes in terms of communication. Some participants ($n = 5$, 21%) that used VR commented that initially it was ‘odd’ not being able to see the other participant or see their hands and follow which switches/buttons their co-pilot was accessing. However, some of these participants ($n = 3$, 13%) also added that they didn’t feel that it interfered with communication since it caused them to focus more on the semantics of communication, while others ($n = 2$, 8%) reported

that they initially felt disconnected from their co-pilot and required constant reassurance that the other participant was sat next to them.

In the following narrative, the participants are discussing their experience in the VE:

Scenario:	2 – Sydney
Configuration:	‘Dual VR’
PNF (VE):	<i>It was odd not being able to see my partner. Sometimes I forgot he was there.</i>
PF (VE):	<i>Yeah. Especially [not being able to see] the hands.</i>
PNF (VE):	<i>But it makes you think more about what you’re saying.</i>
PF (VE):	<i>Yeah. Like no hands, no nothing, just a voice. So I just have the voice. Yeah. And that’s it, yeah. I thought about the voice. The words.</i>

Two international participants in the ‘Dual VR’ configuration, whose first language was not English (and different to each other), reported that they had to concentrate hard on the words used in communication, both as the transmitter and the receiver. Like most participants, they were observed to adopt a closed communication loop to overcome the disconnect felt in the VE.

4.3.4 Non-verbal communication

Non-verbal communication includes deictic gestures and body language. Observed deictic gestures predominantly consisted of pointing, although some instances of giving and reaching were also seen in the ‘Dual Desktop’ configuration. Body language observed included body shifts, raised eyebrows, lip pursing, pouting and shoulder shrugs. It is beyond the scope of this paper to interpret the exact meaning of such body language, but in the context they were made, they can broadly be likened to frustration, disagreement or confusion. Pointing was predominantly observed in the ‘Dual Desktop’ configuration and consisted of participants indicating their own screen, the other participant’s screen or a hardcopy document (e.g. a chart, or scenario guide). While working through the checklists, some participant pairs ($n = 4$, 17%) tended to point and share the same big screen, even though both participants had their own screen, and only one of them had control of the view on the big screen.

Pointing was observed in the ‘Mixed VR/Desktop’ and the ‘Dual VR’ configurations, as demonstrated by the following narrative, where the participants are discussing taxiing onto the takeoff runway:

Scenario:	1 – Heathrow
Configuration:	‘Mixed VR/Desktop’
PF (VE):	<i>... I was pointing at this aircraft on final approach [an AI controlled aircraft]. I wasn’t sure if I should stop [taxiing the aircraft onto the runway].</i>
PNF (Real-World):	<i>But I couldn’t see it. You were pointing at the wall.</i>
PF (VE):	<i>I realised that in the end. But the aircraft went around [performed a ‘go-around’ i.e. aborted the landing]. It was ok in the end.</i>

Body language was also observed indicating frustration (shaking of the head, blowing, fist clenching). Such episodes sometimes included pointing out of frustration, even though one or both participants were immersed in the VE. The most observed pointing was while working through checklists in the VE, especially when accessing the B738 overhead panel.

In the following narrative, the participants are discussing the engine start and using the checklists:

Scenario:	<i>2 – Sydney</i>
Configuration:	<i>‘Dual VR’</i>
PNF (VE):	<i>During engine start, at the beginning, I know I kept pointing at the overhead panel and I know she couldn’t see me.</i>
PF (VE):	<i>And you hit me on the head.</i>
PNF (VE):	<i>I think I stopped after that. It was a little frustrating to begin with. Probably better if we were more familiar with a [Boeing] 737. It takes a bit of getting used to.</i>

4.3.5 Locating switches and checklist usage

A software design limitation caused four of the switches between the slave computer and the master computer not being mapped correctly. This issue was identified pre-experiment which allowed participants to be forewarned.

Many participants ($n = 13$, 54%) reported that they felt some level of frustration when one or both of them were in the VE and trying to describe the location of a button or switch. In response to the question, ‘What was hard/difficult/challenging (Why? Describe. . .)’, the majority of participants ($n = 15$, 63%) mentioned that it took some time to locate the appropriate switches, especially on the overhead panel. This might also be the case during a real type rating course where pilots would be unfamiliar with the cockpit layout. In this study, only three participants had any familiarity of the B738 (see Fig. 9) which was in a recreational capacity.

In the following narrative, one participant is discussing their difficulty with checklists:

Scenario:	<i>2 – Sydney</i>
Configuration:	<i>‘Mixed VR/Desktop’</i>
PF (VE):	<i>. . .even when he [the PNF] described the location to me a few times, I still had problems finding it [the switch] on the upper panel [overhead panel].</i>

4.3.6 Overall visual experience

Towards the end of the experiment, a PF and PNF pair reported observing different static aircraft (not the AI controlled aircraft). This occurred while they were parking the aircraft when the PF observed an occupied gate while the PNF observed the gate to be vacant. The participants (correctly) concluded that this was a software issue and proceeded to a different gate.

In response to the questions,

‘What was easy (Why? Describe. . .)’ or ‘Any other things you would like to talk about?’, a few participants who operated in the VE ($n = 4$, 17%) commented that it was easy to look around which provided a better overall ‘picture’ of the cockpit, the location of various switches and controls, and the outside environment. This is in comparison to participants operating in the real-world and using a

traditional desktop simulator in which the mouse (and sometimes the keyboard cursor keys, pre-defined numeric keypad keys or any combination thereof) had to be used to change the view. This type of two-dimensional representation of the environment is less similar to an actual B738 cockpit which does wrap around the crew to ensure controls are readily within reach.

5.0 Discussion

The study demonstrated that EBT and observation of the eight ICAO core pilot competencies (Table 1) can be accomplished in both a multi-crewed VRFS and a multi-crewed desktop simulator. The competencies, coupled with the post-experiment debrief, highlighted the extensive use of deictic gestures. This is a paradigm of Mehrabian's 7-38-55 rule [41], which states that 7% of meaning is communicated through spoken word, 38% through tone of voice and 55% through body language. Although the limitations of the rule have been subsequently exposed and it is not a measure of cockpit communication, it still serves to emphasise the relative insignificance of the actual words used in normal communication. It was observed that by creating a barrier to the normal communication channel (i.e. donning of the HMDs), non-verbal communication was replaced by verbal communication, the language of which became more precise during the experiment. This finding is synonymous with [42] who state that 'Collaborators used significantly less deictic gestures (in a VE) in favour of more unambiguous verbal references' (para. 1). Therefore, participants recognised the various barriers to communication and compensated by replacing the majority of the 55% of body language with spoken words.

The adoption of precise verbal communication is particularly beneficial since there is currently no commercially available technology that permits collaborators in a VRFS to usefully see each other in a VE, including the use of 'passthrough' functionality [26]. In addition, single-pilot commercial aircraft operations may be introduced in the early 2030s [43], which may ultimately involve physically separated users, and the ability of VRFSs to deliver such remote training would be advantageous.

The scenarios were designed to necessitate collaboration for their successful completion. The fact that all participant pairs managed to fully complete their assigned scenarios in all configurations suggests that the scenarios effectively facilitated collaboration. This implies that users: agreed on the shared goals; planned, allocated responsibility, and coordinated; shared context; communicated; and adapted and learnt [14].

The study has shown that behaviours related to the ICAO professional pilot training competencies can be invoked within a multi-crewed VRFS and a desktop simulator, indicating its applicability for EBT. The scenarios in this research replicated two typical simulator sessions for initial First Officer type rating training, resulting in varying degrees of competency demonstration across different areas. For example, despite none of the participants having any MCC experience flying a B738, the ICAO competency Leadership/Teamwork (i.e. collaboration) was successfully demonstrated. In a similar manner, task completion also demonstrated that the competencies of Problem Solving and Decision Making, Application of Procedures (i.e. checklists) and Workload Management can successfully be employed in a VE. For example, all participants successfully followed checklist procedures to manage the engine failure (fire) before takeoff (Sydney, Scenario 2) by extinguishing the fire and shutting down the engine.

When contrasting a VRFS with a desktop simulator, this study not only demonstrated the viability of integrating EBT into a multi-crewed VRFS but also highlighted that, once initial communication hurdles were overcome, VRFS fosters a stronger sense of presence. This, in turn, may amplify the benefits of immersive environments mentioned in the introduction, such as enhanced psychomotor skills, knowledge retention and spatial awareness.

Limitations of the study have been discussed in the main body of the text, and include restrictive checklists, inability to take notes and the unavailability of scenario guides and charts in the VE (Section 4.3), switch mapping between the slave computer and the master computer (Section 4.3.5), and the small sample size (Section 4.1).

6.0 Conclusions

The present research contributes to a growing body of evidence suggesting that VRFSs can be used to augment professional pilot training methods, such as computer-based training and full flight simulators. Furthermore, the research has demonstrated that a multi-crewed VRFS using a complex jet, with scenarios aligned to a typical initial First Officer airline training programme, can be used to develop and assess pilot core competencies. However, to ensure the efficacy of VRFSs for professional pilot training, much more research needs to be undertaken, especially in the areas of knowledge acquisition, development of procedures and flying skills, transfer of training and the application of human factors principles. Although this study has demonstrated that a VRFS can elicit the eight-pilot core-competencies, further research needs to be conducted on specific measurement of each ICAO competency during EBT.

Competing interests. The authors declare none.

References

- [1] Hays, R.T., Jacobs, J.W., Prince, C. and Salas, E. Flight simulator training effectiveness: A meta-analysis, *Mil. Psychol.*, 1992, **4**, (2), pp 63–74. https://doi.org/10.1207/s15327876mp0402_1
- [2] Cross, J.I., Boag-Hodgson, C.C. and Mavin, T.J. Measuring presence and situational awareness in a virtual reality flight simulator. *Aviat. Psychol. Appl. Human Fact.*, 2023, **13**, (2), pp 83–94. <https://doi.org/10.1027/2192-0923/a000250>
- [3] Tadeja, S.K., Seshadri, P. and Kristensson, P.O. AeroVR: An immersive visualisation system for aerospace design and digital twinning in virtual reality, *Aeronaut. J.*, 2020, **124**, (1280), pp 1615–1635. <https://doi.org/10.1017/aer.2020.49>
- [4] Airbus Virtual Procedure Trainer Offers An Innovative Way for Pilots to Learn Procedures Using Virtual Reality – Lufthansa Group becomes Launch Customer. [Online]. Available: <https://aircraft.airbus.com/en/newsroom/press-releases/2022-11-airbus-virtual-procedure-trainer-offers-an-innovative-way-for> (accessed 8 November 2023).
- [5] Cross, J.I., Boag-Hodgson, C.C., Ryley, T., Mavin, T.J. and Potter, L.E. Using extended reality in flight simulators: A literature review, *IEEE Trans. Visual. Comput. Graph.*, 2022, **29**, (9), pp 3961–3975. <https://doi.org/10.1109/TVCG.2022.3173921>
- [6] Abich, J., Parker, J., Murphy, J.S. and Eudy, M. A review of the evidence for training effectiveness with virtual reality technology, *Virtual Reality*, 2021, **25**, (4), 919–933. <https://doi.org/10.1007/s10055-020-00498-8>
- [7] Aslandere, T.I., Dreyer, D. and Pankratz, F. Virtual hand-button interaction in a generic virtual reality flight simulator. In *Proceedings of the 2015 IEEE Aerospace Conference*, 2015, pp 1–8. <https://doi.org/10.1109/AERO.2015.7118876>
- [8] Oberhauser, M., Dreyer, D., Braunstingl, R. and Koglbauer, I. Pilots' flight performance in a virtual reality flight simulator. In M. Schwarz and J. Harfmann (Eds.), *Proceedings of the 32nd Conference of the European Association of Aviation Psychology*, 2017, pp 322–335. Groningen, The Netherlands: European Association for Aviation Psychology.
- [9] Tran, T.H., Behrend, F., Funning, N. and Arango, A. Single pilot operations with AR-glasses using Microsoft HoloLens, In *Paper presented at the 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, London, pp 1–7. <https://doi.org/10.1109/DASC.2018.8569261>
- [10] Hamman, W.R. Line oriented flight training (LOFT): The intersection of technical and human factor crew resource management (CRM) team skills. In B.G. Kanki, R.L. Helmreich and J.M. Anca (Eds.), *Cockpit Resource Management*, 2010, pp 233–263. Burlington: Elsevier. <https://doi.org/10.1016/B978-0-12-374946-8.10008-1>
- [11] Helmreich, R.L. and Foushee, H.C. Why crew resource management? Empirical and theoretical bases of human factors training in aviation, In E.L. Wiener, B.G. Kanki, and R.L. Helmreich (Eds.), *Cockpit Resource Management*, 1993, pp 3–45. San Diego, CA: Academic Press.
- [12] Salas, E., Fowlkes, J.E., Stout, R.J., Milanovich, D.M. and Prince, C. Does CRM training improve teamwork skills in the cockpit?: Two evaluation studies, *Hum. Factors*, 1999, **41**, (2), pp 326–343. <https://doi.org/10.1518/001872099779591169>
- [13] ICAO. Manual of Evidence-Based Training (Doc 9995) (1st ed.). [Online]. Available: <https://store.icao.int/en/manual-of-evidence-based-training-doc-9995> (accessed 26 March 2022).
- [14] Terveen, L.G. Overview of human-computer collaboration, *Know.-Based Syst.*, 1995, **8**, (2–3), pp 67–81. [https://doi.org/10.1016/0950-7051\(95\)98369-H](https://doi.org/10.1016/0950-7051(95)98369-H)
- [15] Geremek, A., Greenlee, M. and Magnussen, S. Perception beyond gestalt: Introduction. In A. Geremek, M. Greenlee, and S. Magnussen (Eds.), *Perception Beyond Gestalt: Progress in Vision Research*, 2014, pp 1–5. London: Taylor & Francis Group.
- [16] Salas, E., Sims, D.E. and Burke, C.S. Is there a “Big Five” in teamwork?, *Small Group Res.*, 2005, **36**, (5), pp 555–599. <https://doi.org/10.1177/1046496405277134>
- [17] Salas, E., Dickinson, T.L., Converse, S.A. and Tannenbaum, S.I. Toward an understanding of team performance and training, In R.W. Swezey and E. Salas (Eds.), *Teams: Their Training and Performance*, 1992, pp 3–29. Ablex Publishing.
- [18] Crawford, E.R. and Lepine, J.A. A configural theory of team processes: Accounting for the structure of taskwork and teamwork, *Acad. Manag. Rev.*, 2014, **38**, (1), pp 32–48. <https://doi.org/10.5465/amr.2011.0206>

- [19] Landman, H.M., van Oorschot, P., van Paassen, M.M., Groen, E.L., Bronkhorst, A.W. and Mulder, M. Training pilots for unexpected events: A simulator study on the advantage of unpredictable and variable scenarios, *Hum. Factors*, 2018, **60**, (6), pp 793–805. <https://doi.org/10.1177/0018720818779928>
- [20] BEA. Final report on the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro – Paris. Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. Paris, France. [Online]. Available: <https://bea.aero/docs/2009/f-cp090601.en/pdf/f-cp090601.en.pdf> (accessed 12 January 2024).
- [21] Casner, S.M., Geven, R.W. and Williams, K.T. The effectiveness of airline pilot training for abnormal events, *Hum. Factors*, 2012, **55**, (3), pp 477–485. <https://doi.org/10.1177/0018720812466893>
- [22] Rosa, E., Dahlström, N., Knez, I., Ljung, R., Cameron, M. and Willander, J. Dynamic decision-making of airline pilots in low-fidelity simulation, *Theor. Issues Ergon. Sci.*, 2020, **22**, (1), pp 83–102. <https://doi.org/10.1080/1463922X.2020.1758830>
- [23] White, M.D., Padfield, G.D., Lu, L., Advani, S. and Potter, M. Review of flight simulation fidelity requirements to help reduce 'rotorcraft loss of control in-flight' accident rates, *CEAS Aeronaut. J.*, 2021, **12**, pp 701–721. <https://doi.org/10.1007/s13272-021-00542-6>
- [24] Cameron, M., Dahlström, N. and Kennedy, R. Competency scoring in simulation for experiential training (SET). Airside, 2022. [Online]. Available: <https://www.airside.aero/magazine/articles/competency-scoring-in-simulation-for-experiential-training> (accessed 12 January 2023).
- [25] Franks, P., Hay, S. and Mavin, T.J. Can competency-based training fly?: An overview of key issues for *ab initio* pilot training, *Int. J. Train. Res.*, 2015, **12**, (2), pp 132–147. <https://doi.org/10.1080/14480220.2014.11082036>
- [26] Cross, J.I. and Boag-Hodgson, C.C. A collaborative virtual reality flight simulator: Efficacy, challenges and potential, 2024. Submitted for publication.
- [27] Dapica, R., Hernández, A. and Peinado, F. Who trains the trainers? Gamification of flight instructor learning in evidence-based training scenarios, *Entertain. Comput.*, 2022, **43**. <https://doi.org/10.1016/j.entcom.2022.100510>
- [28] Dyer, J.L. Team research and team training: A state-of-the-art review. In F.A. Muckler, A.S. Neal and L. Strother (Eds.), *Human Factors Review*, 1985, pp 285–323. Santa Monica, CA: Human Factors Society. [https://doi.org/10.1016/0003-6870\(85\)90133-4](https://doi.org/10.1016/0003-6870(85)90133-4)
- [29] Harrison, D.A., Mohammed, S., McGrath, J.E., Florey, A.T. and Vanderstoep, S.W. Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality, *Person. Psychol.*, 2006, **56**, (3), pp 633–669. <https://doi.org/10.1111/j.1744-6570.2003.tb00753.x>
- [30] Marks, M.A., Mathieu, J.E. and Zaccaro, S.J. A temporally based framework and taxonomy of team processes, *Acad. Manag. Rev.*, 2001, **26**, (3), pp 356–376. <https://doi.org/10.2307/259182>
- [31] Valentine, M.A., Nembhard, I.M. and Edmondson, A.C. Measuring teamwork in health care settings: A review of survey instruments, *Med. Care*, 2015, **53**, (4), pp 16–30. <https://doi.org/10.1097/MLR.0b013e31827feef6>
- [32] Wright, M.C., Segall, N., Hobbs, G., Phillips-Bute, B., Maynard, L. and Taekman, J.M. Standardized assessment for evaluation of team skills, *Simul. Healthc.: J. Soc. Simul. Healthc.*, 2013, **8**, (5), pp 292–303. <https://doi.org/10.1097/SIH.0b013e318290a022>
- [33] CASA. Human factors principles and non-technical skills training and assessment for air transport operations, Advisory circular AC 119-12 v1.0. [Online]. Available: <https://www.casa.gov.au/human-factors-principles-and-non-technical-skills-training-and-assessment-air-transport-operations> (accessed 9 August 2023).
- [34] Flin, R. and Martin, L. Behavioral markers for crew resource management: A review of current practice, *Int. J. Aviat. Psychol.*, 2009, **11**, (1), pp 95–118. https://doi.org/10.1207/S15327108IJAP1101_6
- [35] ICAO. CBTA Instructional System Design. [Online]. Available: <https://www.icao.int/safety/OPS/OPS-Normal/Pages/CBTA-Instructional-system-design.aspx> (accessed 9 August 2023).
- [36] Braun, V. and Clarke, V. Using thematic analysis in psychology, *Qual. Res. Psychol.*, 2008, **3**, (2), pp 77–101. <https://doi.org/10.1191/1478088706qp0630a>
- [37] Bazeley, P. Analysing qualitative data: More than 'identifying themes', *Malaysian J. Qual. Res.*, 2009, **2**, (2), pp 6–23.
- [38] ICAO. *Evidence-Based Training Implementation Guide* (1st ed.). [Online]. Available: <https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/ebt-implementation-guide.pdf> (accessed August 2023).
- [39] CASA. Getting an aircraft type rating. [Online]. Accessed: 9 Aug. 2023. Available: <https://www.casa.gov.au/licences-and-certificates/pilots/ratings-reviews-and-endorsements/ratings-and-endorsements/getting-aircraft-type-rating#Differencetraining>
- [40] ICAO. International Standards and Recommended Practices. Annex 1 to the Convention on International Civil Aviation Personnel Licensing (12th ed.). [Online]. Available: https://www.icao.int/APAC/Meetings/2019%20COSCAP%20SEAEASA%20PEL/AN01_cons.2019_compressed.pdf (accessed 9 August 2023).
- [41] Mehrabian, A. *Silent Messages*. Wadsworth, 1971.
- [42] Müller, J., Radle, R. and Reiterer, H. Virtual objects as spatial cues in collaborative mixed reality environments: how they shape communication behavior and user task load. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems – CHI '16*, San Jose, California, USA, pp 1245–1249. <https://doi.org/10.1145/2858036.2858043>
- [43] Harris, D. Single-pilot airline operations: Designing the aircraft may be the easy part, *The Aeronautical Journal*, 2023, **127**, (1313), pp 1171–1191. <https://doi.org/10.1017/aer.2022.110>