Research Paper

INFORMATION AND SYSTEMS IN EDUCATION, Vol. 17, No. 1 2018
pp. 36–49

Improving the Flying Skills of Pilot Trainees by Reviewing and Reflecting on Line-of-Sight Data

Akira Haga*, Shinji Endo, Tosaku Shibata**, Shinji Morita**, Kazuyoshi Arai** and Yuji Tokiwa***

(Received 7 June 2017 and accepted in revised form 27 November 2017)

Abstract We developed a method of quantitatively assessing a pilot's flight-simulator performance to improve simulation-based education using flight-training devices (FTDs). As a quantitative index of performance, we focused on the pilot's line-of-sight while scanning flight instruments, and tried to characterize his/her performance level based on visual focus tendencies. Two experiments were conducted. The first was a comparative analysis of line-of-sight data recorded for flight instrument-rated (IR) pilots and non-instrument-rated (NIR) pilots in order to verify that line-of-sight data can be used as a quantitative assessment benchmark. The second was conducted to verify the effectiveness of reviewing and reflecting on a NIR pilot’s own recorded line-of-sight data – as well as that of a IR pilot – in the form of recorded images, for improving his/her flying skills. In each experiment, a steep turn, which is a common in-flight maneuver, was conducted using a flight-training device (FTD). Six cockpit instruments were chosen as the targets of line-of-sight measurements. The number of times that the pilot's eyes became fixated on each instrument was measured and the cumulative fixation time was calculated. In this study, we observed different visual focus tendencies for IR pilots and NIR pilots. Flying skills were found to markedly improve by reflecting on the pilot’s own line-of-sight, as well as that of an IR pilot using recorded images.

Keywords: skills training, line-of-sight data, training of aircraft pilots, flying skills, eye tracking

1. Introduction

In recent years, a shortage of professional pilots – associated with the mass retirement of the baby-boom generation and the worldwide increase in demand for air travel(1) – has presented increasingly serious problems for Japan’s domestic air-travel industry. Japan’s Ministry of Land, Infrastructure, Transport and Tourism has predicted that, by around the year 2030, some 400 new pilots will need to be hired each year, and the Ministry has announced plans to establish training courses for pilots at a total of 7 private universities, including Hosei University, making use of our aviation technology course(2).

The primary objective of these training courses is to improve the flying skills of pilot trainees in efficient ways, and one component of the training strategy is simulation-based education using flight-training devices (FTDs).

Simulation-based education shortens the time required for learning, reduces the costs by not flying actual airplanes, and can contribute significantly to improving flying skills(3). However, assessments of the flying skills of trainees during the training process have primarily been entrusted to subjective evaluations on the part of instructors, and the scarcity of objective methods of characterizing trainee performance is a problematic aspect of this approach. Unfortunately, the recording and organization of operational logs has been insufficient, creating a situation in which trainees are informed of the operational history of their flights only through the comments of instructors in briefing sessions after the flights. Under these circumstances, it is difficult for trainees to review their performance with adequate thoroughness.

Commercial airlines have implemented a program known as *Flight Operational Quality Assurance (FOQA)* to record and evaluate data on airplane flights and deliver feedback on the results to the flight crew. However, no such system currently exists for pilot training courses.

In today’s world, with services such as e-Portfolio enabling the collection and longitudinal assessment of large quantities of educational data, there is a need for quantitative, evidence-based assessments of flying performance for trainees in pilot training courses. It is also

*Research Center for Computing and Multimedia Studies, Hosei University, Japan
**Faculty of Science and Engineering, Hosei University, Japan

important that trainees be able to review their own flying performance based on objective benchmarks in order to identify and address areas needing improvement.

In this study, we focus on one specific candidate for use as a quantitative benchmark of the flying performance of pilot trainees: information on the pilot's line-of-sight (that is, where the pilot tends to direct his or her visual focus) while scanning flight instruments.

Commercial airline pilots are assumed to be capable of instrument-based flight (flight in which the pilot depends solely on instruments for measurements of the aircraft's attitude, altitude, position, and course) and to have the ability to obtain information on various flight statistics by visual scanning (reading numbers on instruments). Moreover, being able to make appropriate decisions based on that input is an important skill tied directly to flight quality.

In general, a scanning method known as *cross-check* – in which the pilot scans instruments from top to bottom and left to right with the horizon indicator in the center – is recommended for all pilots as a means of keeping abreast of accurate flight conditions. Figure 1 shows a flowchart diagramming the steps between the instant a discrepancy in a flight statistic arises (*the trigger*) and the time at which corrective action is taken. The high speed at which airplanes travel ensures that, if the discovery of a trigger is delayed, the discrepancy in question expands, increasing the time required to respond by taking corrective action. Thus, triggers must be detected quickly in order to restore proper flight conditions. This is a challenge that demands effective instrument-scanning techniques.

As the results of visual scanning are directly reflected by flight conditions and in-flight assessments, in this study we use line-of-sight information as an objective benchmark for assessing flying skills.

Measurements of pilot line-of-sight data have been used before in workload measurement studies[4,5], operational behavior analyses[6], and visual design of instrument monitors[7]. Nishi and Ohkubo[8,9] obtained line-of-sight information, muscle potentials and heart rates for Japan Air Self-Defense Force pilots during FTD operations and used these data for analyzing correlations with aircraft response data.

Among research studies conducted outside of Japan, the study of Yu, et al.[10] analyzing the visual-focus patterns of fighter pilots is relatively recent. In a joint study, Glaholt and the Toronto Research Centre[11] worked with Defence Research and Development Canada and other organizations to conduct an exhaustive survey of research over the past 60 years in order to establish correlations between pilot line-of-sight data and a variety of other factors including performance and workload. Dubois, et al.[12] conducted an experiment in which pilot trainees at military schools were given real-time audio or video instructions, such as "look outside", and reported visual-focus habits during this process.

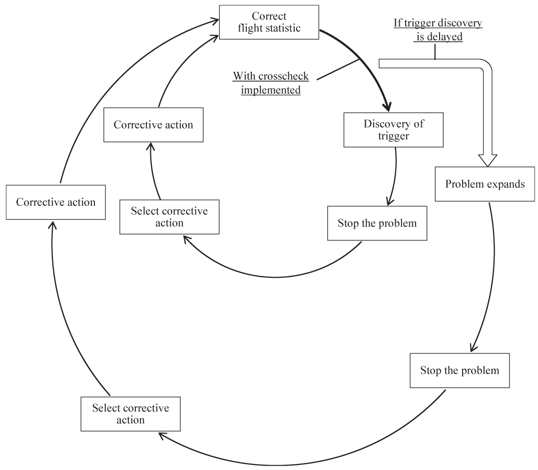

Figure 1. Flowchart of Pilot Corrective Action.

In the studies mentioned above, the primary test subjects were fighter pilots. However, the flight methods, flight tasks, and operational procedures related to visually-based military flights are completely different from those related to instrument-based commercial airline flights. In addition, these studies focused primarily on the observation of trends in line-of-sight information and did not mention its utilization in education and training.

With respect to commercial airline pilots, Haslbeck, et al.[13] carried out a comparative analysis of flying skills among captains of long-distance flights and first officers of short-distance flights based on line-of-sight data. In addition, Weibel, et al.[14] reported on trials of a system for displaying data on pilots' line-of-sight within the flight deck, together with data obtained from video images, audio recordings, digital pens, and other sources.

These studies also focused primarily on observation and did not mention its use in educational settings. This is important because, since there are no case studies that utilized line-of-sight data for educating inexperienced pilot trainees, the practical use of line-of-sight information in pilot education fields can be expected to be valuable.

Line-of-sight information has recently been utilized in the learning support field as well. In a typical example, Gofuku, and Hoshimoto [15] analyzed movement patterns in the visual focus of operators at a chemical plant and noted the differences between beginners and skilled operators.

Many other studies have focused on modeling the causal relationship between line-of-sight information and subject behavior by treating line-of-sight information as a parameter for understanding the cognitive process[16]. In contrast, the objective of this study is to improve the effectiveness of practical education by using line-of-sight information as an index for evaluating performance, with a focus on exploratory line-of-sight movements to detect the instantaneously changing external state of the flight caused by weather conditions such as air currents, as well as airplane position, engine output, and airspeed. Accordingly, herein we proposed and tested the efficacy of line-of-sight data as an objective, quantifiable benchmark for performance assessments in order to assist in improving the flying skills of trainees during simulation-based education in pilot-training courses for commercial aircraft.

We conducted two experiments. The objective of Experiment 1 was to test whether or not line-of-sight data are an effective assessment benchmark that measures a pilot's characteristic flying skill level. To accomplish this, we separated pilots into two groups of different skill levels: *instrument-rated* (IR) pilots, who have earned an instrument-based flight certificate, which is a well-established credential indicating competence in instrument-based flight, and *non-instrument-rated* (NIR) pilots, who have not earned that certificate. Using line-of-sight data as an assessment benchmark, we then performed a comparative analysis of flying skills during which we could determine whether or not our method detects characteristic visual-focus habits indicating a pilot's skill level, and assess whether or not line-of-sight data are an effective benchmark for evaluating pilot flying skills. In Experiment 2, we used the results of Experiment 1 regarding the visual-focus habits of pilots to offer instruction to trainees based on characterization of line-of-sight data, and then assessed the effectiveness of this educational technique in actual practice.

We believe that line-of-sight information can provide an effective indicator if trainees are able to utilize the assessment and feedback based on such data to improve their flying skills by correcting their visual-focus habits.

In this study, we focused on reflection as an educational strategy for modifying visual-focus habits. Reflection encourages self-awareness, which leads to learning improvement by compelling learners to review their own actions and their consequences. Setting aside all other possible elements for reflection other than line-of-sight information, we studied and measured actual performance improvements while considering "visual-focus habit" alone as a subject for reflection.

The two objectives of this instruction were (a) to acquire effective information about flight states by keeping an eye on appropriate meters at the proper timing, and (b) making suitable judgments and performing proper operations based on the above-mentioned information.

In the actual experimental phase, we recruited a group of NIR pilots to serve as test subjects, divided the full group into a test group and a control group, and then asked each pilot to carry out an identical sequence of flight tasks a total of two times. After the completion of the first flight-task sequence, pilots in the test group were given the educational opportunity to review recorded images of their own flight and to view recorded images of an exemplary IR pilot carrying out the identical flight-task sequence, together with images of that pilot's line-of-sight information. The control group was not offered this educational opportunity. We then measured the improvement in flying skills between the first and second flights and assessed the practical effectiveness of assessment and instruction based on line-of-sight data.

For the flight task assigned to pilots in all experiments, we chose a steep turn during flight work (which is a subject that offers practice in learning basic aircraft-operating techniques). Our reason for choosing this task was that it is the most typical flight task within the subjects of flight work. The specific task was to execute a continuous turn with the aircraft at an inclination angle of 45°. Pilots were required to carry out this procedure while stabilizing all three types of aircraft flight control surfaces.

2. Methods

2.1 Effectiveness of Line-of-Sight Data as an Assessment Benchmark

2.1.1 Overview

Experiment 1 to test the effectiveness of line-of-sight data as an assessment benchmark was conducted over a total of 7 days between July 22, 2016 and August 12, 2016. The participants were 5 pilots employed by commercial airlines and possessing instrument-based flight certificates (whom we denoted as the *IR group*) and 5 trainees not in possession of instrument-based flight certificates (the *NIR group*). All participants were male.

Before pilots operated the FTD, we provided both spoken and written overviews of our study and of instrument-based flight; we also discussed safeguards for private personal information and obtained the consent of all participants.

Participants were equipped with measuring instruments to record the position of their line-of-sight during operation of the Model-G58 certified FTD (Japan BTA Co., Ltd. Tokyo, Japan). Participants operated the controls from the left seat; when necessary, an instrument-based-flight-certificate-holding copilot occupied the right seat and received instructions regarding tasks such as raising and lowering flaps.

The FTD settings were configured as follows for our measurements (for the chart, see Figure 2).

Weather conditions: Wind 300°/8 knots
Range of visibility: 1,200 meters
Cloud height: 500 feet
Barometric pressure: 29.92 inches

Operating under these conditions, pilots performed a steep turn upon reaching an altitude of 5000 feet, during which we made line-of-sight measurements. We took the time at which the aircraft begins tilting in the direction of the initial turn as the starting point. The pilots directed the aircraft through a complete 180° rotation at

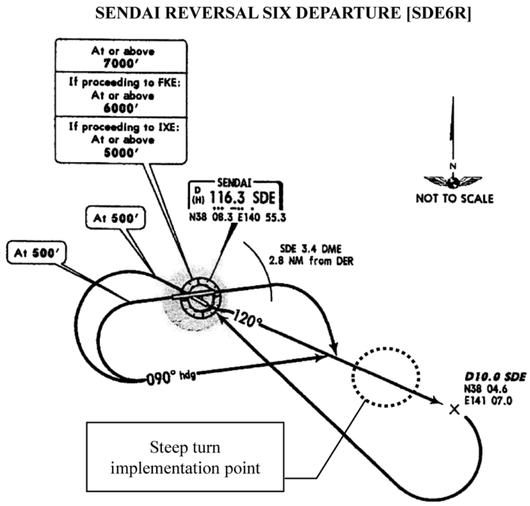

Figure 2. Chart Used during Measurements.

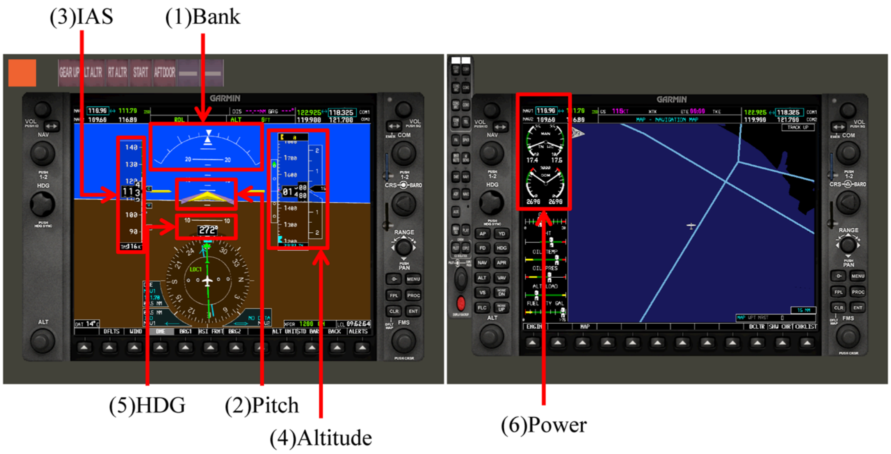

Figure 3. Important Measurement Instruments.

Table 1. Criteria for Assessing Flying Skill in Simulation.

Score	Criteria
5	The flight was stable. No deviation in the Speed, Pitch, Altitude, Bank and Heading were found.
4	The flight was stable, but deviation in Speed less than ±5 kt, in Pitch less than ±2.5, in Altitude less than ±50 ft, in Bank and Heading less than ±5° were found.
3	Some portions of the flight were somewhat unstable. Deviation in Speed between ±5 kt and ±10 kt, in Pitch between ±2.5° and ±5°, in Altitude between ±50 ft and ±100 ft, in Bank and Heading between ±5° and ±10° were found.
2	Some portions of the flight were unstable. Deviation in Speed between ±10 kt and ±20 kt, in Pitch between ±5° and ±7.5°, in Altitude between ±100ft and ±200ft, in Bank and Heading between ±10° and ±15° were found.
1	Some portions of the flight were unstable. Deviation in Speed greater than ±20 kt, in Pitch greater than ±7.5°, in Altitude greater than ±200 ft, in Bank and Heading greater than ±15° were found.

an orientation of 45°, and then rotated the aircraft 180° back in the opposite direction. The point at which the aircraft returns to its original course was taken as the end of the procedure.

For the instruments considered important for carrying out the flight task described above, we measured the *frequency* with which pilots directed their line-of-sight at the instrument and the total *duration* of time the line-of-sight remained on the instrument. We considered a pilot to have directed their visual focus on a target when their line-of-sight remains trained on that target for 0.1 s or longer.

The instrument screen (Figure 3) was captured as an operation log for evaluating the participant's skill. The target area was set just above each instrument, and involuntary eye movements and eye movements while switching to another target were ignored during the assessment. The visual focus was identified by instructors based on line-of-sight logs.

Assessments of the flying skills of trainees in simulations were made by instructors in Hosei University's Flight Operation Training program according to a 5-step assessment table (Table 1) based on existing practical testing guidelines for aircraft operators(17). In general, for commercial airlines, an assessment score of 3 or higher is considered passing for these practical tests.

2.1.2 Instruments Monitored

The measurement instruments considered necessary to complete the flight task are listed below; numbers refer to Figure 3, "Important Measurement Instruments."

(1) Bank
Shows the angle of left-right tilt about the aircraft's front-to-back axis when the control wheel is pointed to the left or right.

(2) Pitch
Shows the angle between the aircraft's neck and the horizontal – that is, the degree to which the aircraft is tilted in the horizontal direction – when the control wheel is moved forward or backward.

(3) Indicated Air Speed (IAS)
Shows the instantaneous speed of the aircraft with respect to the air in knots.

(4) Altitude
Shows the altitude of the aircraft. The units are feet (one foot = 0.305 meters). In this study, we included the GS indicator, which is an instrument that measures the rate at which the aircraft is climbing or descending, in our definition of altitude instruments. .

(5) Heading (HDG)
An angle display that indicates the angle of the direction in which the aircraft is proceeding, with magnetic north at 0°.

(6) Power
An instrument that indicates engine output. On the G58 model FTD, power is determined from the manifold pressure indicator.

2.1.3 Tools and Equipment

For line-of-sight data collection, we used the EMR-9 eye-mark recorder (NAC Image Technology, Inc. Tokyo, Japan). This instrument uses the corneal reflection method to detect the line-of-sight position by measuring the relative distance between the pupil center position and an image reflected from the cornea using near-infrared illumination. The instrument can collect data at a frequency of 60 Hz, allowing detection of line-of-site focusing events of 0.1 s or greater duration.

The FTD we used was equipped with a G1000 flight meter (Garmin Ltd. Olathe, KA.). The charts inserted in the instruments were printed for each flight.

For statistical processing we used SPSS Statistics 24 (IBM, Armonk, NY) software.

2.2 Effectiveness of Instruction Based on Line-of-Sight Data

Experiment 2 to test the effectiveness of instruction was conducted over a 9-day period from December 7, 2016 to December 17, 2016.

Based on the results of Experiment 1 discussed in Section 2.1, we first confirmed a certain problematic tendency, characteristic of NIR pilots, and we related it to the placement of visual focus during the execution of flight tasks. We also confirmed the more appropriate visual-focus habits of IR pilots during flight. Next, we conducted Experiment 2 to evaluate the practical effectiveness of instruction based on assessments of line-of-sight data in improving the flying skills of pilot trainees.

To accomplish this, we recruited a new pool of 14 NIR pilots not in possession of instrument-based-flight certificates (12 males and 2 females) and randomly divided this pool into two groups of 7, one as a test group and one as a control group. Flight tasks and measurements were carried out using the procedures described in Section 2.1.1.

Experimental measurements were made for the same flight-task sequence repeated two times. After the first flight, pilots in the test group were given an opportunity to review and reflect on line-of-sight data and flight logs from their own first flight performance, as well as an opportunity to review and reflect on line-of-sight data and flight logs for the same flight-task sequence executed by an exemplary IR pilot. In the process of reviewing and reflecting on flight records, trainees had an opportunity to identify —in recorded images— the points at which problems arose. Trainees were then asked to write down the causes and some proposed remedies for these problems, after which they repeated the flight-task sequence a second time. Next, from the recorded images, the trainees identified the differences between their visual scanning technique and that of the exemplary IR pilot. After confirming the trainees' review and reflection, two instructors provided advice about the trainees' scanning method, but not about their flight operations. The control group was not given an opportunity to review and reflect on the previous performance. The interval between the first and second trials was 2 weeks.

Based on the results of Section 2.1, we concluded that, in assessing flying skills during the execution of flight tasks, it would be desirable to obtain a more detailed characterization of line-of-sight transitions for NIR pilots during the various stages of execution of the flight-task sequence. To this end, we augmented our 5-stage evaluation rubric by creating a new assessment benchmark based on line-of-sight data, thereby defining clear targets for what pilots should be seeking as they scan instruments while completing flight tasks (Figure 4). We divided the "steep turn" into a 10-step process and considered that pilots directed their visual focus when their line-of-sight remained for 0.1 s or more on a target area, one or more times in each process. The assessment included both routine scanning criteria —to be in place at all times in flight— as well as recommended specific scanning targets to be used during the progress of a flight.

The frequency with which pilots failed to demonstrate desirable visual scanning practices was converted into a demerit penalty that reduced the pilots' flying skill assessment score. The instructors then calculated the number of times that acceptable scanning was not performed in each process using the demerit mark system. More specifically, the cases where the pilots did not direct their line-of-sight to the applicable instrument, or where they did not carry out an acceptable operation, was calculated as −2 points. Cases where the pilots directed their line-of-sight to the applicable instrument but did not perform an appropriate corrective operation were calculated as −1 point.

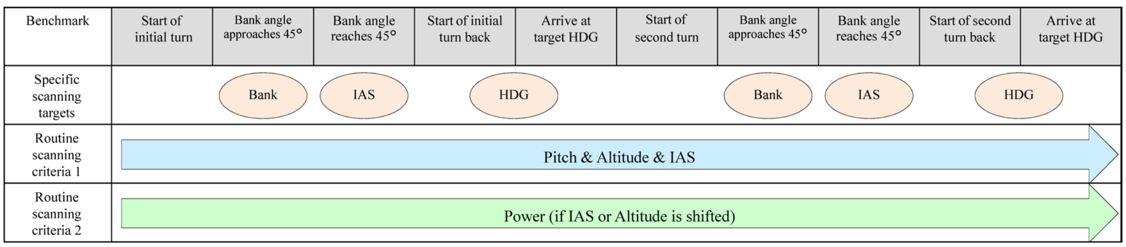

Figure 4. Evaluation Rubric Based on Line-of-Sight Data.

3. Results

3.1 Effectiveness of Line-of-Sight Data as an Assessment Benchmark

Within our pool of test subjects, the mean age of the IR pilots was 32.4 years. The mean age of the NIR pilots was 20.2 years.

Figures 5 and 6 show typical examples of transitions in visual focus with respect to the rubric of the IR and NIR groups. Compared to the IR group, the NIR group tended to focus more on the Bank indicator, with very little visual focus on other flight instruments. In contrast, the IR group showed a tendency to focus evenly on all six instruments.

To check the normality of each datum, we conducted Kolmogorov–Smirnov tests of the frequency and total duration of the visual focus displayed by pilots while carrying out flight tasks, as well as of the flying skill scores of the pilots in simulations and confirmed that the data obeyed a normal distribution.

Next, we analyzed the data using a parametric test

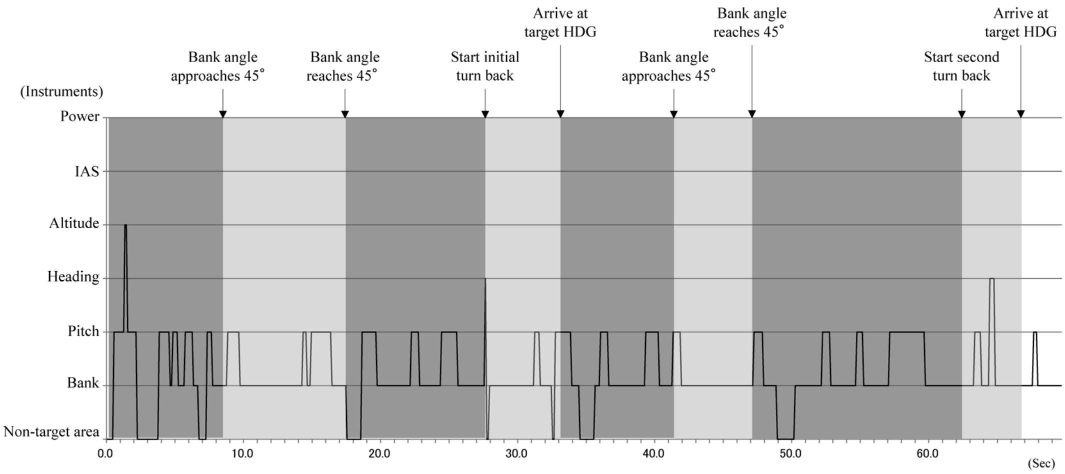

Figure 5. Visual Focus Transitions of a NIR Pilot.

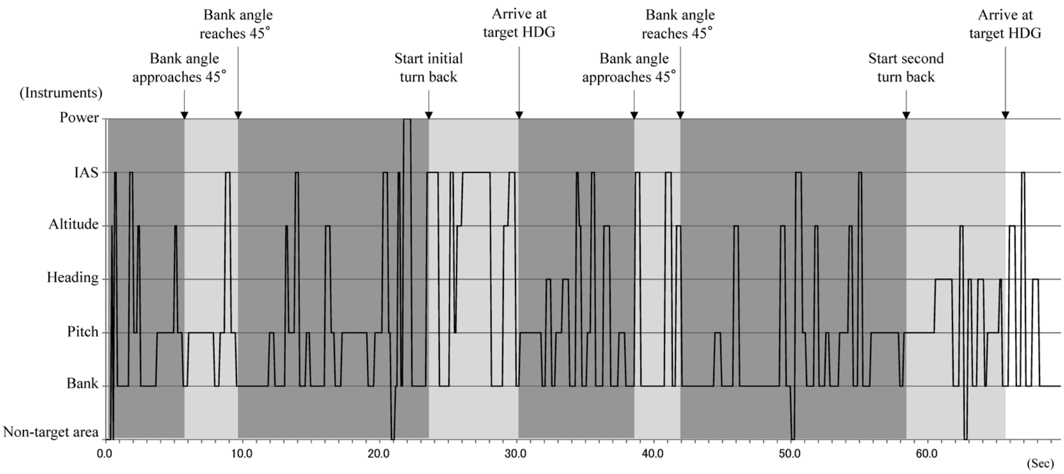

Figure 6. Visual Focus Transitions of an IR Pilot.

Table 2. Frequency and Duration of Visual Focus on Various Flight Instruments while Executing a Steep Turn.

Flight task	Instrument	Gaze	Instrument-rated pilot (IR) n=5	Non-instrument-rated pilot (NIR) n=5	t-test	
Steep Turn	Bank	Frequency	59.00 (SD±2.35)	89.80 (SD±4.98)	t (8) =−5.59	IR<NIR**
		Duration	18.73 (SD±3.13)	31.31 (SD±3.57)	t (8) =−2.65	IR<NIR*
	Pitch	Frequency	59.80 (SD±10.40)	54.60 (SD±4.39)	t (8) =0.46	N.S
		Duration	18.09 (SD±1.72)	16.68 (SD±1.01)	t (8) =−0.71	N.S
	Heading	Frequency	21.60 (SD±5.70)	11.40 (SD±3.67)	t (8) =−3.05	N.S
		Duration	7.02 (SD±1.51)	2.80 (SD±1.03)	t (8) =−2.31	NIR<IR*
	Altitude	Frequency	31.40 (SD±4.46)	11.40 (SD±5.70)	t (8) =2.77	NIR<IR*
		Duration	7.48 (SD±1.37)	2.78 (SD±1.35)	t (8) =−2.45	NIR<IR*
	IAS	Frequency	21.40 (SD±2.93)	3.60 (SD±1.36)	t (8) =5.51	NIR<IR**
		Duration	5.42 (SD±0.84)	1.21 (SD±0.34)	t (5.29) =−4.65	NIR<IR**
	Power	Frequency	3.40 (SD±1.50)	1.20 (SD±0.58)	t (5.11) =1.36	N.S
		Duration	0.83 (SD±0.42)	0.25 (SD±0.13)	t (8) =−1.307	N.S

※Mean notation. Significance level **=0.01 *=0.05

(t-test). Table 2 shows a comparison of frequency and total duration of visual focus on flight instruments, while Table 3 shows a comparison of flying skill assessment scores. In these tables, we see first that the frequency and duration of visual focus on the Bank indicator were greater, with statistical significance, for NIR pilots than for IR pilots. In contrast, both the frequency and duration of visual focus on the Altitude and IAS indicators were greater for IR pilots than for NIR pilots. For the Heading indicator, the duration of visual focus was greater for IR pilots than for NIR pilots. For the other instruments, no statistically significant differences were observed between the groups. We can also see that flying skill assessment scores were higher, with statistical significance for the IR group than for the NIR group.

3.2 Effectiveness of Instruction Based on Line-of-Sight Data

Among the NIR pilots in our study, the mean age of the test group was 20.2 years, while the mean age of the control group was 20.1 years.

Figures 7 and 8 show examples of transitions in visual focus of a typical subject in the test group before and after reflection. Here, we see that, compared to the data before reflection, the issue of excessive visual focus on the Bank indicator decreased after reflection. Moreover, the frequency of visual focus also increased for flight instruments that were looked at less frequently before reflection, such as the Heading and Altitude indicators.

With respect to the rubric, we can see that the pilot

Table 3. Flying Skill Assessments for IR and NIR Groups.

No.	Group	Score
1	IR	5
2	IR	5
3	IR	5
4	IR	3
5	IR	3
6	NIR	1
7	NIR	1
8	NIR	2
9	NIR	1
10	NIR	1
t-test	t (8) =5.67	NIR<IR**

Significance level **=0.01 *=0.05

had started visually focusing on the flight instrument that requires visual focus at each step such as, for example, directing his/her visual focus on the IAS at the point where the bank angle reaches 45°, or on the HDG at the time of turning back.

Our statistical data for scores within the 5-step assessment of flying skills, the number of demerits, and the frequency and duration of visual focus on various measurement instruments were then subjected to Kolmogorov–Smirnov tests to check normality and were found to obey a non-normal distribution. Next, we analyzed the data using non-parametric methods. In this stage, the differences in flying skill scores and the number of demerits between the two groups were compared. Since the experiment was conducted between test subjects, the Mann–Whitney U test was used, and the results are shown in Table 4.

Next, we compared the differences in visual focus frequency and duration between the first and second trials for the test and control groups (Table 5). Here, since data for the same subjects were compared, the Wilcoxon signed-rank test was used.

In assessments of flying skills, scores for pilots in the test group improved to higher values than was observed for the control group. Within the test group, two students were observed to achieve a score of 3, which is considered passing by commercial airlines. The number of demerits received tended to be lower for the test group compared to the control group. For test group members, both the frequency and duration of visual focus on the Bank indicator were reduced on the second flight compared to the first flight. Similarly, for the test group, the frequency and duration of visual focus on the Heading indicator were greater on the second flight than the first flight. The duration of visual focus on the Altitude instrument also increased from the first to the second trial for the test group. For the control group, no statistically significant differences in visual focus frequency or duration were observed between the two trials for any instrument.

Tables 6 and 7 show the sheets filled in as refer-

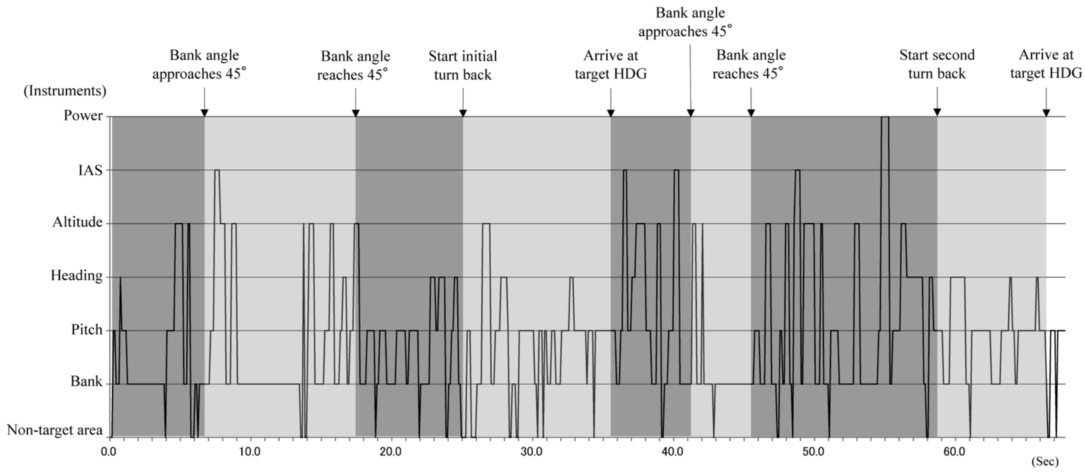

Figure 7. Transitions in Visual Focus before Reflection.

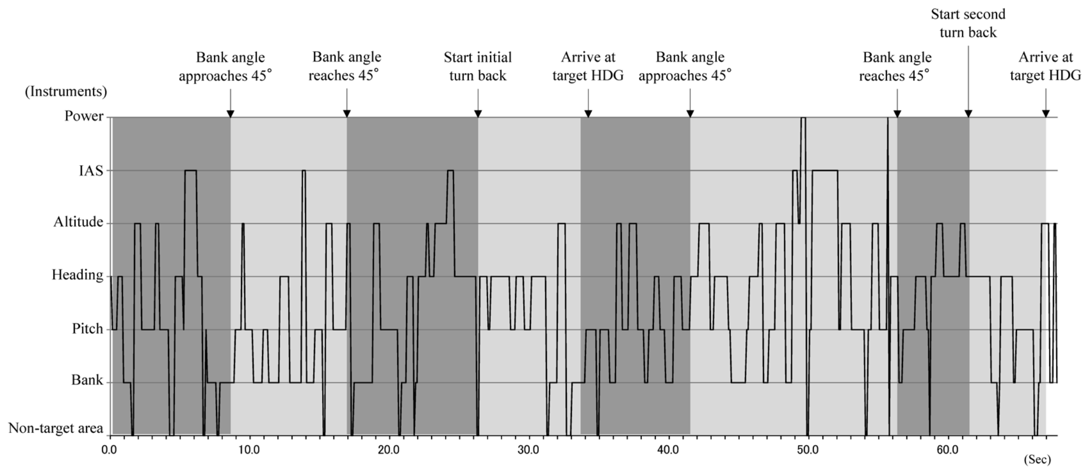

Figure 8. Transitions in Visual Focus after Reflection.

Table 4. Scores and Demerit Frequencies for Test and Control Groups.

Group	No.	1stScore	2ndScore	Difference of Score	1st Demerit Point	2nd Demerit Point	Difference of Demerit Point
Test	1	1	3	2	25	4	−21
	2	1	3	2	22	11	−11
	3	1	2	1	11	7	−4
	4	1	2	1	16	5	−11
	5	2	2	0	15	8	−7
	6	1	2	1	15	14	−1
	7	1	2	1	13	10	−3
Control	8	1	1	0	16	14	−2
	9	1	1	0	12	14	2
	10	2	2	0	2	12	10
	11	1	1	0	21	20	−1
	12	1	1	0	31	30	−1
	13	1	1	0	12	15	3
	14	1	2	1	13	10	−3
Mann–Whitney *U* test		Test < Control**			Control < Test**		

Significance level **=0.01 *=0.05

Table 5. Results Regarding Visual Focus on Instruments for Test and Control Groups.

Instrument	Gaze	Test (*n*=7)		Wilcoxon signed-rank test	Control (*n*=7)		Wilcoxon signed-rank test
		1st	2nd		1st	2nd	
Bank	Frequency	84.00	60.00	N.S	64.00	77.00	N.S
	Duration	28.26	21.29	1st>2nd**	32.25	25.51	N.S
Pitch	Frequency	47.00	55.00	N.S	49.00	57.00	N.S
	Duration	18.63	17.37	N.S	20.06	19.48	N.S
Heading	Frequency	17.00	27.00	1st<2nd**	32.00	28.00	N.S
	Duration	7.10	11.61	N.S	13.09	11.71	N.S
Altitude	Frequency	15.00	32.00	N.S	24.00	26.00	N.S
	Duration	4.42	11.87	1st<2nd**	7.93	7.78	N.S
IAS	Frequency	3.00	8.00	N.S	1.00	4.00	N.S
	Duration	1.46	1.52	N.S	0.37	1.15	N.S
Power	Frequency	1.00	3.00	N.S	0.00	0.00	N.S
	Duration	0.14	0.42	N.S	0.00	0.00	N.S

※Median notation. Significance level **=0.01 *=0.05

ence materials for reflection by the subject in Figures 7 and 8.

In Table 6, where the subject himself reviewed the first flight, we see that the subject was aware of his tendency to direct excessive visual focus on the Bank indicator. The subject was also aware of his insufficient visual focus on Altitude from the time the bank angle reaches 45° until initial turn back, and the need to visually focus on other instruments when turning back. This indicated he was able to gain a certain degree of improved awareness just by reviewing his own video recording.

In Table 7, where the subject compared himself with IR pilots, we see that the subject had become aware of the fact that his own eye movements were scattered compared to the systematic scanning of the IR pilots.

In the advice on directing visual focus provided by instructors in Table 8, we can see that the subject had become aware of the need to steadily observe the Pitch indicator, which is an important element that should always be kept in mind when scanning. Additionally, the instructor's advice touched on points that the student himself had already noted, as shown in Table 6.

Table 6. Self-Review of First Flight.

Step	Problematic visual focus point	Issue	Solution
After Bank angle reaches 45° until Start of initial turn back	Bank	Visually focused on Bank indicator and did not look at Altitude indicator	Look at Altitude indicator more to keep it steady
After Bank angle reaches 45°until Start of second turn back	Bank	Forced to change the Pitch indicator excessively because of my preoccupation with the Bank indicator	Execute each action precisely
After Start second turn back until Arrive at target HDG	Bank	Missed the Altitude indicator because I wasn't looking at the Pitch indicator	Scan all the other instruments when returning to the Bank indicator

4. Discussion

4.1 Effectiveness of Line-of-Sight Data as an Assessment Benchmark

The test results in Table 2 corroborated the tendencies in visual focus transitions shown in Figures 5 and 6.

During the execution of the flight task, NIR pilots directed their visual focus at the Bank indicator more frequently and for a greater total duration than IR pilots. We attribute this to the likelihood that, compared to IR pilots, NIR pilots must concentrate more intensely on maintaining a 45° bank angle while turning.

The frequency and duration of visual focus for the Altitude and IAS indicators were higher for IR pilots than for NIR pilots. This indicates that when NIR pilots are forced to concentrate most of their focus on maintaining the bank angle, they inevitably pay less attention to maintaining the aircraft's altitude and speed.

Compared to NIR pilots, IR pilots exhibited a higher duration of visual focus on the Heading indicator. However, no meaningful discrepancy between the groups was observed for the frequency of visual focus on that instrument. This suggests that, whereas NIR pilots look somewhat absent-mindedly at the Heading instrument when reading the number it displays, IR pilots recognize the significance of this number and correspondingly devote significant time to observing it.

Based on these observations, we can see that our method has identified differences in the visual-focus habits of IR and NIR pilots. Specifically, whereas IR pilots scan many instruments with uniform attention and focus, NIR pilots direct the majority of their attentions to maintaining the aircraft's bank angle and consequently have little focus remaining to direct to the task of maintaining other flight indicators such as altitude or speed.

The fact that our method detects both differences in skill level and specific areas of problematic behavior in visual-focus habits confirms its effectiveness as a quantitative assessment benchmark.

Table 7. Line-of-Sight Movement in Comparison with IR Pilots.

Compare your own line-of-sight movement with those of IR pilots
My line of sight was scattered and deviated from the instruments. I was not able to look at the instruments with composure. I will scan more thoroughly and with more presence of mind. I did not look at the IAS sufficiently.

4.2 Effectiveness of Instruction Based on Line-of-Sight Data

Based on the results in Table 4, the fact that the group of test subjects who reviewed and reflected on their performance both (a) improved their assessment scores, and (b) decreased their numbers of demerits suggests that this method of instruction indeed contributes to improving the flying skills of trainees. The decrease in the demerit points suggests trainees learned how to scan and what they should scan, and modified their visual-focus habits through the comparisons with an exemplary IR pilot.

The test results in Table 5 corroborated the differences in visual focus transitions found in Figures 7 and 8.

We see that the tendency to be consumed by the task of controlling the aircraft's bank angle, and thus to direct excessive visual focus to the Bank indicator — which we noted in Section 4.1 was a problematic behavior that NIR pilots tended to display— was reduced among the trainees who reviewed and reflected on their performance.

The visual focus time directed toward the Altitude and Heading instruments also increased, suggesting that reviewing and reflecting on recorded images caused pilots to focus consciously on executing the crosscheck approach.

No differences were observed for IAS, thus indicating that, compared to the differences between the IR and NIR groups, the differences among the NIR group

Table 8. Advice on Line-of-Sight Movement from Instructors.

Advice from instructors			Solution considering the advice
Step	Important visual focus point	Issue	
After Start of initial turn until Bank angle approaches 45°	Pitch, IAS, Bank	Insufficient attention to Pitch and IAS indicators, excessive attention on Bank indicator	Incline bank more slowly and maintain scan of other instruments
After Bank angle approaches 45° until Start of initial turn back	Pitch, Altitude, IAS, Bank	Insufficient attention paid to Pitch, Altitude, and IAS indicators, excessive attention on Bank indicator	Scan more precisely because I focused needlessly on Bank indicator
After Start of initial turn back until Arrive at target HDG	IAS	Generally good, but insufficient attention paid to IAS indicator	Control the throttle
After Arrive at target HDG until Bank angle approaches 45°	Altitude	Insufficient attention paid to Altitude indicator while returning	Look at Pitch indicator
After Bank angle reaches 45° until Start of second turn back	IAS	Generally good, but somewhat insufficient attention paid to IAS indicator	Control the throttle
General Remarks	Pitch, Altitude, IAS, Bank	Altitude and IAS are unsteady because you did not look at the Pitch indicator steadily	Look at Pitch indicator steadily

were not so great as to yield clearly significant performance gaps for this benchmark.

The results in Figure 8 and Table 5 show that corrections required to achieve the visual focus targets, presented as solutions by the trainee in Tables 6, 7, and 8, had been carried out. Between the time to arrive at the target heading and when the bank angle approaches 45°, the solution to "look at pitch" given by the trainee for the issue of "insufficient visual focus on altitude", suggests that the trainee had become aware that a steady pitch would result in more time to visually focus on altitude.

5. Conclusion

In this study, with the goal of improving the educational effectiveness of simulation-based training, we performed quantitative assessments of flying skills using line-of-sight data as an assessment benchmark.

First, to test the effectiveness of line-of-sight data as a quantitative assessment benchmark, we made experimental measurements of trends in line-of-sight information during flight work for two groups of pilots with differing skill levels —one group consisting of pilots possessing instrument-based flight certificates (IR pilots) and another group not possessing these certificates (NIR pilots)— and compared the results.

Our results for the frequency of line-of-sight focusing and for total duration of visual focus indicated that IR pilots used the crosscheck approach, and thus tended to direct their visual focus toward all instruments uniformly. In contrast, NIR pilots were consumed by the task of maintaining a constant angle on the Bank indicator during a steep turn —which is the primary task of the flight sequence— and consequently paid less attention to other instruments, which is a problem that our method revealed in full quantitative detail. This demonstrates that line-of-sight data offer insights into the visual-focus habits of pilots of varying skill levels, thereby demonstrating its effectiveness as an assessment benchmark.

Next, to test the practical effectiveness of using line-of-sight information to improve the flying skills of trainees, we offered some NIR pilots an opportunity to review and reflect on their flying performance, as indicated by the line-of-sight data, and then compared flying skill scores and visual-focus habits before and after this review both for participants who did and who did not review and reflect on their performance. Our experimental results showed that the group of NIR pilots who reviewed and reflected on their performance exhibited improvements in visual-focus habits— an aspect of performance in which significant problems were present in the first experimental flight trial— and improved their flying skill scores, thereby reducing the number of demerits incurred. This shows that assessment and instruction based on line-of-sight data improves the flying skills of trainees, thus demonstrating the efficacy of our proposed method.

Furthermore, the results suggested that improvement in flying skills can be obtained by instructions limited to line-of-sight movement alone.

In future work, we hope to design an automated line-of-sight pattern-learning system using virtual-reality techniques in which our rubric for line-of-sight data is taken as a conditional formula, and thus contribute to continued improvements in the efficacy of pilot training.

Trademark

The names of systems and products used in this paper are, unless otherwise specified, trademarks or registered trademarks of the corporation in question.

Acknowledgement
This study was supported by the Japan Society for the Promotion of Science (JSPS) research grant #15K00494

References

(1) International Civil Aviation Organization: Global and Regional 20-year Forecasts: Pilots, Maintenance Personnel, Air Traffic Controllers, International Civil Aviation Organization, Montreal (2011).
(2) Japan Civil Aviation Bureau: "Current Status and Issues of the Pilots, Maintenance Personnel and Air Traffic Controllers", https://www.mlit.go.jp/common/001019364.pdf, (accessed 2017.9.20) (in Japanese).
(3) Hays, R. T., Jacobs, J. W., Prince, C. et al.: "Flight Simulator Training Effectiveness: A Meta-Analysis", Military Psychology, Vol. 4, No. 2, pp. 63–74 (1992).
(4) Katoh, Z., Kadoo, A. and Nishi, S.: "Heart Rate Difference between Actual and Simulated Flight", Japanese J. of Ergonomics, Vol. 32, No. 3, pp. 123–129 (1996).
(5) Tsuruhara, A., Arake, M., Ogawa, T. et al.: "A Study on Pilot Mental Workload under a Flight Task in a Simulator: Assessment by ASSR". Japanese J. of Ergonomics, Vol. 51, Suppl., pp. 342–343 (2015).
(6) Nomiyama, T.: "Measurement of Situation Awareness in Simulated Flight: Analysis of Eye Tracking Data and Heart Rate Data", Japan Ergonomics Society 51th Conference, pp. 160–161 (2010).
(7) Sarter, N. B., Mumaw, R. J. and Wickens, C. D.: "Pilots' Monitoring Strategies and Performance on Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data", J. of the Human Factors and Ergonomics Society, Vol. 49, No. 3, pp. 347–357 (2007).
(8) Nishi, S. and Ohkubo, T.: "A Study of Pilot's Visual Scanning Behavior on a Visual Flight Simulator", The Japanese J. of Ergonomics, Vol. 31, No. 3, pp. 225–233 (1995).
(9) Nishi, S. and Ohkubo, T.: "A Study on Pilot Visual Scanning, Control Action and Flight Performance during Simulator Flights", J. of Japan Industrial Management Association, Vol. 47, No. 6, pp. 327–334 (1997).
(10) Yu, C-S., Wang, E. M-Y., Li, W-C. et al.: "Pilots' Visual Scan Patterns and Attention Distribution During the Pursuit of a Dynamic Target", Aerospace Medicine and Human Performance, Vol. 87, No. 1, pp. 40–47 (2016).
(11) Glaholt, M. G. and the Toronto Research Centre: "Eye Tracking in the Cockpit: A Review of the Relationships between Eye Movements and the Aviator's Cognitive State", Defense Research and Development Canada Scientific Report DRDC-RDDC-2014-R153 (2014).
(12) Dubois E., Blättler C., Camachon C. et al.: "Eye Movements Data Processing for Ab Initio Military Pilot Training", Intelligent Decision Technologies 39 of the series Smart Innovation, Systems and Technologies, pp.125–135 (2015).
(13) Haslbeck, A., Schubert, E., Gontar, P. et al.: "The Relationship between Pilots' Manual Flying Skills and Their Visual Behavior: A Flight Simulator Study Using Eye Tracking. A Flight Simulator Study Using Eye Tracking", in Advances in Human Aspects of Aviation (Advances in Human Factors and Ergonomics), eds. Laundry, S., Salvendy, G. and Karwowski, W., pp. 561–568 (2012).
(14) Weibel, N., Fouse, A., Emmenegger, C. et al.: "Let's Look at the Cockpit: Exploring Mobile Eye-tracking for Observational Research on the Flight Deck", Proc. of the Symposium on Eye Tracking Research and Applications, pp.107–114 (2012).
(15) Gofuku, A. and Hoshimoto, T.: "A Technique to Analyze Transition Patterns of Eye Fixation Points for Extracting Operation Skill and its Application to a Model Plant", Transactions of Human Interface Society, Vol. 14, No. 2, pp. 159–166 (2012).
(16) Tsubokura, A., Imai, H., Nishiki, T. et al.: "The Modelling of Cognition-Motion Process Using Eye Mark Tracing Date", Transactions of Japanese Society for Information and Systems in Education, Vol. 14, No. 5, pp. 191–200 (1998).
(17) Japan, Civil Aviation Bureau, Engineering Department, Personnel Licensing Division: "Practical Testing Guidelines for Aircraft Operators", Hobun Shorin Co., Ltd., Tokyo (2008). (in Japanese)

Akira Haga completed his Ph.D. at Waseda University, Graduate School of Human Sciences in 2012. After working as an assistant professor, since 2016 he has worked as a lecturer at the Research Center for Computing and Multimedia Studies at Hosei University. He was also a visiting research scholar at the Center for Higher Education Studies at Waseda University.

Shinji Endo graduated from the Department of Mathematics, Faculty of Science, at Tokyo Metropolitan University in 1970 and began working at Japan Airlines the same year. He worked as a B747–400 captain and ground instructor. In 2008, he began working as a professor in the Course of Aviation Technology in the Faculty of Science and Engineering at Hosei University. Since 2016, he has worked as a lecturer at the same university. He is also a member of the Japan Society for Aeronautical and Space Sciences, and serves on the technical committee at the New Energy and Industrial Technology Development Organization.

Tosaku Shibata graduated from the University of Electro-Communications in 1973 and began working at Japan Airlines the same year. He worked as a DC-10 captain and a flight engineer for B747s. In 2003, he completed his Master's degree in the Graduate School of Psychology at Tokyo Seitoku University. Since 2008, he has worked as a professor in the Course of Aviation Technology in the Faculty of Science and Engineering at Hosei University.

Shinji Morita graduated from the Civil Aviation College in 1975 and began working the same year at Japan Airlines. He has worked as a B747 captain, a deputy manager in the department of the cockpit crew of B747–400s and Mission Director Vice President of Airport Operations. He has worked as a Senior Director at Napa Flight Crew Training Center in CA, USA. After working as a deputy general manager for Aero Asahi Corporation, in the Business Jet Division, from 2016, he worked as a professor in the Course of Aviation Technology in the Faculty of Science and Engineering at Hosei University. He has served as Department Chair since 2017.

Kazuyoshi Arai received his Ph.D. from the Graduate School of Science and Engineering at Tokyo Institute of Technology, in 1988 and began working the same year in the Faculty of Engineering at Yokohama National University as a research associate. From 1995, after working as a research associate in the Department of Industrial Engineering and Management at Nihon University, he worked as a senior lecturer and then associate professor in the Faculty of Science and Engineering at Hosei University. Since 2002, he has worked as a professor at the same university.

Yuji Tokiwa completed his Master's degree in 1978 in the Faculty of Engineering at Keio University and began working at Ishikawajima-Harima Heavy Industries (now IHI Corporation) the same year. In 1984, he began working at IBM Japan. Since 2005, he has worked as a professor at the Research Center for Computing and Multimedia Studies at Hosei University, engaged in the development of systems for university education, research, administration, and management. He is a member of the Information Processing Society of Japan and the IEEE.