Master's Thesis

Quantitative analysis on the effectiveness of CRM training in aviation industry

Takumi Okada

Supervisor Lam Chi Yung

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Knowledge Science)

March 2024

Acknowledgements

I would like to appreciate many people who provided guidance and encouragement in the preparation of this paper.

Especially, I am indebted to Professor Lam Chi Yung, my supervisor. Thank you for giving me so many feedback and advice. In addition to them, I also need to say thank you for buying expensive equipment for my experiment. It would not be possible to complete this thesis without your financial support. I sincerely appreciate overall support you gave me.

I would like to appreciate Professor Kazushi Nishimoto and Professor Hideomi Gokon too. Both of the professors gave me advice whenever I sent e-mails. The advice was critical and enabled me to think more deeply about my theme. In addition to the appreciation. I am especially indebted to Gokon sensei who was in charge of supervisor on my sub-research. The knowledge and insight obtained in the discussion with Gokon sensei was crucial to this main research too.

Finally, I appreciate all participants of this research. Even though many participants were busy on writing on their own thesis, they did not hesitate to help me. I truly appreciate about the cooperation.

Thank you for the people involved in this research.

Abstract

This study is conducted to quantitatively evaluate the effectiveness of training in CRM (Crew Resource Management). At first, CRM is a management framework to maximize the use of human resources by training pilots, mainly in the aviation industry. The following two benefits are expected from this training. One is the improvement of communication skills. The second is to improve management techniques based on current and future situational awareness. Thus, the essence of CRM is a smooth utilization of the knowledge and experience among crew members in their work.

However, the recent environment surrounding CRM is problematic. Because when it comes to evaluation methods on CRM training, current aviation industry cannot help but rely on the subjective experience of the evaluator. This is because CRM deals with vague concepts such as communication and management. Thus, it makes difficult to evaluate CRM effectiveness quantitatively. Therefore, different evaluators may give different ratings to the same training. This situation is undesirable. According to this situation, the objective of this study is to develop and verify an objective evaluation method that outputs the same results regardless of who does the evaluation.

As a novelty of evaluation method, this study proposes an evaluation method focusing on aircraft trajectories. Previous evaluation studies in CRM training have focused on changes in pilot behavior. This is not surprising from the perspective that CRM drives change in pilots. However, since the goal of conducting CRM training is ultimately to improve operational performance. Therefore, it is worth to be considered.

Based on the above ideas, this study proposes a five-axis evaluation framework that adds a quantitative evaluation axis for communication in addition to the three evaluation axes for aircraft trajectory. Then, comparative experiment will be conducted to verify the effectiveness of this evaluation method. Specifically, the CRM-trained experimental group and control groups will be given a task using a flight simulator and the results will be compared. In this experiment, an artificial authority gradient shall be formed by giving false information to each subject to reproduce the actual cockpit situation. This will be created as an impediment to communication by giving them the role of a supervisor and his subordinate.

The results of the experiment, conducted in pairs, were generally favorable for the validity of this evaluation method. Specifically, the results reflect the effectiveness of CRM training with respect to course-focused evaluations. In the other words, the effectiveness of CRM training, such as task management and planning, was reflected.

However, contradictory results were obtained for some of the evaluation items related to communication. The control group, which had not received CRM training, recorded better performance. In conclusion, the results of the framework verification were generally appropriate, but some of the results left room for improvement.

Table of contents

Table of Figures

Table of tables

Chapter.1 - Introduction

1.1 - Research background

In recent years, Aircraft operations have been very safe. Many people tend to think that airplanes are dangerous vehicles whenever airplane accident happens. Because it is broadcasted widely on the news. However, this is not true. For example, the probability of an accident during flight is 1 in 1.2 million flights, and the probability of it being fatal is 1 in 11 million. On the other hand, the probability of death in a car accident is more than 200,000 times higher [1]. This is because of the development of technology used in aircraft, legislation, and training systems. This study focuses on the most developed one in these advanced systems. It is a training system for pilots. This is because the primary cause of aircraft accidents is neither of aircraft malfunction nor terrorist attack. The most probable cause of aviation accidents is pilot error. Therefore, a large number of educational methods to reduce errors have been studied from various aspects. One element of these research is worth noting. It's called CRM (Crew Resource Management).

A brief overview of CRM will be given. This is a concept to maximize aircraft safety by exploiting the knowledge and experience of all crew members. Modern aircraft are larger than before to carry more passengers. Thus, the number of crew members required for operation has also increased. However, the increase in headcount was not proportional to the performance of the operation. In other words, the number of mistakes did not decrease as the number of crews increased.

Case of accident	Catalyst	Probable cause
Eastern Air Lines Flight 401	Gear position light failure Poor task management	All crews were obsessed by small light problem and forgot to maintain altitude
United Airlines Flight 173	Gear position light failure Poor threat error management	Captain postponed landing to deal with the gear problem and ran out of fuel
Tenerife airport disaster	Poor visibility Congested air traffic Poor communication	KLM captain started take-off without clearance and collided with another Jumbo jet on the ground

Table 1 : Typical aircraft accidents

The above are examples of typical accidents in the modern aviation industry[2][3][4]. The common cause in all accidents were poor management practices and authoritarian

old habits that inhibited communication between crews. CRM has therefore developed in a way that removes these factors. In the other words, CRM was created as a system to maximize the use of human resources in the cockpit by facilitating communication between crew members.

Because of this background, CRM training in airline companies now became a mandatory one for pilots. In addition, this CRM concept has been applied to other industries such as marine transportation and nuclear power powerplant operation to reduce mistakes. These facts show its effectiveness and importance.

However, the current environment surrounding CRM is fraught with problem. The problem is rooted in its ambiguity. Specifically, as noted above, CRM deals with the abstract concepts of communication and management. Therefore, it makes difficult to quantitatively evaluate trainee`s skill. This is because items such as communication include very comprehensive and complex content. Therefore, CRM training in the aviation industry today completely relies on subjective evaluations from the viewpoint of an evaluator, an experienced pilot. But the people who evaluate them are also human and make mistakes. Suppose the evaluator makes a mistake on evaluation. Then the CRM trainee will conduct operations based on the wrong experience, which can produce another mistake. Such a negative cycle could occur. In the worst-case scenario, it would be a situation that could result in the loss of human life.

Figure 1: Burning fuselage of JAL (Cited from Nikkei)

For example, the JAL plane collision accident on January 2, 2024, which killed 5 people. Although this accident is still under investigation, it has already been pointed out that probable cause in this case was malfunction of CRM. Therefore, we can say that there is an urgent need to establish safer evaluation methods to prevent such accidents.

According to these reasons, this study aims to address the following two items. One is to build an objective method of evaluation in CRM training, regardless of who is doing the evaluation. The second point is to verify whether the methodology constructed above actually works effectively for evaluating the content of CRM training.

In the next section, we will first provide a more detailed description about CRM. Then, description about the issues identified in the literature review will be explained.

1.2 - What is the CRM?

First of all, the definition of CRM is "the effective utilization of all available human resources, hardware and information to achieve safe and efficient operations" (cited from AIM-J Chap 9-2).

According to the book of Aeronautical information manual – Japan 2019(written as AIM-J below), it defines CRM as below.

【CRM Skills】
The following non-technical skills that crew need to operate safely and efficiently are referred to as CRM skills.

a） Situational Awareness
Skills to not only recognize what is happening inside and outside the aircraft, but also to analyze it and predict how it will change in the future. The objects to be recognized are “The aircraft”, “The external environment, “Time” and etc. In many accidents, the loss of Situational Awareness of the crew is cited as a factor.

b） Decision Making
Skills in identifying problems, devising solutions to them and reflecting on post-decisional actions, which are necessary for the decision-making process. The process of resolving problems that arose in operations and reflecting on the results is Decision Making.

c） Workload Management
Skills to handle tasks arising in a variety of situations appropriately and to maintain performance levels above a certain level at each stage. It is important to note that

overloading increases errors because it exceeds human processing capacity, and if the workload is too low, the human brain goes into standby mode.

d) Team Building

Skills of the crew to form an effective functioning team. In aircraft with multiple crews on board, each crew member needs to participate in the operation and function as a team. This requires maintaining an appropriate 'authority gradient' for the relationship between the captain and co-pilot. Each crew member provides leadership appropriate to his or her role, cooperates by offering assistance to other crew members when necessary, and constructively resolves conflicts of opinion.

e) Communication

Skills to communicate information, intentions and opinions about operations clearly and without misunderstanding. In multi-person operations, proper communication between crew members is essential to ensure that they are aware of each other and that they are making the best decisions as a team (Decision Making).

The above is the definition of CRM by AIM-J. As CRM skills are not globally standardized and are set by individual airlines based on their operating patterns and policies, the expression of CRM skills varies from airline to airline and consists of descriptions and behavioral indicators that are somewhat ambiguous. Figure 9-2 is an example of CRM skills with behavioral indicator published by JAXA in 2003.

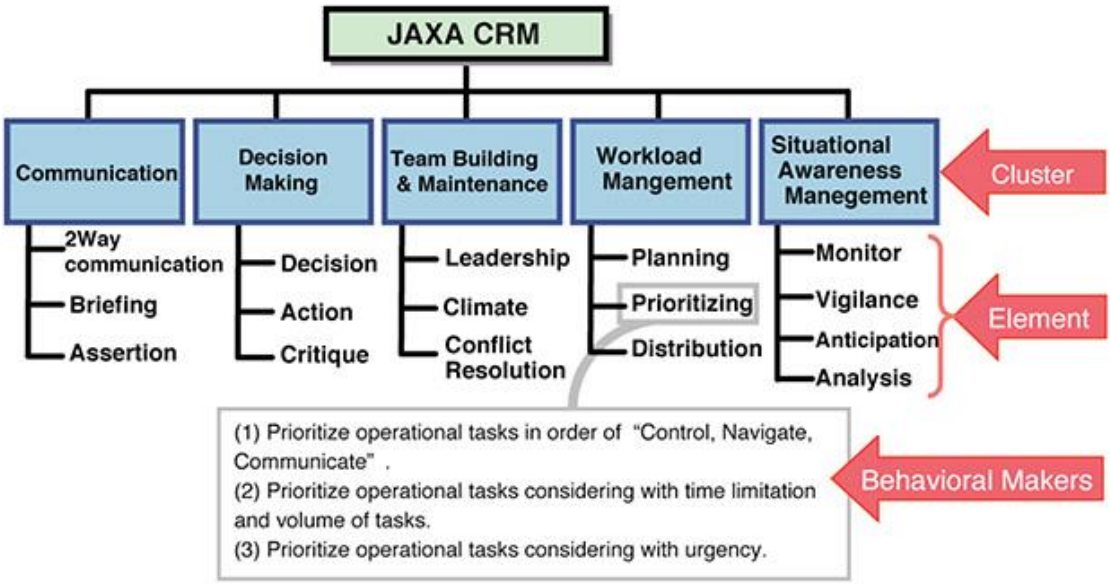

Figure 2: CRM defined by JAXA

The content of this CRM as defined here is generally consistent with the definition of CRM cited earlier. This trend is not confined to domestic inter-organizational one. In fact, CRM has been established as a guideline of similar content in other countries. The following diagram illustrates each skill area defined by the UK's Civil Aviation Authority (CAA) as contents to be covered by CRM training.

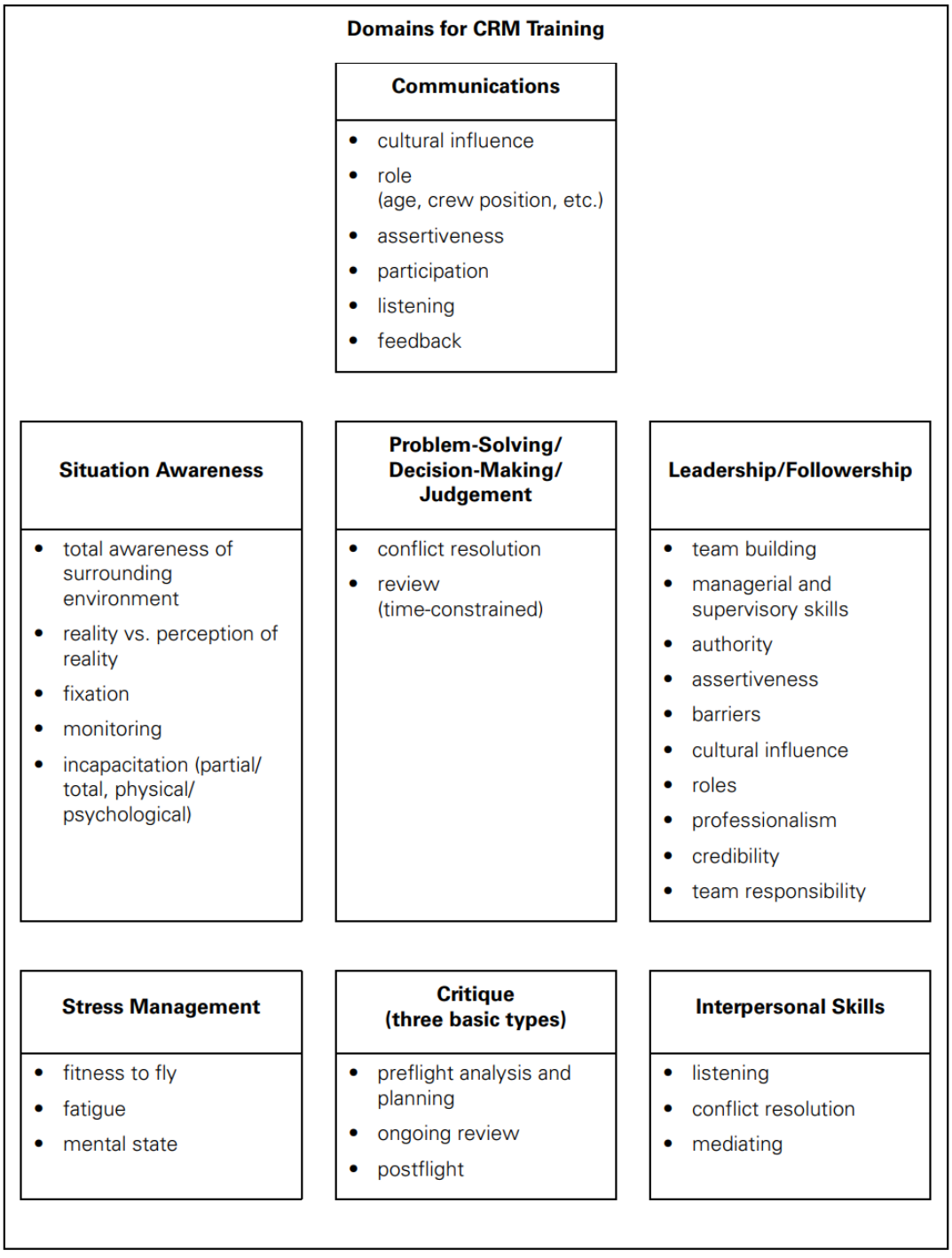

Figure 3: CRM defined by CAA

In both of these diagrams, there appears to be some degree of definitional agreement in all organizations in the areas that are fundamental to CRM, such as management skills and communication skills related to problem-solving.

Chapter 2 - Literature review

2.1 - Outline of literature review

This chapter will examine whether any studies have so far analyzed CRM skills.

Let's start with a broad overview of the relevant studies. CRM evaluation methods studied to date have evolved along with three main steps (2002, O'Connor)[5]. The first attempted to analyze the effectiveness of the training itself by examining crew members' reactions to CRM training. This is a study that attempted to measure its effectiveness mainly in the form of a questionnaire (1984, Helmreich)[6]. The second study measured changes in crew member attitudes and knowledge regarding CRM training. Again, measurements were often made in the form of questionnaires, in line with the first study. Finally, the third method of measuring the effectiveness of CRM training differed from the two previously mentioned methods in that the experimenter checked the CRM skills of the flight crew during actual training or on the flight. The following sections will detail what results these three CRM skills analysis methods have produced or what problems they have shown.

2.2 - CRM evaluation based on reaction

At first, let us discuss the first study described in the overview. As mentioned earlier, this was an attempt to analyze the effectiveness of the training itself by examining crew members' responses to CRM training (2001, Salas et al)[7]. This is literally a study that attempted to measure the effect of training, albeit indirectly, by comparing questionnaires at two time points, before and after training on CRM. For example, let's assume that someone who was skeptical about CRM before the training strongly felt its effectiveness after the training. If training is in fact effective, then the responses on the questionnaire should change noticeably. This study was based on such a claim. In fact, this prediction was correct. Please see the chart below. This figure shows the results of CRM training for US Naval Aviation and whether the training was effective on a scale of 1 (strongly disagree) to 5 (strongly agree) (2012, O'Connor)[8]. The figure shows that while there are exceptions in some groups, the overall trend is that the majority generally feel the benefits of training over a long period of time. Although the experiment in this study was conducted with Naval Aviation rather than a civilian airline, it could be applied to civilian airlines because it was measured in military

training, a group in which teamwork skills are more important. However, there is a problem with this study. Because the study focused solely on the perceived effectiveness of CRM training, it did not measure each of the specific skills that CRM includes. CRM, as we have mentioned, covers a variety of areas, such as task management skills and consensus-building skills. However, they have not examined the effectiveness of each component of CRM in this report but did only in the ambiguous sense of the term. Critically, the study uses a similar measurement approach to customer satisfaction surveys conducted by normal companies. Thus, even though this study found that the concept of CRM had a positive impact on team building, it did not know which elements in the training had such an impact.

Figure 1 The percentage of respondents 'agreeing' or 'strongly agreeing' that CRM is helping to improve mission performance and safety

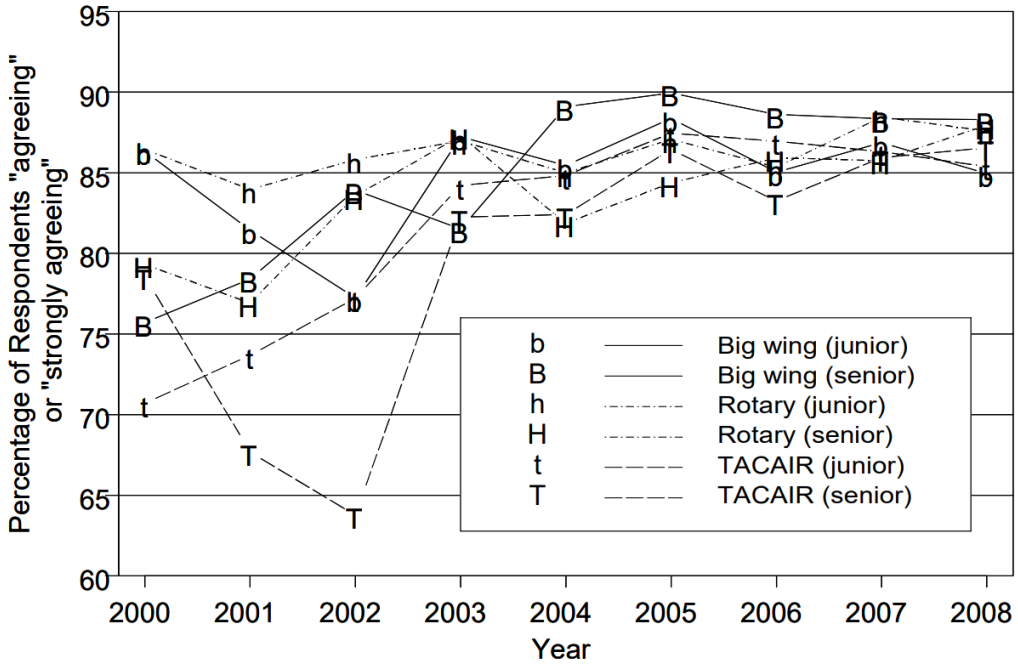

Figure 4: Pilot`s good reaction before and after CRM training

2.3 - Multimodal CRM evaluation

After the problems described in the first study was recognized, a study was conducted, this time dividing CRM into several components and surveying them (2001, Holt et al)[9]. Although these groups of questionnaires are classified in detail, they can be broadly divided into two types. One is the change in attitude through the CRM training of the crew. This is comprehensive, including psychological factors that focus on

personal aspects, such as stress felt by the subjects themselves and by others, as well as team factors, such as ease of communication and sense of leadership (2012, O' Connor)[8]. The second was a survey on knowledge about the concept of CRM itself. The method used here was to select answer from multiple choices on questions like what skills are important in CRM. This method can be easily explained by comparing it to a written test in the subject of CRM. Now, the results obtained using these methods were very useful. For example, a questionnaire survey on changes in the attitudes of participants through CRM training revealed a significant negative correlation between "stress felt toward others" and "communication" among young crew members. In other words, if a young crewmember has something to say to another crewmember (e.g., the captain) but if unable to say it, this can be a major stressor. This finding has made a significant contribution. Because it demonstrated the need for the implementation of assertive behavior training in young crew members. At the same time, the CRM knowledge survey yielded significant results. One example is that the results of the "written test" revealed a low level of awareness among the crews regarding situational awareness and decision making. This is because many of the human errors that lead to aviation accidents are caused by the lack of situational awareness and decision-making mentioned earlier. Therefore, it was significant to discover the tendency of many crew members neglecting these important factors. However, despite these achievements, problems remained with the research to this point. That is, both questionnaires were answered based on the subjective criteria of the crew members who participated in the experiment. Vague concepts such as stress and communication vary widely from individual to individual. Therefore, these questionnaire-based evaluation methods lacked objective persuasiveness. This was the remaining problem.

2.4 - CRM evaluation by “Evaluator”

Finally, we will introduce a study that attempted to solve these problems and its methodology. The main problem in the previous section was that the data were subjectively influenced by each individual participating in the experiment. Therefore, some evaluation methods were developed to bring uniformity to the standards held by these separate people (2001, Klampfer)[10]. There are several types of methodology, but they all have one thing in common: the evaluator is on board the actual training or flight. The common evaluator checks the subject during the actual training or flight. This would allow for a standardized evaluation without the results being influenced by subjective criteria held by the subject. Following three studies are the examples of it.

We introduce the outline of them in a few sentences.

Non-Technical Skills (NOTECHS) : (1998, Avermaete)[11]
The assessment items fall into two main categories: social skills and cognitive skills. From there, the subject is further divided into 15 items to be evaluated.

Line Operational Evaluations (LOE) worksheet : (1999, Ikomi et al)[12]
Crew performance was evaluated based on an evaluation axis classified into 20 items. The usefulness of CRM training was detected in 13 of the 20 items in an actual evaluation at a U.S. regional airline.

Line/LOS Checklist : (2001, Helmreich)[13]
Evaluates crew performance based on an evaluation axis classified into 14 items, and has received high acclaim, including being adopted by ICAO in the form of the Line Operations Safety Audit.

2.5 - Deficiencies in current methodology

There is no doubt that all of the methods described above are carefully constructed research methods in terms of assessing the essential aspects of CRM skills. However, it is difficult to say that the essential problem has been solved. The reason is that while these studies have succeeded in reducing the subjectivity of the evaluators from the previous research methods, they have yet to completely eliminate it. In other words, the experimenter or the person in charge of evaluating the training will evaluate the CRM skills of the target person, and this is always subject to human subjectivity. In addition to the point, the evaluation methods use discrete numbers on a Likert scale, so it is also difficult to say that it accurately measures the performance of the individuals. In light of these remaining problems, there is still room for improvement in this area of research. In summary, two points are written below about the deficiencies in current research on CRM evaluation.

1. Current methodology relies on subjectivity of evaluator.
2. Accurate and precise evaluation is difficult, because it's based on Likert scale.

2.6 - Research purpose

In the previous section, we introduced the studies that have been conducted so far to assess CRM skills. As can be seen from the description so far, CRM skills include ambiguous elements such as communication skills and leadership. Therefore, there is a problem that the evaluator's subjectivity becomes the basis for evaluation when checking these factors. The objective of this study is to eliminate these subjective evaluation criteria and to develop a framework that provides uniform evaluation criteria. At the same time, considering the fact that CRM training is now widely adopted in the airline industry, based on what was mentioned in the introduction of the previous study, it is clear that CRM training has the ability to improve the relevant skills. Therefore, it can be said that the framework described above should quantitatively capture the performance gains made by subjects in CRM training through the training.
To summarize what we have written so far, the objectives of this study can be divided into the following two categories.

1. To develop a quantitative analysis framework that eliminates subjective elements to assess CRM skills.
2. To verify whether the improvement of CRM skills and operational performance can be detected, and quantitatively evaluated in experiments using the framework.

These are the objectives of this study. When these objectives are achieved, we will be able to provide a framework that will produce the same results regardless of who evaluates the product, instead of the subjective human evaluation methods currently used in the airline industry. Thus, it shows that it is possible to provide a more standardized method in the sense that it not only eliminates the subjective gap between the training personnel doing the evaluations, but also eliminates the gap in evaluation criteria between airline companies. It is highly likely that the proposal and provision of such a standardized methodology will contribute to further safety in aircraft operations in the aviation industry, where the recent labor situation is in an unstable condition because of corona disaster.

In summary, this research has a certain social significance. Because it proposes a standardization method in the field of CRM, which is one of the most fundamental factors to safe aircraft operations.

Chapter 3 – Experiment method

3.1 - Experiment Overview

This section provides an overview of the experiments conducted in this study. The purpose of this study is to develop and verify a framework for evaluating the usefulness of CRM training. Therefore, a comparative experiment using subjects was employed for verification. Specifically, a comparison was made between an experimental group that had received CRM training and a control group that had no such CRM knowledge.

In the following section, specific methods will be described. In the first phase of the experiment, recruited personnel were assigned to four groups and individually trained on piloting skills and CRM skills. In the experiment, each subject is given false information about the education he or she will receive. For example, experimenter tell a subject that he or she is receiving longer and higher quality training than other subjects despite the shorter duration of training that person receives. Or vice versa. The subjects are then reassigned to the experimental and control groups, and then paired with each other. The pair consist of one subject playing the role of captain and one subject playing the co-pilot. A subject with low skill is selected to play the role of the captain, and a person with high skill is selected to play the role of the co-pilot. Then, after pairing up, each pair performs a task related to flight using the simulator. A detailed description of this task will be given in the Experimental Tasks section below. Data collected when each group completed the task was analyzed to determine how effective the CRM training actually was using a framework for CRM evaluation. The above is a summary of the experiments conducted in this study.

3.2 - Grouping

This section describes the methodology of the experiments conducted in this study.
The flow of the experiment is as follows. First, 16 subjects are recruited to participate in the experiment. Although the subjects here are just students but not professional pilots, this is not a problem. Because whether he or she has license is irrelevant to the improvement of CRM skills. In addition, for comparison purposes, it is easier to verify the effectiveness of the experiment by selecting subjects who has no knowledge about CRM. The recruited subjects are then classified into four final groups through two processes in the end. The first process classifies the recruited subjects into two groups, A and B. The difference between these two groups is the content of the pilot training

they receive. Group A is the group that receives advanced pilot training compared to Group B, its training is longer in duration and superior to that taken by Group B in terms of technical improvement. Conversely, the content that Group B receives is the minimum required. Therefore, after attending pilot training, two groups will be formed. Group A, which has superior piloting skills, and Group B, which has inferior abilities. The second process following the first is to further classify these two groups into two groups each. First, group A is divided into two groups, A1 and A2. The A1 group is then given CRM training, while A2 is not given any training or knowledge matters related to CRM. Similarly, do the same for Group B. The B1 group would be required to take CRM training, while the B2 group would be given nothing. Finally, the state of each group will be shown in the graph below.

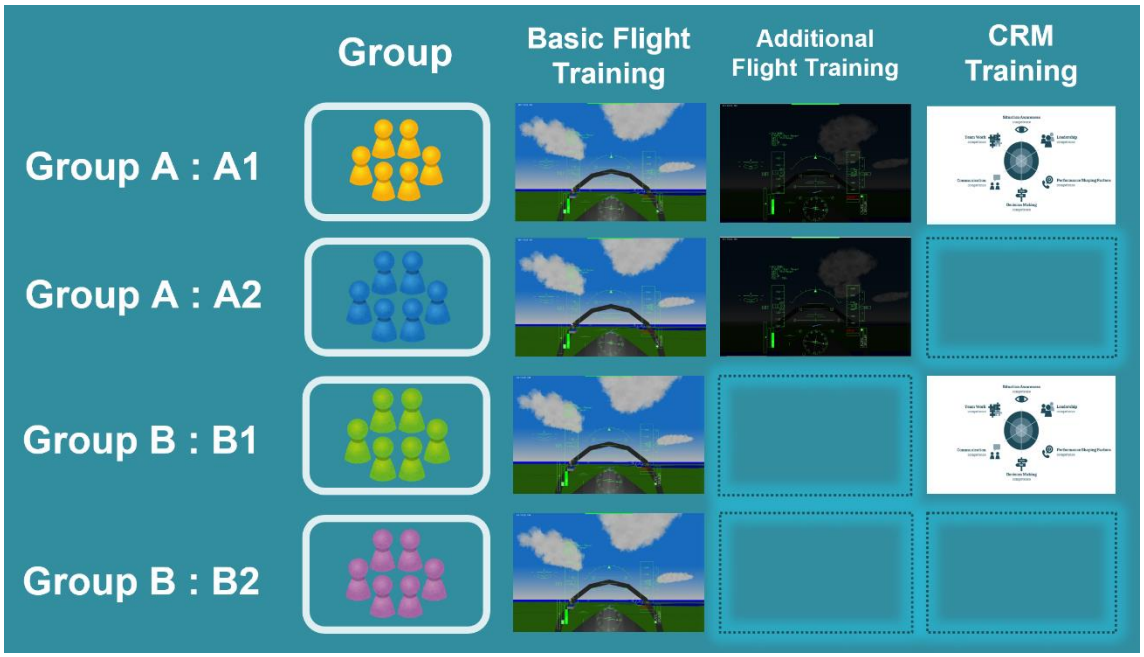

Figure 5: Classification based training subject receive

Each of the four groups formed through the process so far is then reorganized into two groups. Specifically, an experimental group with CRM training and a control group without CRM training will be formed. Accordingly, Group 1, the experimental group, is formed by combining members of Group A1, who have received advanced piloting and CRM training, and Group B1, who have learned minimal piloting and CRM skills. Then, Group A2, which is highly technically competent but has no knowledge about CRM, and Group B2, which has only minimal piloting skills, are combined to form Group 2, which has no knowledge of CRM training. This process will eventually result

in the formation of two groups, Group 1 and Group 2, as shown below.

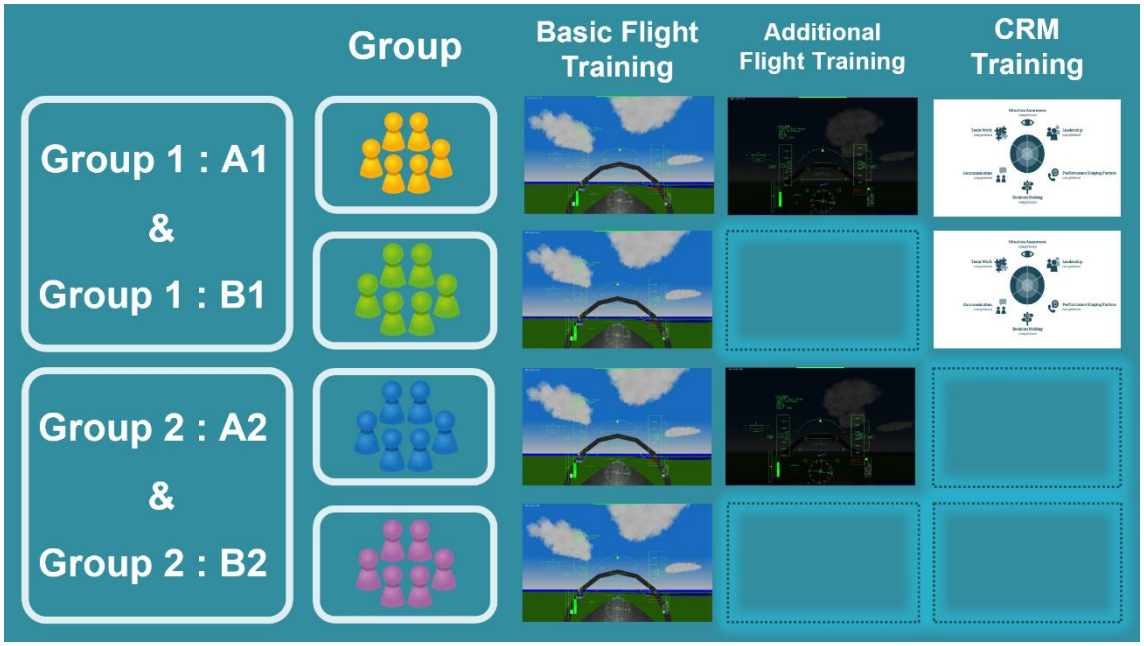

Figure 6: Classification into experiment group and control group on CRM

Group 1 formed via the process is the experimental group that received CRM training in this experiment, and Group 2 is the control group that did not receive CRM training.

3.3 - Pilot training

Above, we described the classification regarding the groups of subjects and the education each group receives. This section describes the training that all groups will receive in common regarding piloting. As mentioned in the experimental overview, this experiment involves a simulator-based task.

First on the premise, this experiment uses a flight simulator to educate the subjects on how to operate a helicopter. In the following, we will explain how the task will proceed through the process from the subject's point of view.

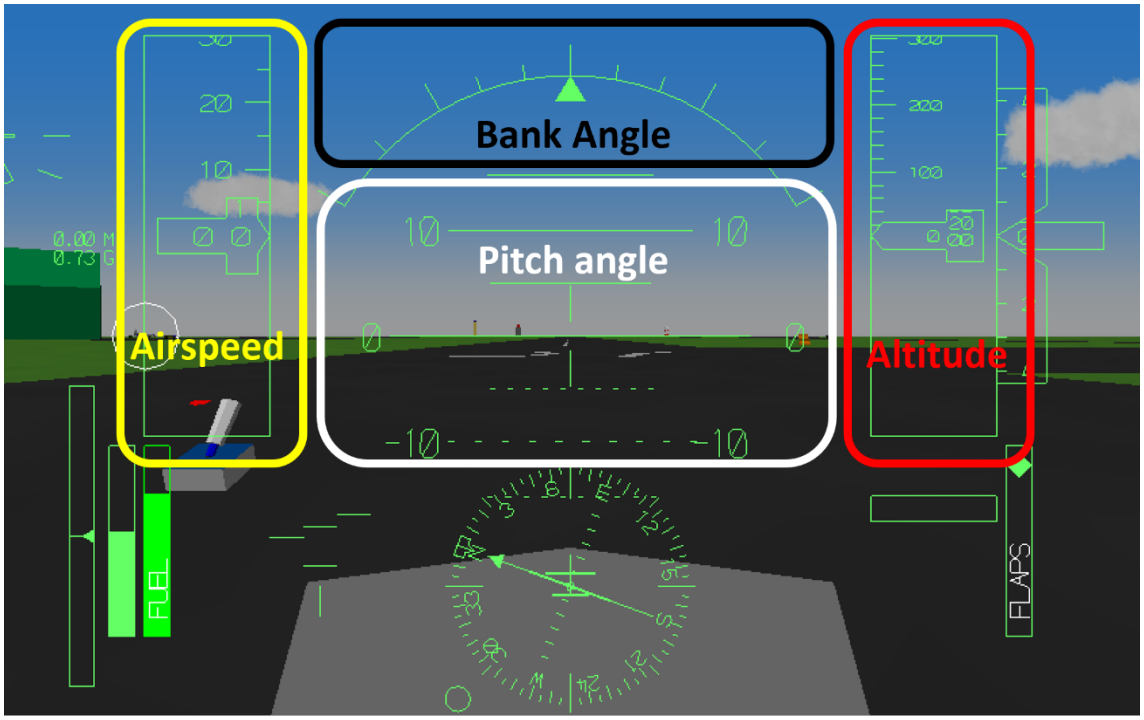

Figure 7: Screen displayed for pilot trainee

The first thing I will explain is the meaning of the screens displayed through the simulator. The figure shows the instruments and their meanings just after the start of the task, when the aircraft is on the ground. Let's look at each instrument one by one in turn. The yellow-circled instrument on the left indicates the speed. This instrument, called a speedometer, literally displays the speed of the aircraft. Next, the instrument in the upper center surrounded by black is the bank angle, which indicates the left-right tilt of the aircraft. If the aircraft were to tilt to the left, the arrow indicated by the triangle in the center would also tilt to the left. In other words, the current state in which the arrow points completely to the center means that the fuselage is horizontal with no tilt to the left or right. At the same time, the white area in the lower center is explained. This instrument, which has a series of horizontal lines, indicates the pitch angle, or the vertical tilt of the aircraft. On the screen, you can see numbers such as 0, 10, and -10 placed next to horizontally displayed lines. This indicates how much the aircraft is tilted from the horizontal line. And when it combined with the bank angle indicator described earlier, they are called an attitude indicator that indicates literally the attitude of the aircraft. Finally, we will explain the instruments that also play an important role in this experiment. The area circled in red on the right is the altimeter, which, like the speedometer, literally indicates the altitude of the aircraft. If the value indicated by this instrument were to drop to zero after takeoff, it would mean a landing

or crash, so it would not be an exaggeration to say that it is one of the most important instruments.

Figure 8: Environment and equipment used in experiment

The following is a description of the methods used to operate the aircraft. The two control rods shown in the figure are used in this experiment. By moving each of these sticks of equipment up, down, left, and right respectively, the aircraft in the simulator space is manipulated. In the image, the control stick has minor switches, buttons, etc., but they are not used extensively in this study. For this reason, the next section will focus only on stick movements and explain specific operations in the task.

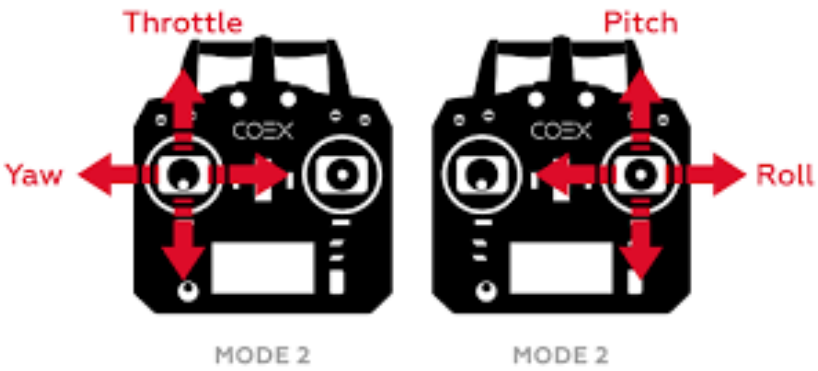

Figure 9: How to control

This section explains how the left and right stick operations are linked to the movement of the aircraft. First, let's talk about the left stick operation. As shown in Figure 8, the left side is a little more complicated to operate. This is because the simulator requires a subject to operate the aircraft in the XY plane in the simulator space in addition to the Z axis by only moving single stick. Let me explain in detail. First, the up/down operation of the left stick is linked to adjusting the lift output of the aircraft. In other words, moving the stick up will raise the aircraft. Conversely, moving the stick down will cause the aircraft to descend. Next, we will discuss the case of moving the same stick to the left or right. This one corresponds to a left turn and a right turn. Here, it is important to mention that a turn is a maneuvering that does not involve movement. However, many readers may have difficulty grasping the image from the word "turning" alone, which is used in an aeronautical term. So let me explain this word with an example from daily life. Imagine yourself standing in front of a mirror in the morning, and a voice calls out to you from the side, and you turn to face the person who spoke without moving from where you are standing. This is a turn. Note that turning left at the next intersection is not defined as a turn because it is a composite movement of a turn and a translation. It may be easier to understand if you think of it only as indicating movement about the Z axis in local coordinates.

Next, we will discuss the operational implications of the right stick. This one is relatively easy to understand because it is linked to the tilt of the aircraft regardless of whether it is moved up, down, left, or right. For example, if the right stick is tilted to the right, the aircraft will also tilt to the right. Conversely, if you tilt it to the left, it will likewise tilt to the left. And if you tilt it up and down, you can look up the sky or look down to the ground. In other words, the aforementioned bank angle corresponds to the left and right operations of the right stick, and the pitch angle is linked to the up

and down operations.

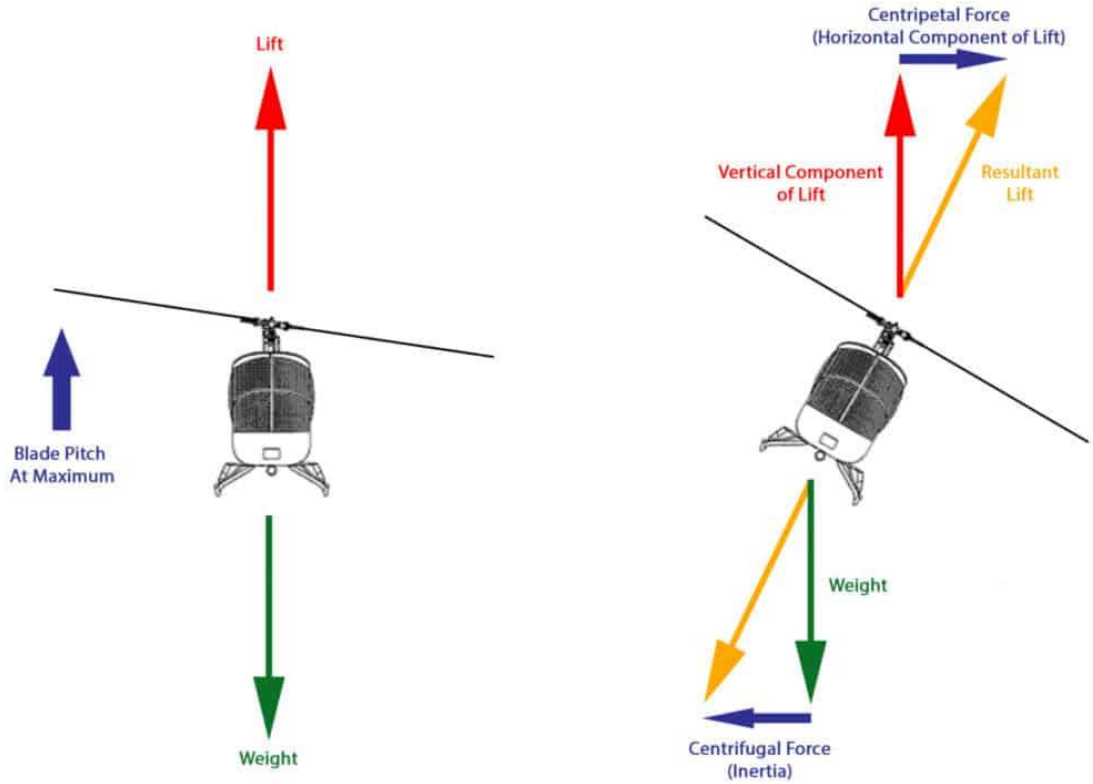

Figure 10: Mechanism of movement in a rotorcraft

Now, we know that the right stick is linked to the tilt of the aircraft. In addition to this. There is another important point to mention. That is, the inclination of the aircraft simultaneously implies a horizontal movement in that direction. Both drones and helicopters are characterized by the fact that they generate lift directly above the aircraft. In other words, as shown on the left in Figure 9, as long as the aircraft is horizontal, it cannot move horizontally. However, when the tilt of the fuselage is added, the lift force that was previously used only in the vertical direction now works in the horizontal direction as well. This is how rotary wing aircraft move. This is the reason why we stated that the right stick corresponds to the horizontal movement of the plane.

3.3.1 - Description of tasks in pilot training

In the previous section, we described the simulator used to conduct the experiments and its operation. In this section, we will describe the tasks to be trained using the environment.

The task in this experiment can be summed up in one word: "level flight. Simply explained, it is a method of flight in which the aircraft flies straight ahead on a predetermined course while maintaining a certain altitude. In the following, the flight process is divided into several sections, each of which is explained below.

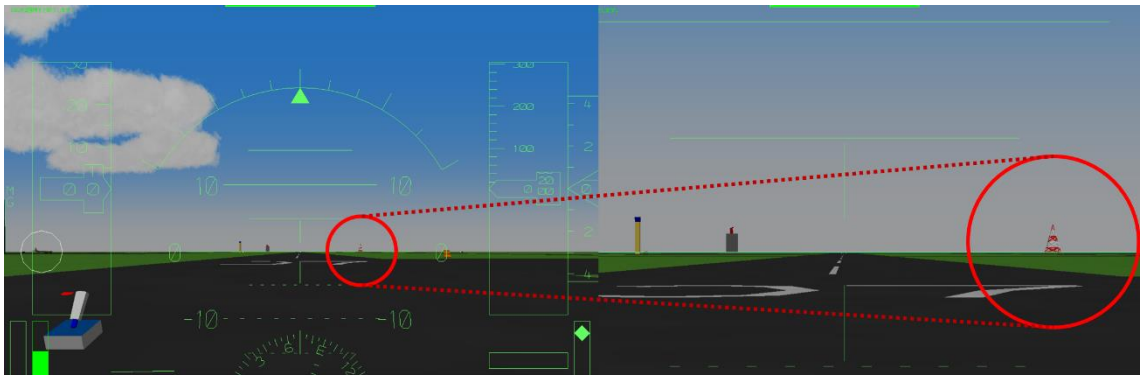

Figure 11: Enlarged image of the destination (target) in the assignment

Let's begin by dividing the flight process into five stages. The first step is to get the destination object in sight. This is because it is impossible to set a course without recognizing the goal object in flight. Therefore, subjects are asked to view the target object as shown in Figure 10 first as soon as the experimental screen is displayed. This is the first step. The second process followed is takeoff and climb. This is true for both helicopters and drones, but rotary-wing aircraft cannot basically move while grounded, so takeoff and the accompanying ascent are essential. Since this assignment requires the aircraft to be moved, this process was included as an item in the assignment. Then there is the third process, which is the maintaining of altitude. In real-life helicopter operations, helicopters often fly near the ground where there are many obstacles. Therefore, the skill to manage altitude is extremely important. In addition, failure to maintain altitude can impede landing, even in the absence of the obstacles mentioned above. Therefore, we incorporated this process into the assignment because we felt it was necessary to validate skills to maintain a certain level of altitude.

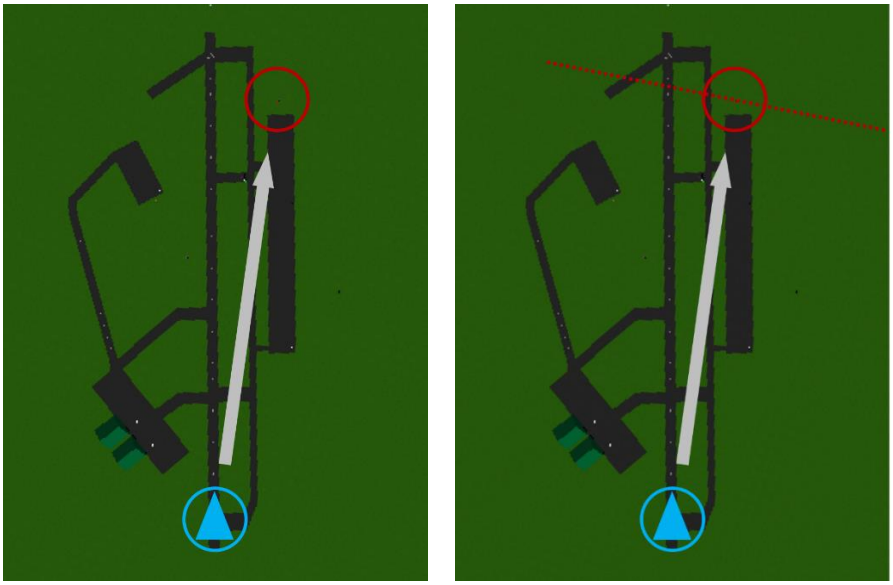

Figure 12: Course from initial position to destination (left), experiment end line perpendicular to the course (right)

The fourth is the process of flying to the destination while maintaining altitude and course. This probably needs no explanation, but an aircraft can be dangerous if it cannot fly the course intended by the pilot. Therefore, it is necessary for the subject to have the skill to set his or her own course to the goal and follow it, while maintaining the altitude mentioned earlier. This is why the fourth process, maintaining altitude and course was included in the assignment. The left image in Figure 11 shows the ideal course on this assignment. As you may be able to see from this figure, the course in this assignment does not follow a line along the initial position of the runway, but rather a line with a goal slightly off to the right from there. This was the result of setting up to more comprehensively evaluate the subject's maneuvering skills. This is because a simple course that can be accomplished by simply taking off and moving forward from the initial state can be achieved without using any of the previously mentioned "turning" maneuvers necessary for aircraft operation. Therefore, we set a diagonal course that requires turning movements as shown in Figure 11.

Finally, the fifth step of the process, which includes the termination condition of this assignment, is described. The goal in this task is to reach the destination, as mentioned at the beginning of this section. Note that the definition of arrival here is contact with the target object, not landing in the vicinity. This study is only intended to verify the effectiveness of CRM in aircraft operations, and there should be no need for subjects to spend a lot of time learning a skill that requires extremely difficult maneuvers such as

landing. This is the reason why the termination condition of the experiment was specified as reaching (contacting) the destination (target). However, it is naturally assumed that the target cannot be reached forever due to lack of the skill of the subject. Therefore, in this assignment, as shown in the right image in Figure 11, we set a line that is perpendicular to the course from the destination and add a condition that the experiment ends when the line is crossed. This has the advantage of reducing the burden on subjects due to the long duration of the experiment, as well as simplifying the trajectory data by relaxing the end conditions, making it easier to investigate deviations from the course. In consideration of theses advantages, we decided to terminate the experiment when the above conditions were met. These are the five processes in this assignment. Please refer to Figure 12 below, which reiterates the flow of the assignment.

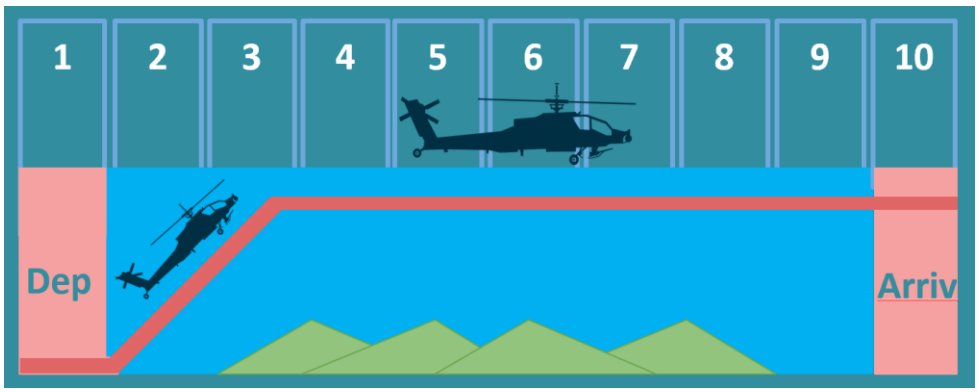

Figure 13: Flow of tasks in an assignment

3.3.2 - Pilot training received by each group

So far, we have explained about the method to operate aircraft and the task content of the flight. Based on what has explained so far, this section explains what kind of training each of the four groups created by sorting the subjects will receive. As mentioned in the first part of the experimental methodology regarding the classification of groups, there is a clear difference between Group A and Group B in terms of education on pilot training. However, it is not enough to conduct a comparative experiment when you do not even know how to fly and have no knowledge of the goals of the task. Therefore, the knowledge of how to operate the aircraft and the tasks described so far will be educated in the same way in Group A and Group B.

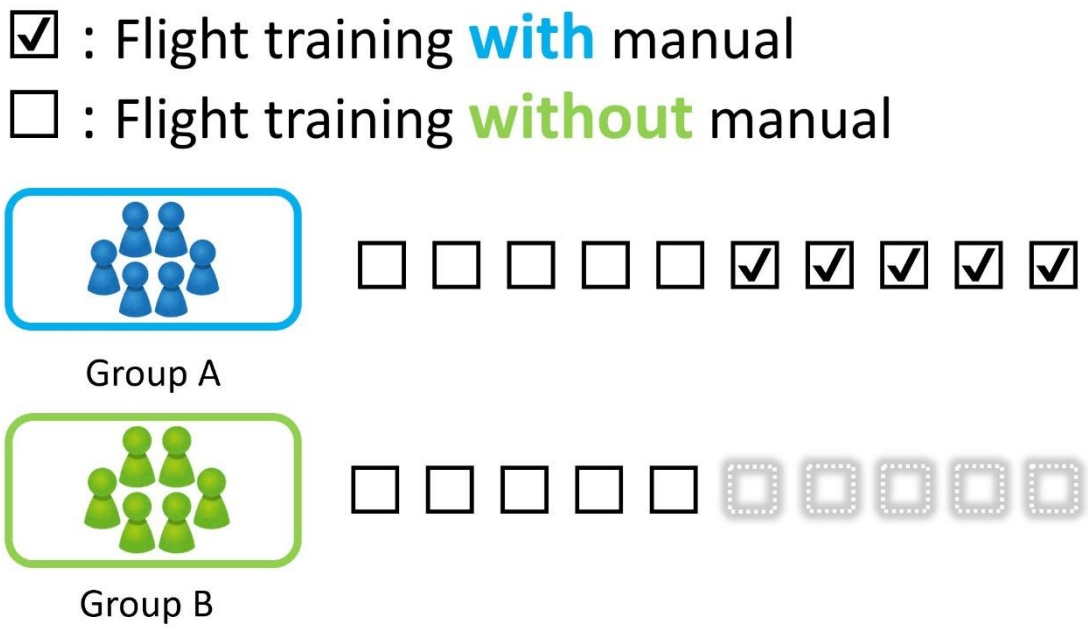

Figure 14: Difference of training each group receive

The part that differentiates between the groups is the amount of practice and its efficiency. In this experiment, the piloting skills of Group A must be better than those of Group B. To do so, first, each member of Group A will be asked to perform the described task 10 times. In addition, the training shall be conducted using manuals to further improve the efficiency of the training. The reason for using manuals here will be explained later. Conversely, Group B is given only the minimum amount of training necessary for the comparative experiment to be successful. Specifically, Group A received 10 flight training sessions while Group B received only 5 flight training sessions. This can create a noticeable difference in maneuvering technique between Group A and Group B.

3.3.3 - Reasons for using manual

Above we mentioned that in Group A training, we use manuals as well as giving relatively long training time. This section explains the reasons for using this manual for training.

Manuals are generally known for or give a strong impression of their role in preventing errors in practice. It is not wrong at all. Actually, aviation industry employs manuals in such purpose in a perspective.

However, the advantage of using manual is not limited to the prevention of errors described above. A specific example is explained based on training in the aviation industry. Normally, the cost of training pilots in the aviation industry is very expensive. This is because the expensive commodity of an aircraft needs to be used only for one trainee, not for transporting cargo or passengers. Therefore, it is fair to say that the aviation industry tends to avoid training with actual aircraft whenever possible. This is where the manuals mentioned above come into play. This is because the manuals enable the user to clearly understand what to do at each point in the operation without having to actually fly the aircraft. As an example, see Figure 4 below. This is part of a manual for a Cessna aircraft used in actual training. The figure clearly and concisely describes what the pilot must do at each step of the process, from takeoff to landing. By reading these manuals in advance, training in the aviation industry can get off to a smooth start without any confusion when boarding the actual aircraft and beginning training. This greatly reduces the time required for training in the actual aircraft.

These are the roles of manuals in the aviation industry. This study used this mechanism to incorporate manuals into flight instruction in simulations to not only differentiate training time, but also to establish a more pronounced difference in flight technique between the A and B groups.

※The reason why the first five training sessions in Group A are conducted without manuals is that it is more effective in improving skills than conducting all training sessions according to manuals.

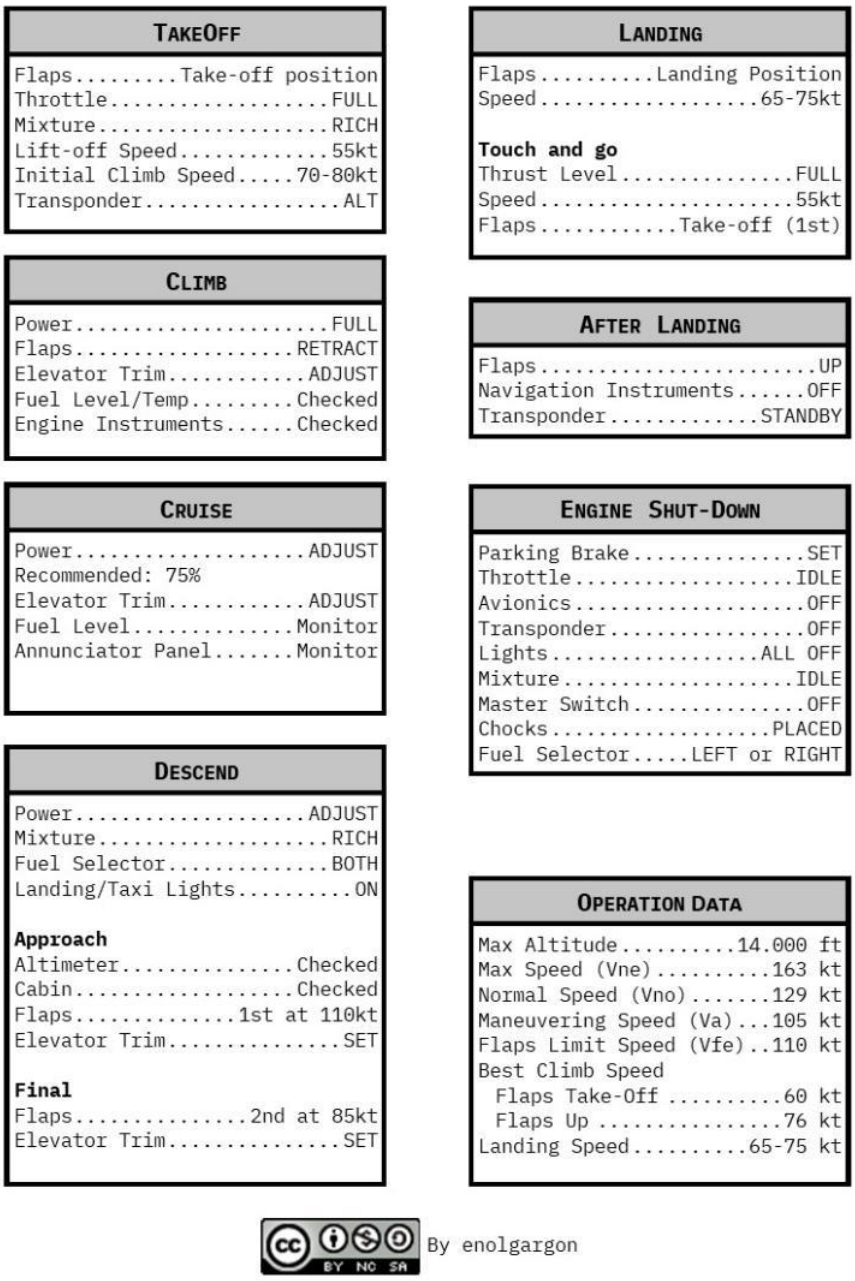

Figure 15: Example of manual (Cessna 172)

3.4 - Direction of training in CRM education

The following section describes the training on CRM, which is the primary focus of this study.

At first, the CRM training in this study will be scenario-based training. More specifically, it is based on the aforementioned pilot training. This is training for a specific situation in aircraft operations. The reason for adopting this method is to enable the trainee (subject) to take the most appropriate action depending on the situation. Basically, for the vast majority of aircraft operations, there is always an optimal solution. For example, consider a situation where one of several engines fails after takeoff. In this case, the optimal solution is to continue climbing after takeoff and maintain a certain altitude to ensure a safety margin before landing at the airport where the aircraft took off. This is because continuing to fly for an extended period of time with one engine out of order is a very risky practice. In addition, the option of aborting takeoff and immediately grounding the aircraft on the runway is also risky. This is because a normal takeoff uses a large portion of the runway, and there is likely to be no space available for a safe stop if an attempt is made to land on the same runway immediately after takeoff. For these reasons, the best solution in the above example is the option of landing after keeping a safety margin while maintaining flight capability. Once again, optimal solutions like this exist for the majority of situations in aircraft operations. Therefore, this study uses scenario-based training with specific situations, such as those used in actual airlines in the form of LOFT (line-oriented flight training).

Then, let`s continue to discuss training focused on CRM. Training in CRM skills is a matter that should normally be considered from a long-term perspective that includes classroom and practical training. In fact, even if we only look at the items of Non-Technical Skills (1998, Avermaete), which has a small number of evaluation items, we can see that the evaluation targets are subdivided into more than 10 items. Below are specific examples of what items are included. The list below is divided into three main categories: Cooperation, Leadership and Managerial Skill, and Situation awareness. The evaluation items in each category are then further classified into 3 or 4 small evaluation items, eventually forming a small group of 15 evaluation items. These are the small items that will be evaluated in the CRM training.

- Category: Cooperation
  - Team building & Maintaining
    - ✧ Establishes atmosphere for open communication and participation
    - ✧ Encourages inputs and feedback from others (lower the barriers)
    - ✧ Does not compete with others
  - Considering others
    - ✧ Takes notice of the suggestions of other CM even if s/he does not agree
    - ✧ Takes condition of other CM into account
    - ✧ Gives personal feedback
  - Supporting others
    - ✧ Helps other crew members in demanding situation
    - ✧ Offers assistance
  - Conflict Solving
    - ✧ Keeps calm in conflicts
    - ✧ Suggests conflict solutions
    - ✧ Concentrates on what is right rather than who is right

- Category: Leadership and Managerial Skills
  - Use of Authority/ Assertiveness
    - ✧ Advocate own position
    - ✧ Takes initiative to ensure involvement and task completion
    - ✧ Take command if situation requires
    - ✧ Motivates crew by appreciation and coaches when necessary
  - Providing and Maintaining Standards
    - ✧ Ensure SOP compliance
    - ✧ Intervenes if task completion deviates from standards
    - ✧ With crew being consulted deviates from standards if situation requires
  - Planning and coordination
    - ✧ Encourages crew participation in planning and task completion
    - ✧ Clearly states intentions and goals
    - ✧ With crew being consulted, changes plan if necessary
  - Workload management
    - ✧ Distributes tasks among the crew; checks and corrects appropriately
    - ✧ Secondary operational tasks are prioritized to retain sufficient resources for primary
    - ✧ Allocates enough time to complete tasks

- Category: Situation Awareness
  - System Awareness
    - ✧ Monitors and reports changes in systems states
    - ✧ Acknowledges entries and changes to systems
  - Environmental Awareness
    - ✧ Collects information about the environment
    - ✧ Contacts outside resources when necessary
    - ✧ Shares information about the environment with others
  - Anticipation
    - ✧ Discusses contingency strategies
    - ✧ Identifies possible/ future problems

These are examples of evaluation items in CRM training. Training in an actual airline company should be conducted through long-term classroom lectures and exercises that cover all of these evaluation items. However, the key issue in this study is not the content of training to improve CRM skills, but rather the development of a framework to measure the effectiveness of the training. Therefore, it would be more meaningful in this study to subject the participants to a relatively simple training in which only the essence of CRM is extracted, rather than to conduct a complex training that can be applied to many evaluation axes as indicated above. This is because the position of the experiment in this study is what is commonly referred to as a proof-of-concept position. The purpose of this study is to demonstrate the possibility of quantitative assessment of CRM skills without subjectivity, but not to try to improve CRM training. Therefore, training with simplified content that can be handled in a simple framework that is easier to analyze should be used rather than conducting experiments using complex training that requires a complex framework. Based on the above ideas, we will select three particularly important elements from the above CRM elements and impose them on the subjects in the experiment.

3.5 - CRM training

As mentioned earlier, the CRM training to be conducted in this study will be limited to three types of training. Therefore, we will consider the common elements of the CRM training. Non-Technical Skills introduced earlier as an example, the Line Operational Evaluations (LOE) worksheet and the Line/LOS Checklist, which are widely used in the aviation industry today. Common factors in these examples.

3.5.1 - Communication

First of all, the most important aspect of CRM when it comes to evaluation criteria is communication skills. There is no doubt about this, as it is a fundamental part of CRM skills. However, the term "communication skills" has a wide range of meanings. For example, the term encompasses a wide range of definitions, including the ability to accept others' opinions and, conversely, the ability to convey opinions to others. However, the three studies just mentioned above all focus on assessing the ability to convey own opinion, despite their nuanced differences. The reason is simple. The CRM concept was originally developed because there were many cases of failure to say what needed to be said in situations where it needed to be said, such as the Tenerife accident. Therefore, assertiveness training will be incorporated in this study as a training to improve communication skills.

3.5.2 - Task management

Next in importance is task management skills. As in the case of the accident described at the beginning of this paper, no matter how many human resources are available for aircraft operations, they are meaningless if they are not properly utilized. In both cases, Eastern 401 and United 173, this lack of task management capability resulted in the loss of many lives, even though the aircraft itself was barely affected. Therefore, it seems reasonable to incorporate task management training to improve this capability.

3.5.3 - Threat and error management

Finally, there is a third important element. It is the ability to manage threats and errors. Like task management, it is related to management in some aspect, but the scope of its coverage is completely different. For example, in task management, the main objective is to list things that need to be prioritized in the "now" situation and to deal with them. However, threat error management has other roles. Specifically, this management skill is useful in organizing the situation with respect to the "future. Please remember the summary of the Tenerife accident once again. In this accident, the crew's failure to communicate the critical situation to the captain was cited as the cause of the accident, but the accident would not have occurred if the captain had been able to detect such danger in the first place. These examples show that situational awareness and planning for the future is definitely important. In addition to these

facts, task management skills for managing "now," threat/error management skills for managing "future," and communication skills to support these skills are the three most important skills at the core of CRM from the perspective of MECE.
Therefore, the CRM training in this study will incorporate the following three.

1: Assertive Action Training
2: Task Management Training
3: Threat and Error Management Training

3.6 - Specific contents of CRM Training

CRM training in this study should include the three elements mentioned earlier. What we propose is a training program that uses video and discussion of pilot training. The following is a detailed explanation.

First of all, we chose two subjects from Group A1 and Group B1, who will be taking the CRM training. Then, the selected two subjects form one pair. At this point, it is assumed that the two subjects have already completed their pilot training.

3.6.1 - Assertive action training

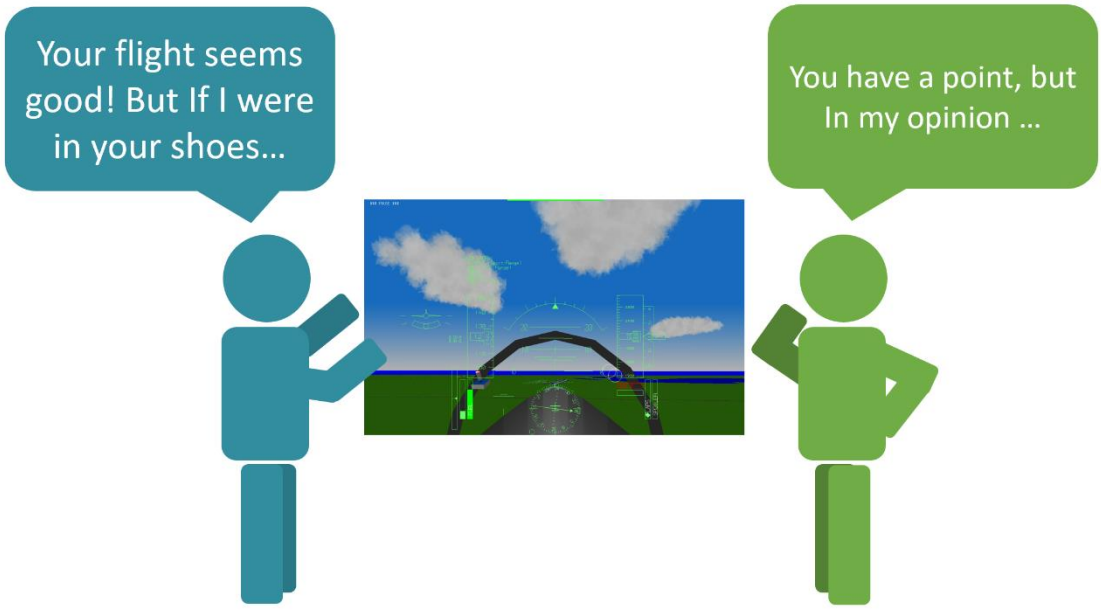

Figure 16: Assertive action training

First, the specifics of assertive action training are described. As mentioned above, this

training is designed to improve the ability to convey one`s opinion, among other communication skills. First, a video recording from one of the two participants' flight training is played and its content is shared between the two participants. Afterwards, the two discuss about the point to improve while reflecting on the content of the project. After this discussion, they will now share another participant's training record and discuss about it. This means that the two participants need to watch each other's training records and discuss their room of performance to improve. This is assertive action training. In this training, subjects need to express critical opinions, albeit in the form of suggestions for improvement, to people who are not very close to you. Normally, it does not frequently happen to exchange ideas directly with a person you are not familiar with. However, by imposing this training, it is expected to improve the ability to communicate one's opinions without being influenced by human relationships. At the same time, the nature of this training is to propose plans of improvement, so it requires subject`s statements to be persuasive. Therefore, it is expected to improve overall communication skills, which are not limited to assertiveness alone, in terms of developing logical thinking to convince others.

3.6.2 - Task management training

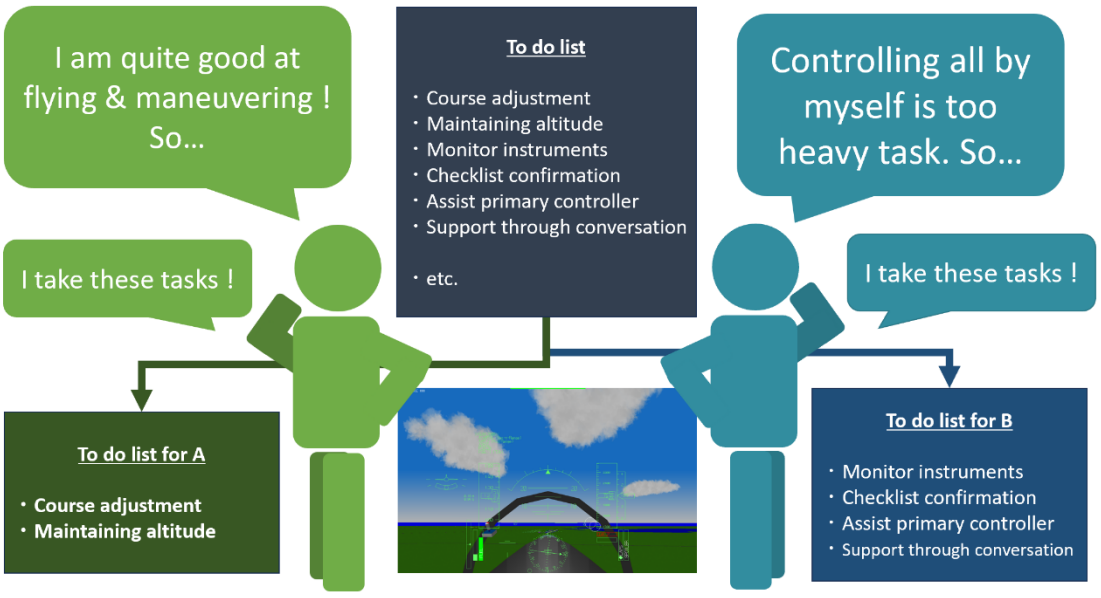

Figure 17: Task management training

The next training is task management. The training used here will be scenario-based, similar to the assertive action training described above. First, we give this pair a scenario of a situation that simulates a pilot training. The "scenarios" referred to here

are not tasks that are actually performed, such as pilot training the pair had taken beforehand, but are more like desk exercises. In the initial phase of the training, we instruct these two people to work together to list the tasks required in the scenario. For example, the tasks in this study could include altitude control, course correction, and instrument monitoring tasks. Then the participants discuss with each other which participant will now be in charge of the tasks that they have listed. Let's take another example. If you are confident in your piloting skills, you would inform your partner of this and accept the piloting-related task; if you are not confident in your piloting skills, you would accept the task required to support your partner. By preparing specific scenarios of flight in this format, this training seeks to improve participants' task management skills.

There are three reasons why we believe that this training is expected to improve participants' task management skills.

One is because it is expected to improve their ability in terms of situational awareness. This training includes the assignment of listing up the tasks required for the scenario. This is nothing more than the verbalization of what was done unconsciously in the pilot training. In other words, by imposing training to verbalize the necessary tasks at any given time, it makes subjects possible to reaffirm one's understanding of the situation.

The second reason is that it allows them to understand their own capacities in terms of their abilities. This training will require the participants to use the experience gained through pilot training to determine whether the tasks listed can be handled by them. Therefore, it is necessary to determine the limits of one's ability by reflecting on one's past experience in pilot training. Therefore, it is expected to improve task management skills in terms of understanding one's own capacities.

Third, it is expected to improve the ability of participants to reach consensus on tasks and their distribution among them. As mentioned in the first reason, the ability to verbalize the critical elements in each situation is essential for consensus building through communication among participants. In addition, the understanding of what you can and cannot do, as indicated by the second reason, will help you to persuade others in discussions on task allocation. Therefore, this training is considered to be useful in improving the ability to understand tasks and their management.

3.6.3 - Threat Error Management Training

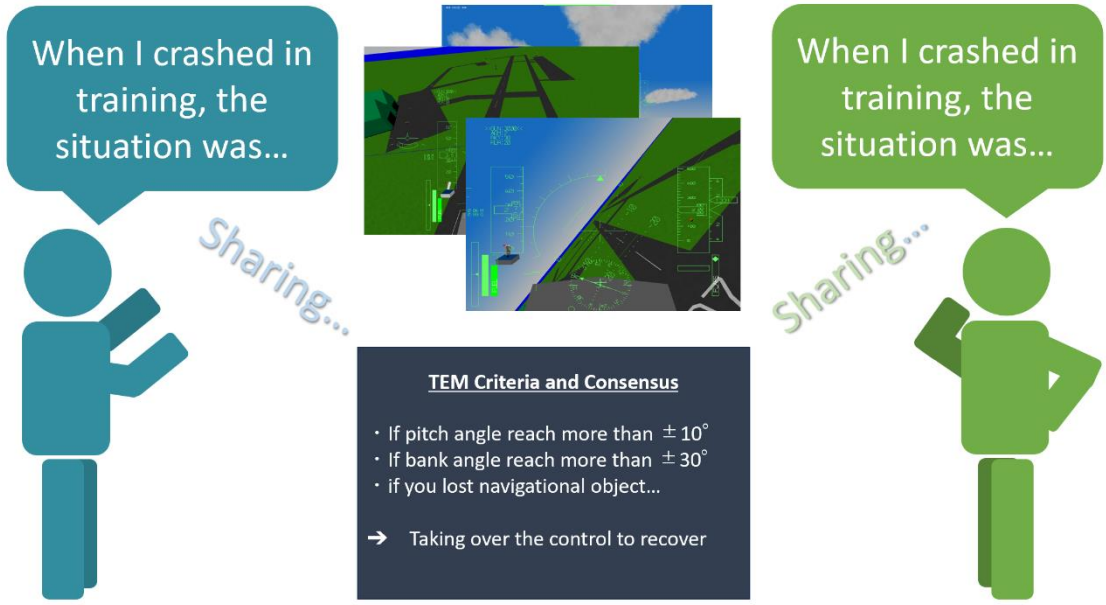

Figure 18: Threat error management training

The last one is Threat Error Management Training. This training has two purposes. One is to increase situational awareness and sensitivity to crisis situations by sharing the knowledge possessed by each trainee. The second is to develop skills to prevent crisis situations and to build consensus among trainees on how to correct such situations.

This training will be conducted in a discussion format similar to assertive action training. First, each trainee will share information about the critical situations they have experienced in their flight training. For example, while concentrating on course correction, one forgot to maintain altitude and crashed. Or experience such as deviating from one's intended trajectory because the aircraft was tilted too far is also good example. Examples of such dangerous situations in flight training are shared among trainees through conversation. The important point here is that, unlike assertive action training, the information is not shared using video recordings, but "photographs" are used instead. More specifically, each trainee is asked to select and prepare in advance several photographs of what he or she perceives to be critical situations during flight training, in order to share information. Then, during the discussion, the participants will share information and build consensus on the operational process regarding the crisis situation, using those photos in conjunction with the conversation. The reason for using photographs instead of moving images is

that it is possible to extract objective information, in the form of numerical values, about subjective situations that subjects perceive as dangerous. Let's use the following photo as an example.

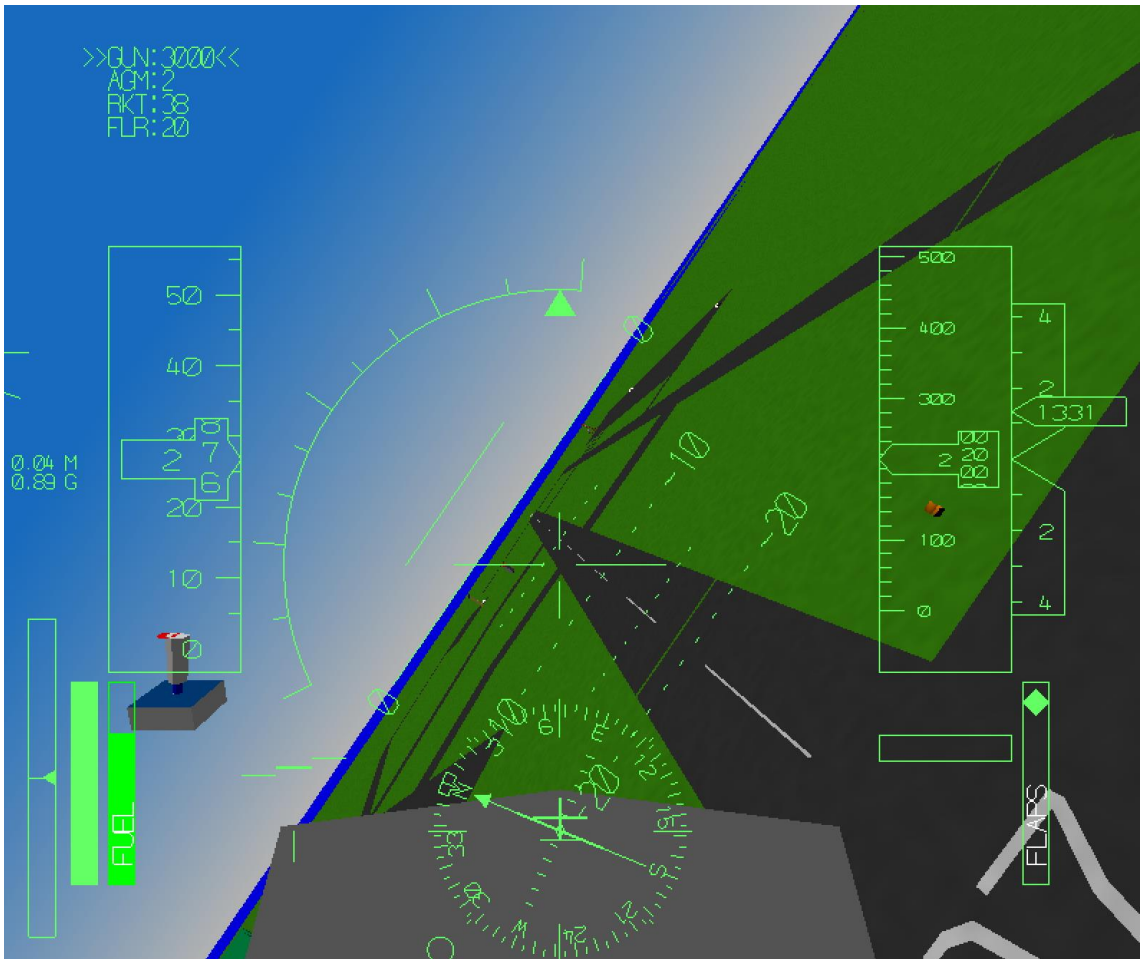

Figure 19: Example of picture used in threat error management training

Suppose that one of the trainees crashed during flight training from the above picture. In this case, it would be possible to share information in terms such as "the aircraft was tilted too far to the right, causing a critical situation." However, this expression is very subjective and ambiguous. This is because it is not clear to anyone other than the speaker how much inclination fits the definition of a critically crisis situation. This is also true for explanations using video media in combination. In the video, the numerical values of the instruments seen in the photo above are constantly fluctuating. This makes it difficult to share clear criteria for defining danger. On the other hand, photographs can provide a numerical cutout of the situation at the moment the trainee perceives it to be most dangerous.

In the example of photo shown above, the trainee felt that the situation at 27 kts, 200 ft altitude, 55 degrees bank angle, and -7 degrees pitch angle as the most critical situation. This training will discuss ways to prevent or recover from crisis situations by sharing information based on such data.

In the previous example,

・Bank angle should not exceed 30 degrees to prevent accidents
・If the bank angle exceeds 30 degrees, a warning will be issued.
・If the warning is not given in time and the bank angle exceeds 45 degrees, another person takes over the control.

These are the plausible results of discussion. The above is an overview of Threat and Error Management training.

3.7 - Experiment task

This section describes about the experiments in this study. In this experiment, two subjects are paired up and given the same tasks as those they had engaged in during pilot training. Although the tasks are the same, the environment of the experiment is very different from that of the training. First, all of the aforementioned tasks related to pilot training are supposed to be carried out by one person and not collaboratively with the others. In addition, although the CRM training was something that two subjects worked on together, they did not actually use the simulator while training. However, the experimental tasks in this study must be conducted using a simulator with two people working in pairs. The following figure shows the actual experiment.

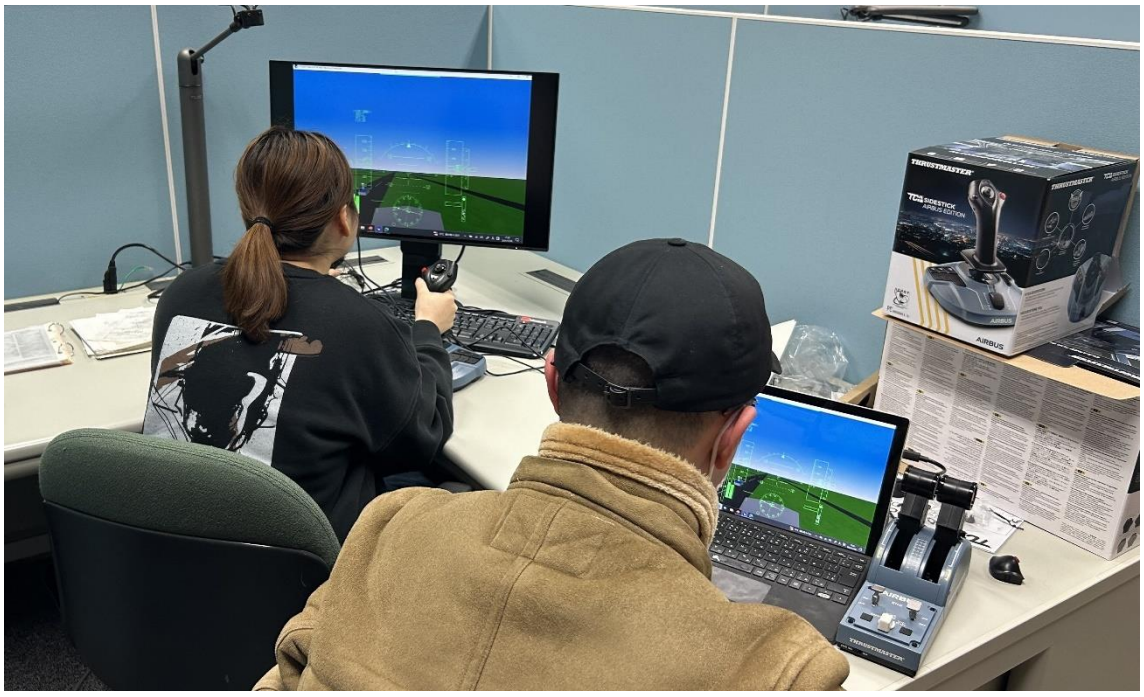

Figure 20: Experiment environment

The tasks to be performed in this experiment are the same "level flights" as those in the training, and the objectives and operating procedures are the same as those in the training. However, from the perspective of two people working on the same task, there are two distinct differences from training in this assignment.

One is to support the pilot through conversation. In this experiment, a member of Group B is first selected to play the role of captain, and a member from Group A is selected to play the role of co-pilot. Therefore, in the early stages of the experiment, the

tasks are basically divided between the pilot and his/her support. This means that you can leave the monitoring of instruments, including altitude, which you alone tend to miss, to the participants who will support you. For example, if the pilot is too focused on correcting the course in an assignment and loses altitude, the supporting participant can provide advice in the form of pointing out the loss of altitude. Therefore, this experiment differs from training in that it is possible to modify the flight by conversations like this.

The second difference is the division of roles in the assignment. The task in training required the pilot to be responsible for all controls by himself. However, this assignment allows participants to share or exchange tasks in the maneuver. For example, if a pilot finds it difficult to maintain altitude, he or she can have another participant take charge of the maneuver. Please take a look at the photo above. The participant in the back of the photo is the main pilot and can be seen using the two sticks to control the aircraft. And the participant in the foreground is the support person. The content of the screen seen by both participants is the same. Now, you can notice that a lever different from a stick is provided on the right side of the screen he is looking at. This is a device that controls altitude, and moving this lever can override the altitude-related operations performed by the main pilot. Thus, if the main pilot wants to concentrate on course correction, for example, he or she can detach only the task of manipulating the altitude and let other participants take charge of it. In the extreme, the entire piloting role can be taken over by another participant if the main pilot feels that he or she cannot complete the task with his or her own piloting skills. These are the main differences from training in this experiment.

Although multiple examples were given above to illustrate the differences between training and experimentation, in reality only a small number of instructions are given by the experimenter. Specifically, participants are simply given the following information.

- Objective: To reach the goal while maintaining a certain altitude like in training. To maximize performance by two people working together to achieve this goal.
- It is allowed for participants to talk each other during the experiment.
- It is allowed to share or switch the role between participants.

To summarize what has been explained so far, the objective of this experiment is simple.

To maximize performance on a task while cooperating in pairs. However, there is a interfering factor in this experiment, a power gradient, that inhibits cooperation between the two.

3.8 - Authority gradient

As noted in the overview of this experiment, each subject is given false information to misunderstand their own competence from the start of the training. For example, members of Group A are highly skilled compared to Group B because they have received long-term, high-quality training. However, during training they have been given explanations that are not true and have not been given the confidence commensurate with their abilities. This is because the experimenter will tell the subject following statement upon failure during training.

"To be honest, your ability is a little lacking compared to the members of Group B. However, there is no need to worry. You will be paired with a member of superior skill when the experiment goes live. He should be able to lead you."

Conversely, experimenter tells Group B members the following statement regardless of their performance during training.

"Your skills are very good compared to the members of Group A. In the actual experiment, you will be working in pairs to complete the tasks, so I would like you to lead your partner."

Of course, both of the above statements are completely false. However, there is no way for subjects to verify the facts of this information. This is because, pilot training is conducted individually. In addition to that, CRM training is scenario-based and discussion-based. There is no way to know what is being done individually, and there is no way to know the skills of the other participants in the discussion. Thus, each subject will have a confidence that is not commensurate with his or her ability. Thus, this experimental environment forms an artificial authority gradient.

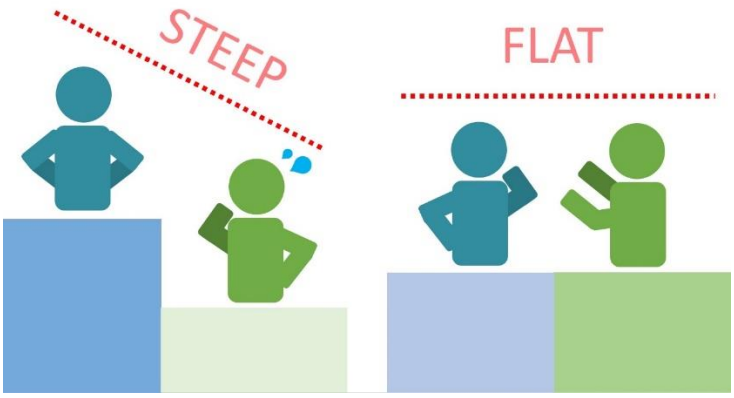

Figure 21: Image of authority gradient

An authority gradient is the power relationship among members within an organization. If the authority gradient is steep, the organization tends to consist of a strong, powerful boss and subordinates who are obedient to his or her instructions. On the other hand, if the authority gradient is flat, the organization is fluid, composed of equal members and not bound by hierarchical relationships. In terms of these authority gradients, we can see that false information creates an artificial authority gradient between Group A and Group B in terms of confidence in their competence. In addition, in this experiment, a member of Group B with relatively inferior skills is first selected to play the role of captain, and conversely, a member of Group A with higher skills is selected to play the role of co-pilot. Although these are roles in name only, such titles would further reinforce the authority gradient among members. Therefore, the point of this experiment we need to check is how CRM training can eliminate this authority gradient.

The above is a description of the experiments in this study.

3.9 - Schedule of the experiment

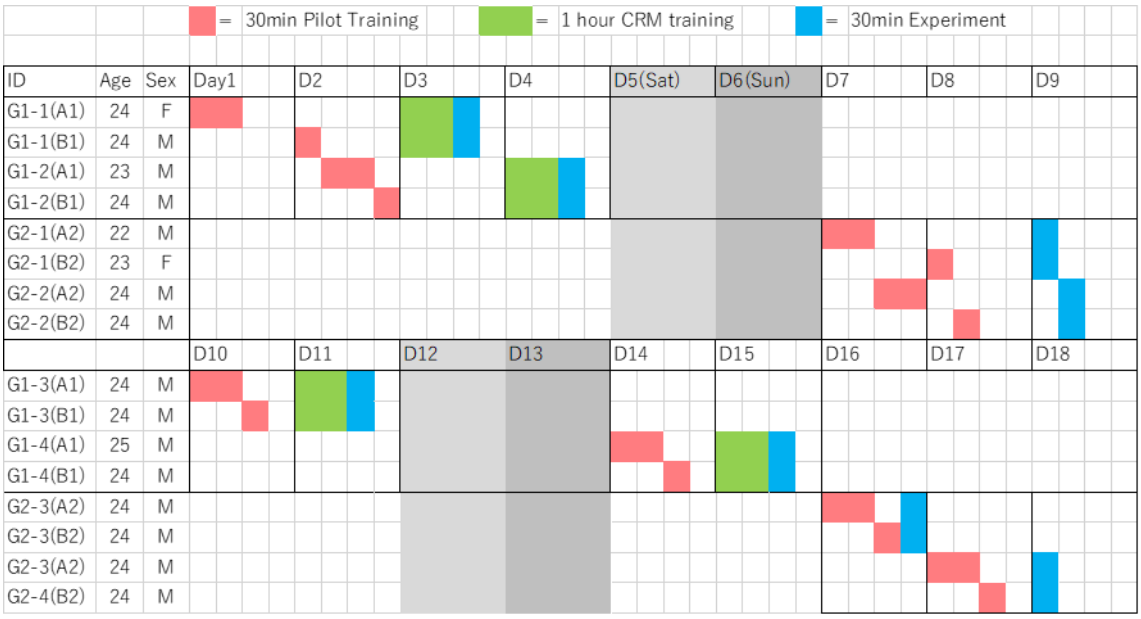

Figure 22: Schedule of the experiment

So far, we have described about the experiments in this study including pilot training, CRM training, and testing methods. Finally, this section describes attributes of subjects participating in this experiment and their schedule.

At first, the subjects in this experiment are mainly students and working adults between the ages of 22 to 25. Two are female and the remaining 16 are male. The schedule of the experiment is shown above. Because subjects participating CRM training and test in this experiment need to know how to use the simulator, they are required to conduct pilot training at first. Thus, each subject first undergoes pilot training. Then, before passing more than two days, members of Group 1 will attend on CRM training. Upon finishing these training, a pair consisted of Group 1 and Group 2 will then be tested.

Chapter 4 – Evaluation methodology

4.1 - CRM Training Effectiveness Measurement Framework

This section describes the evaluation method in the experiment. Since this evaluation method is a key point in this study, we will again review the assumptions that should be recognized in what we have written so far.

As a premise, the purpose of this study is to propose and validate a new framework for measuring the effectiveness of CRM training. This is because there are several problems with the evaluation methods used in CRM training to date. Specifically, the following two points are at issue.

・Only discrete or qualitative data using Likert scales are available
・No standardized evaluation criteria exist because the evaluation method is based on the subjectivity of the evaluator.

These two problems have existed. Therefore, in this section, we would like to propose a new evaluation framework that can obtain and analyze quantitative data with objectivity in order to solve the problems shown above.

This section provides an overview of the evaluation methodology used in this study. The main focus of this evaluation framework is on aircraft trajectories. The reason is simple. If CRM works well, the aircraft will follow an ideal trajectory for the pilot, and if CRM does not work well, the opposite will be true. Studies done so far have focused their evaluations only on pilot behavior to measure the effectiveness of CRM training. This was not a mistake. Because the purpose of CRM training is to facilitate communication among pilots. However, the ultimate goal of CRM training should not be to improve communication skills but to improve overall operational performance. Then, selecting the trajectory of the aircraft as the subject of evaluation is worth considering in terms of returning to the purpose of the training.

4.2 - Deviation from the course

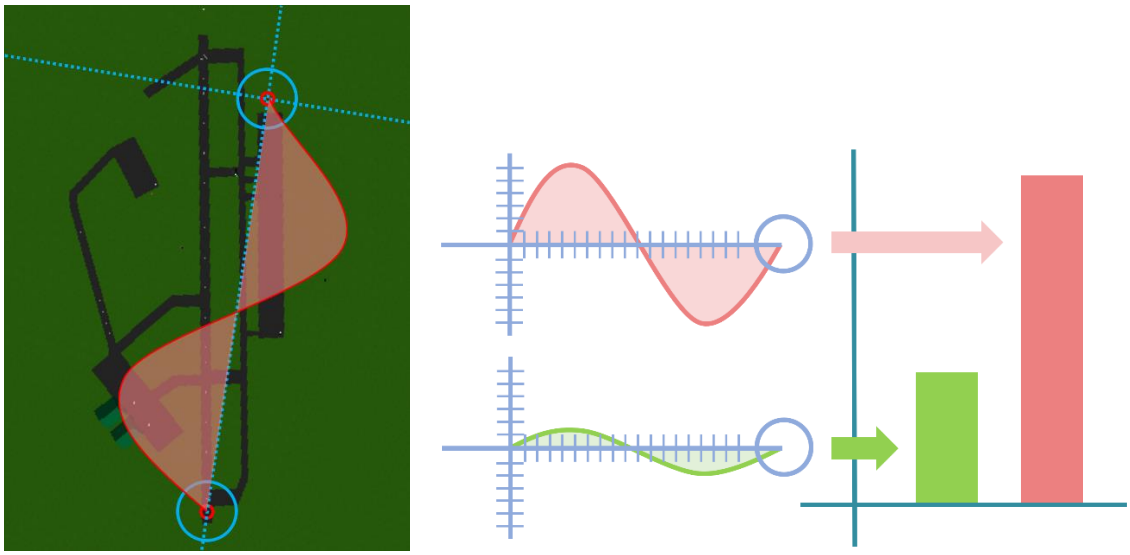

Figure 23: Deviation from the course

A more detailed description of this framework is given below. The first explanation is about deviation from the course. CRM training is often scenario-based with clear answers provided. For this reason, we have also prepared a clear answer to the experimental problem in this study: "The shortest course from the initial position to the destination. And as mentioned earlier, if CRM works well, stabilized flight paths is expected through improved performance. Therefore, a good rating should be given if the trajectory of the aircraft in the task is close to the ideal course, and a bad rating should be given if the trajectory is far from such a course. Therefore, in this evaluation framework, the deviation from the ideal course is the subject of evaluation. Specifically, the deviation from the course is determined as the area, and the evaluation is calculated based on that value. Above is an example of the courses and a graph related to the calculation method of the deviation for reference.

4.3 - Course change points

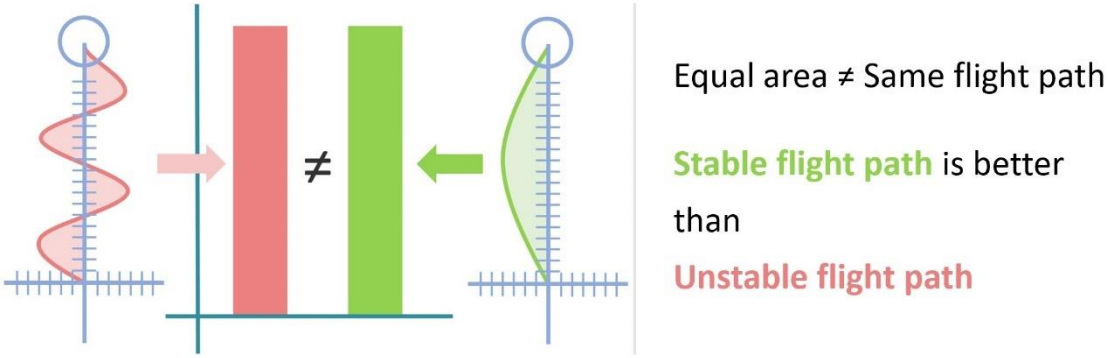

Figure 24: Different flight path with the same area

In the above, we stated that the effectiveness of CRM training is measured by deviation from course in flight. However, this evaluation standard alone is problematic when evaluating flights. This is because even if the areas related to deviation from the course are equal, there can be a significant difference in flight stability. Take a look at the figure shown above. The red course on the left meanders greatly, while the green course on the right is relatively stable. Intuitively, it can be said that the nature of the left and right courses is very different. However, these courses would be treated as the same flight content when evaluated under the previous criteria. The fact that similar ratings are given regardless of the stability of these flights is problematic. Therefore, in this evaluation framework, we added a stability item in addition to the deviation from the course criteria.

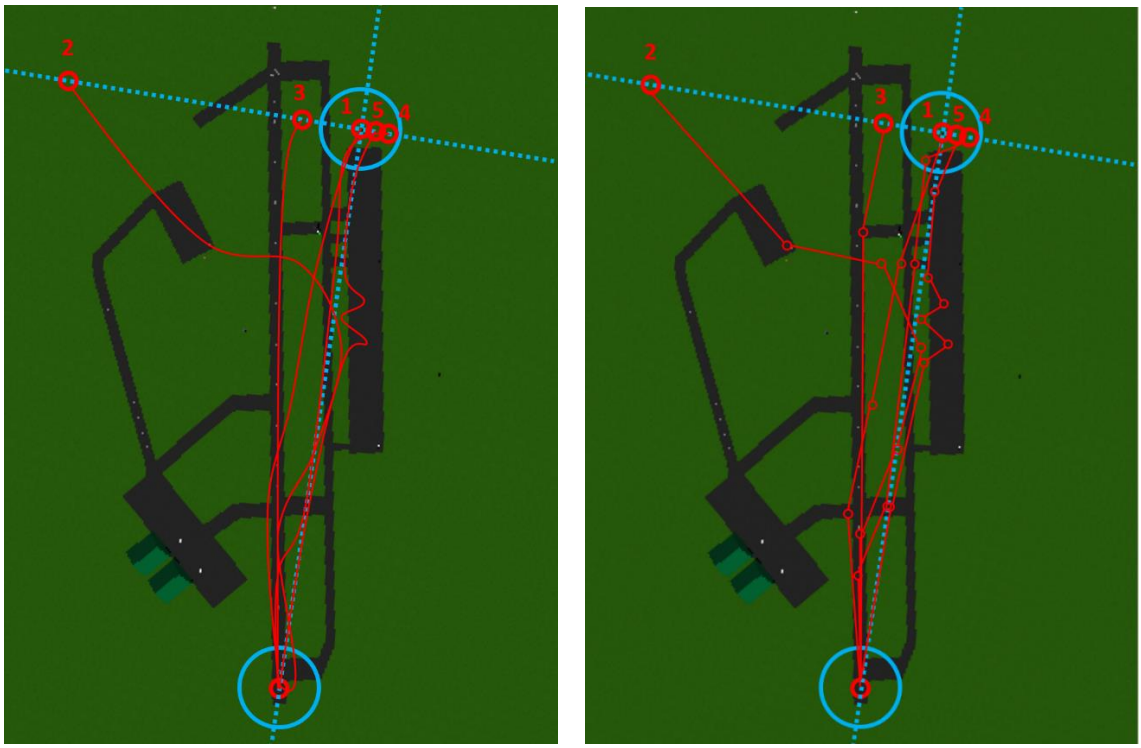

Figure 25: Original flight path data (left), Point connected flight path (right)

The second evaluation criterion mentioned above is explained in this section. In this evaluation criterion, the number of course changes in the trial will be used to judge the quality of the flights and thus the effectiveness of the CRM training. Specifically, the number of points or locations on the course where there is a change of course more than 10 degrees is counted to evaluate the stability of the flight. As an example, the actual data from the experiment is shown above. The figure on the left is the original data of the course. The figure on the right is an image with additional information of the points at which the course changed. In the image on the right, you can see that the stability of the flight is more pronounced, which was difficult to see in the left figure. In this evaluation framework, the number of these points are counted to evaluate stability.

4.4 - Distance from the goal

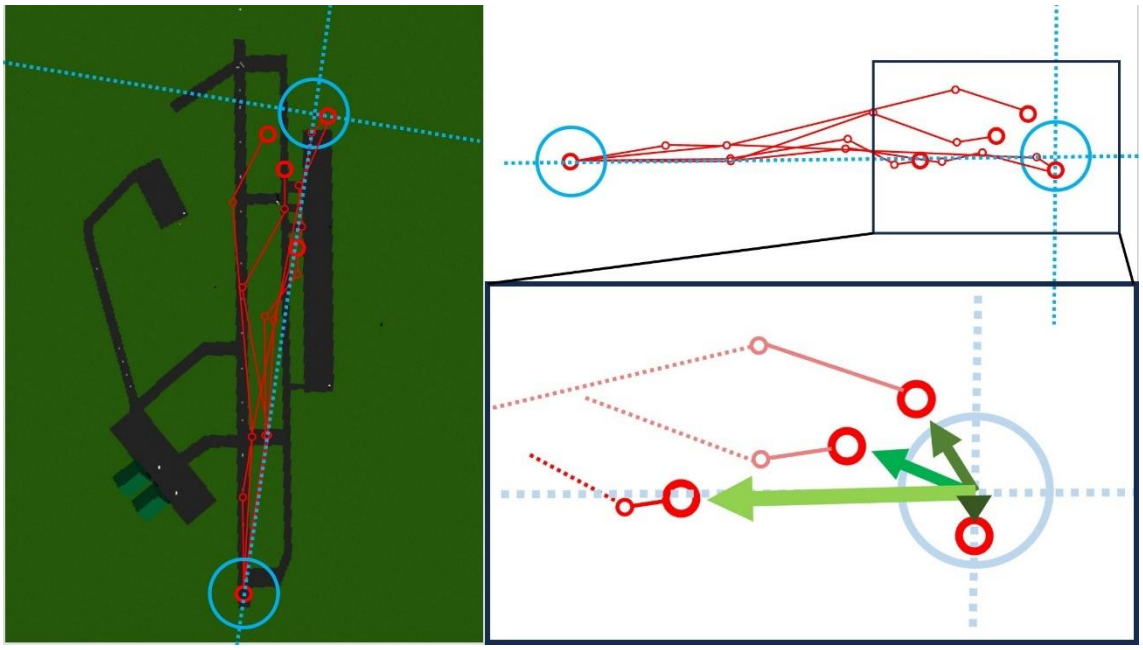

Figure 26: Measuring distance from the goal

In the next, the third evaluation criterion is described. Under this criterion, the distance from the destination at the end of the assignment is the subject of the evaluation. The reason to set this item is the same as the criteria regarding deviation from the course. The closer the point at the end of the experimental task is to the destination, the better. This criterion was established with this in mind. But that is not the only reason. Because the tasks in this study have other conditions for completion, other than the conditions such as reaching the destination or reaching a horizontal line. For example, if the aircraft crashes or becomes incapable of continuing flight, the assignment will end at that point. Therefore, if this criterion was not established to account for situation like this, there is a possibility that a flight will be rated as "good" even though it should have been rated as “bad”. This standard was established to prevent such situations.

Let`s explain with an example. The images shown above are experimental data from one subject. Many of the trajectories in this data are basically passing close to the ideal course, which at first glance appears to be stable flight. However, a closer look at the points where the experiment ends reveals that many of these points cease before the destination. In other words, he failed to manage his altitude and crashed mid-flight. These points cannot be detected by the two aforementioned criteria. Therefore, this

item was established as an evaluation criterion that can be applied to such cases.

4.5 - Control transition

Figure 27: Score given by control transition

In order to examine the effectiveness of CRM training in more detail, two methods other than the evaluation method using aircraft trajectories will be introduced next.
One is an evaluation based on the presence or absence of transitions in operating authority. The reason for using this method is explained below. To put it simply, the reason for adopting this method is that it can adequately detect whether authority gradients between subjects have been resolved. One of the main goals of CRM training is to mitigate authority gradients, and the experiments in this study reproduce this authority gradient artificially. If CRM training were effective enough, these authority gradients could be reduced. This methodology was set up to give an appropriate assessment in this regard.

The following section describes the methodology. As mentioned in the paper, the experiment in this study was conducted with one subject playing the role of the captain and one subject playing the role of the co-pilot. In this case, changes in task assignment in the aircraft operation are allowed regardless of the group to which the subject belongs. Thus, each subject can change his or her task at any point in the experiment by mutual agreement. However, in this experiment, subjects were given false information in advance. For example, long and highly trained members of Group A are informed that their skills are lower than those of Group B and that they have less

training time. Conversely, members of Group B with shorter and lower quality training are given information that they are more skilled and have more training time than the others. Off course, this is false information. However, because the subjects do not have any chance to notice about the facts, a false authority gradient is formed that the members of Group B have better skills than those of Group A. Therefore, it is expected that Group B members will take the initiative when communication through CRM does not work well. Therefore, a significant decrease in performance would be inevitable as a result of this situation. However, if the CRM training is effective and task management is successfully done among the subjects, their performance should exceed that of the control group. Based on this idea, this evaluation framework will target the evaluation of whether or not there is a transition of piloting operation. Based on the above, a score of "0.0" is given if the CRM does not work and a subject with low skill continues to take the initiative. Conversely, "0.5" is given when the task management is done appropriately, and each subject's workload is reduced by half. Finally, after appropriately judging his or her own skills, if member of group B entrusts the piloting of the aircraft to the Group A personnel, then "1.0" is given.

4.6 - Total words count

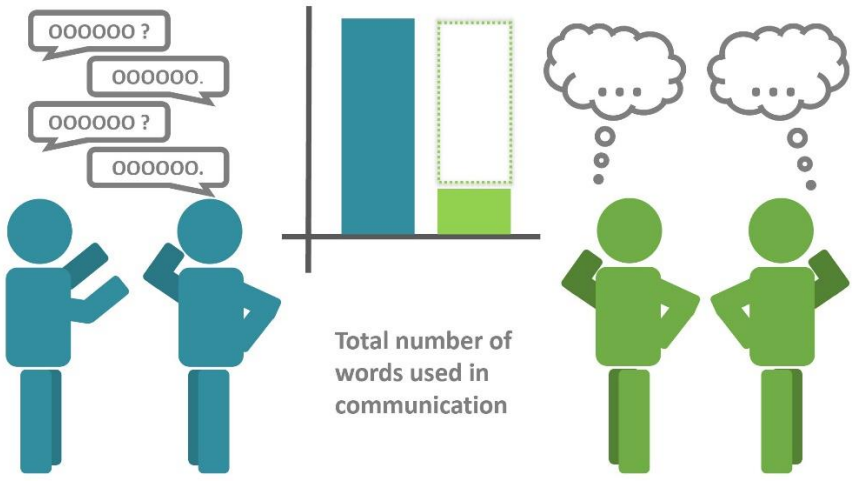

Figure 28: Total words count

In the previous section, we introduced evaluation methods using factors other than aircraft trajectory. Similarly, this section will introduce methods to supplement the information needed to measure the effectiveness of CRM training. This method counts the number of words in a conversation in communication, and the total number of

words in the experiment is used as the index.

The premise is that communication is essential to bring out all the elements included in CRM skills. Even if an individual's task management skills and Threat and Error Management skills are high, there are limits to one person's capabilities. Therefore, it is considered necessary to reach a consensus among the subjects to maximize performance using the resources of the two operating crew members. The means to achieve this is communication. Therefore, this evaluation method focuses on communication. The reason why the total number of words included in a conversation is used as an indicator is simply because the amount of conversation is proportional to the amount of information shared. Because the experiments in this study are issue-oriented, the conversations that take place in them are necessarily limited to two points: information sharing and consensus building. Therefore, the greater the volume of conversation, the better the communication regarding CRM is perceived to be. This is the reason why this evaluation method was adopted.

Chapter 5 - Results

This chapter describes the results obtained in the experiment. In order to facilitate understanding of the main points of the results, we will reiterate the outline of the experiment.

The purpose of the experiment in this study is to verify the framework constructed for CRM evaluation and is a comparative experiment using two groups. The two groups consisted of an experimental group, Group 1, with CRM training, and its comparison group, Group 2. At the same time, each group consists of two types of members: Group A with high piloting skills and Group B with low piloting skills. In the experiment, one member is selected from each of group A and B to form a team of two, but these two are given false information beforehand. This creates an artificial authority gradient so that the skills of members of group B, who actually have lower skills, appear to have higher skill than those of members of group A. In addition, the above pair is allowed to talk with and change the tasks they are responsible for during the experiment. Based on these conditions, the pairs will work on the experimental task. This is the outline of the experiment.

5.1 - Results of experiment: Course

In this section, the results obtained from the experiments are shown. In the following, we compare the results of Group 1 and Group 2 based on five evaluation axes: deviation in flight, stability, distance from the goal, number of operating authority transitions, and total number of words. Note that the assignment was performed 5 times per pair, so each group has 20 data. Thus, a total of 40 data were obtained in this experiment.

Results for the first three course-related data are presented below. From top to bottom, they are deviation from the course, number of course change points, and distance from the goal. Group 1 and Group 2 consist of four pairs of two persons. Therefore, in the graph below, this notation is expressed as follows.

Group 1：pair 1 ➡ 1.1　　Group2：pair 2 ➡ 2.2
Group 1：pair 4 ➡ 1.4　　Group2：pair 3 ➡ 2.3

Therefore, the numbers on the horizontal axis represent pairs of groups. The vertical axis of the graphs regarding deviation from the course and distance from the destination is plotted with the data the maximum value as 100.

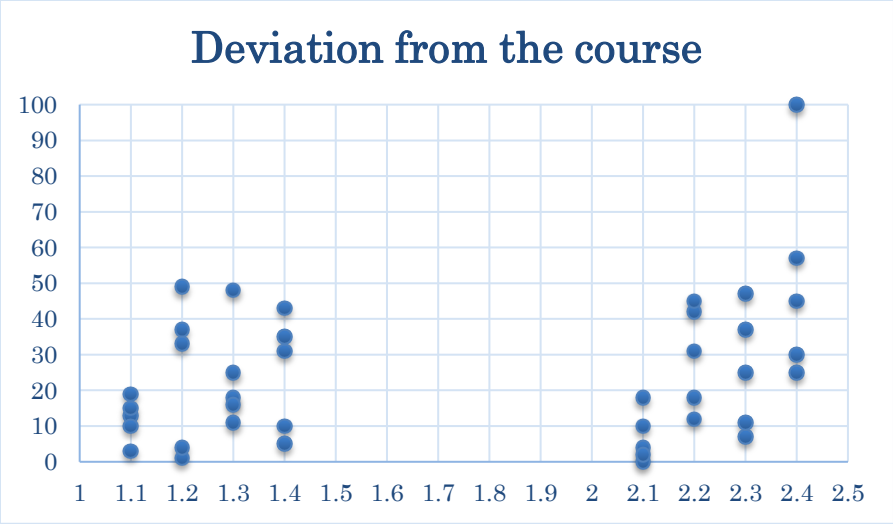

Figure 29: Deviation from the course

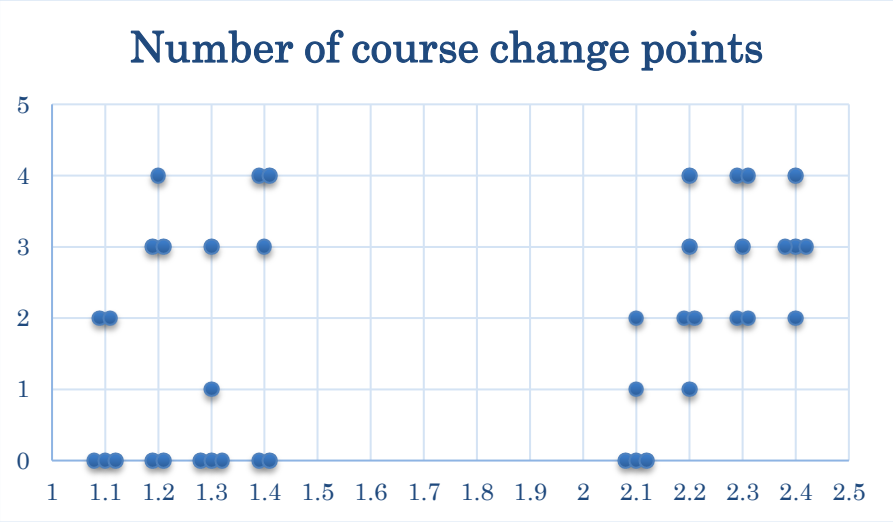

Figure 30: Number of course change points

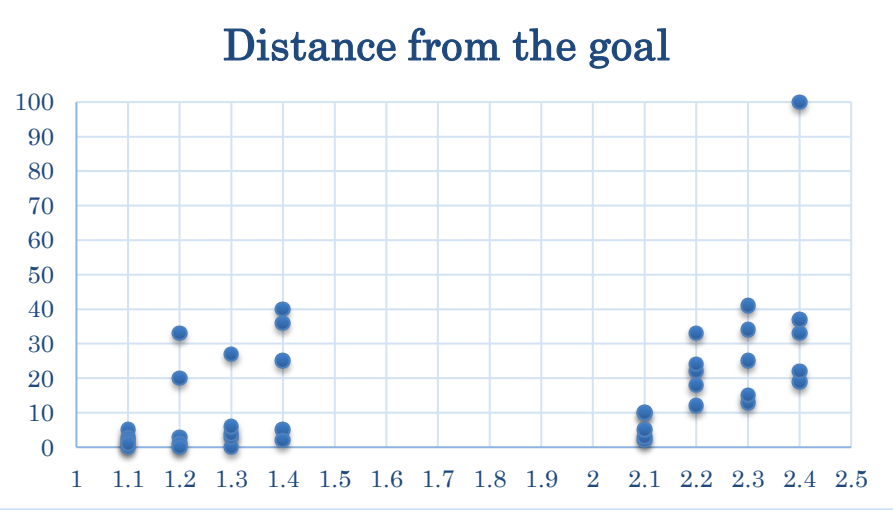

Figure 31: Distance from the goal

5.2 - Results of experiment: Control transition

The following graphs relate to evaluation criteria other than those related to the course. The first graph regarding the number of operation authority transitions is shown below. In this graph, the trials in each group were taken as average data, so they are not subdivided like the graph of courses as shown above. Therefore, we note here that the horizontal axis of the graph below is a number indicating the group number.

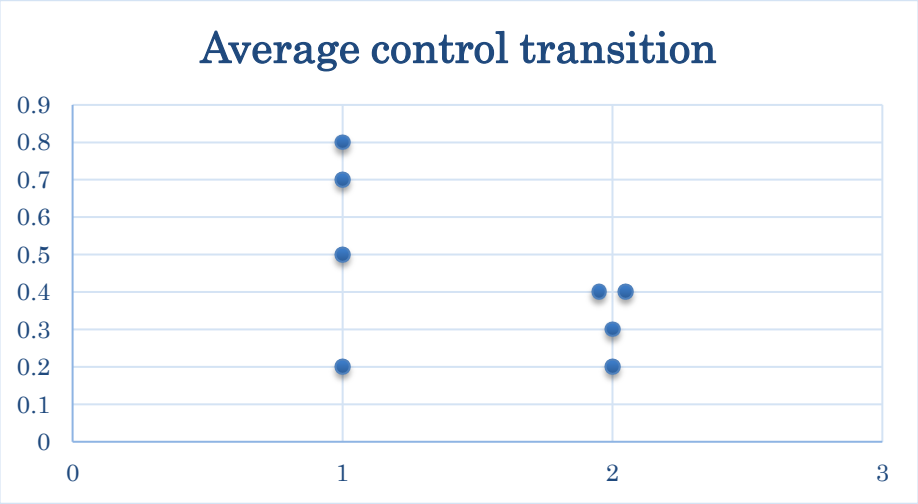

Figure 32: Average control transition

5.3 - Results of experiment: Total words count

Next, a graph regarding the total number of words is shown below. This graph is viewed in the same way as the above graph.

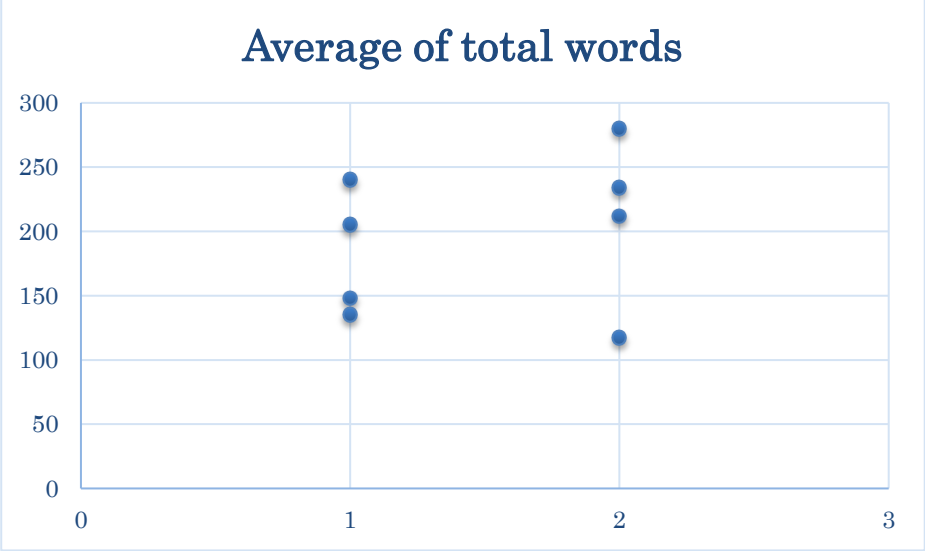

Figure 33: Average total words

Chapter 6 - Framework verification

The above chapter described the results of the experiment. This chapter examines the usefulness of a framework for evaluation of CRM skills. Specifically, we will verify whether we are detecting the fact that subjects' performance is improved by CRM training by comparing experimental results. The first result to be examined is the deviation from the course.

6.1 - Deviation from the course

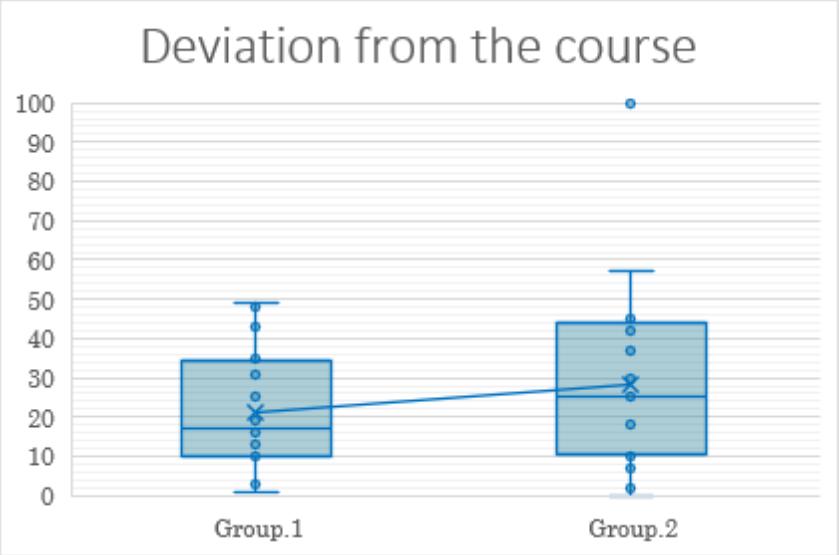

Figure 34: Deviation from the course (Box plot)

The figures shown above are box plots of the deviation from course in each of Groups 1 and 2. As can be seen from the figure, the average performance in Group 1 is lower than the average in Group 2. This means that CRM training worked effectively. Let us explain with concrete numbers. First, the mean for Group 1 is 21.3, while the mean for Group 2 is 28.3. These figures indicate that CRM training improved the performance of the experimental group by an average of 24.7% when compared to the control group. Therefore, one of the evaluation items in this framework, the axis of deviation from the course, is useful in detecting and evaluating the effectiveness of CRM training.

6.2 - Course change points

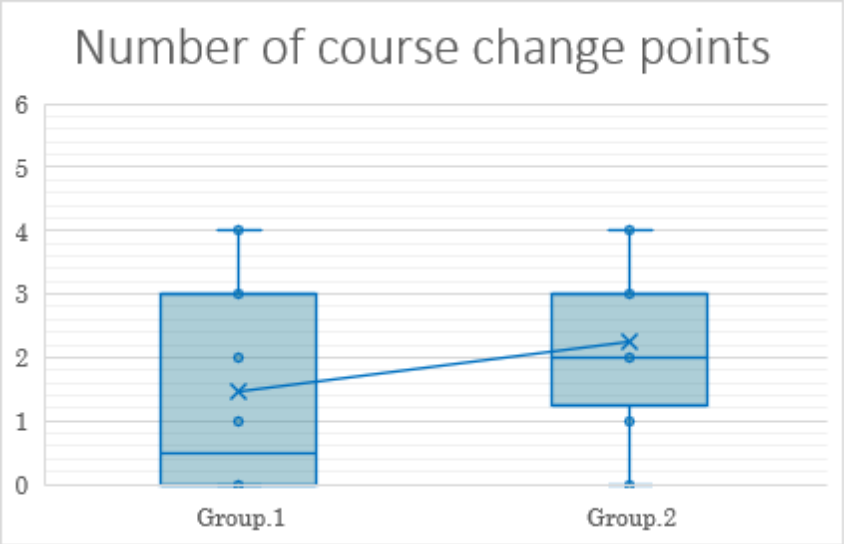

Figure 35: Number of course change points (Box plot)

In the next, we examine the experimental results in the number of course changes, an item included in this evaluation framework. Again, we will use box plots as well as the graph of deviation from course described earlier. The way of performance comparisons is same too. Group 1, the experimental group, is compared to Group 2, the control group. First, the average number of course changes for Group 1 is 1.45. In addition, the median is as low as 0.5. On the other hand, Group 2 recorded a mean value of 2.25 and a median value of 2.0, which is significantly different from the result of Group 1. Thus, it can be said that the performance of the experimental group greatly exceeded that of the control group. Based on specific figures, the performance of Group 1, the experimental group, in terms of the number of course changes, improved by an average of 35.6% from that of Group 2. Therefore, in this evaluation framework, the number of course changes is a useful guideline for detecting and evaluating the effectiveness of CRM training.

6.3 - Distance from the goal

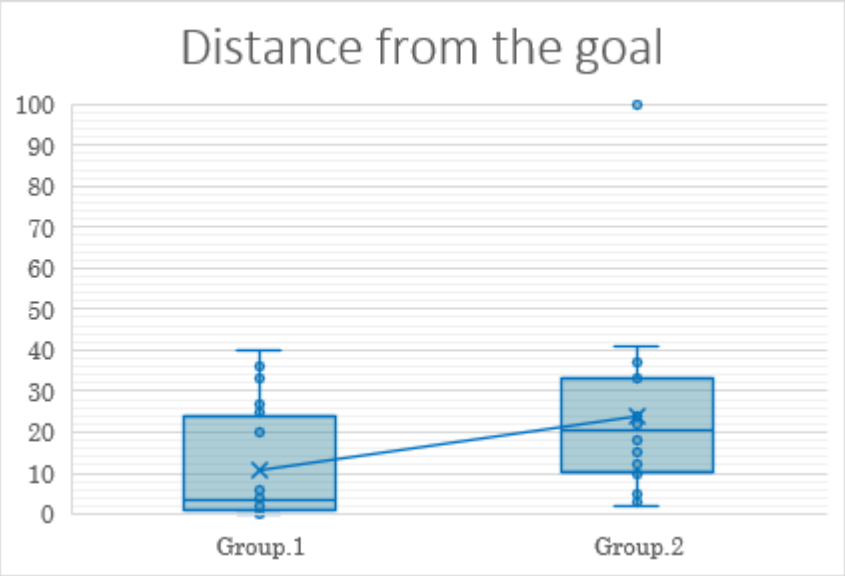

Figure 36: Distance from the goal (Box plot)

The results for the distance from the destination, which is the evaluation axis for the course, are then shown above. Since the contents of the graphs and the method of verifying the results are the same as in the previous two examples, only the results are described below. The mean value for Group 1, the experimental group, was 10.8, while the mean value for Group 2, the control group, was 23.9. Thus, the average improvement in performance in this item due to CRM training is 54.8%. From the facts written above, we believe that the three indicators related to courses, including this evaluation item, distance from the destination, are effective evaluation indicators for assessing and verifying the effectiveness of CRM training.

6.4 - Control transition

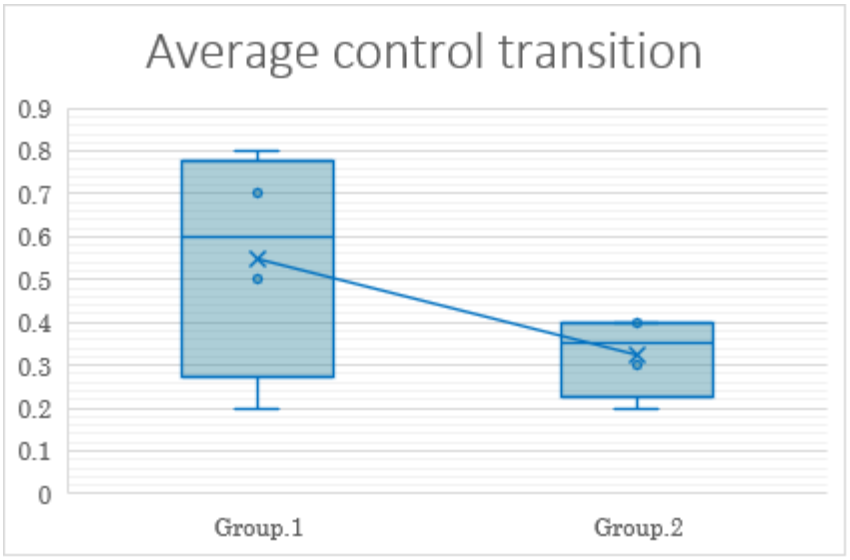

Figure 37: Average control transition (Box plot)

This section examines evaluation items other than those related to the course. It should be noted in advance that the two evaluation items introduced above and below are different from the evaluation axis for courses described earlier. In these figures, the higher index indicates the higher performance is. The first thing to be verified is the number of transitions in piloting operation. The mean value for Group 1 on this graph is 0.55, while the mean value for Group 2 is 0.33. Looking at the numbers alone, there does not seem to be much difference, but in fact the average value for Group 1 is about 1.7 times that of Group 2. Therefore, this item, when considered in conjunction with the course-related evaluation indicators presented above, is considered useful for evaluating the usefulness of CRM training.

6.5 - Total words count

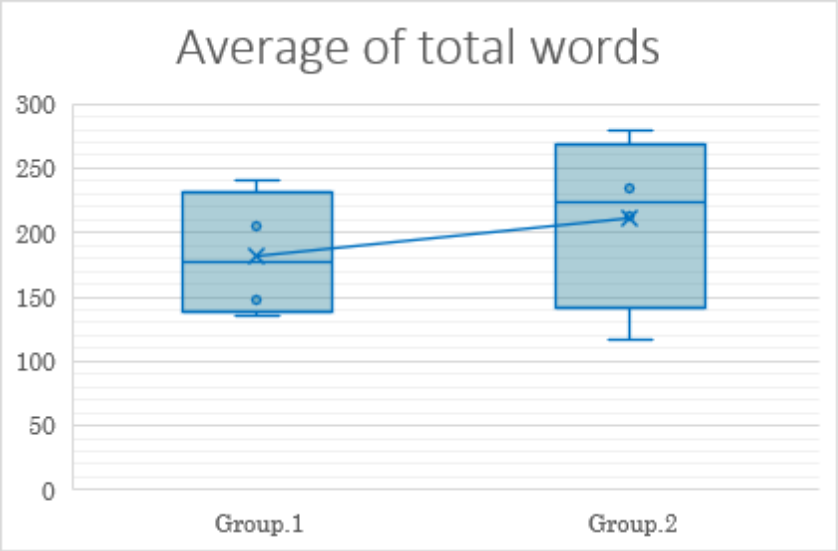

Figure 38: Average of total words (Box plot)

Finally, this section verifies the last item in evaluation framework. What we describe here is the total number of words used in the conversation during the experiment. In conclusion, this item did not produce results that clearly indicate the usefulness of CRM training. The following is a detailed explanation. First, this evaluation item is based on the assumption that performance in flight, such as information sharing and consensus building, increases as the number of words used in communication increases. However, the results of the experiment contradicted this assumption. Group 1, the experimental group, used an average of 182 words, while Group 2, the control group, used an average of 210.8. In this evaluation item, a high value indicates a high performance. Therefore, the fact that the results of the experimental group are 13.6% below the mean of the control group has implications that negate the usefulness of CRM training. However, in light of the validation results other than this evaluation items, it is clear that the CRM training itself was effective. This fact suggests that although this evaluation item was able to detect the impact of CRM training on the subjects, it is questionable to be used as a criterion for measuring its effectiveness.

6.6 - Overall Evaluation

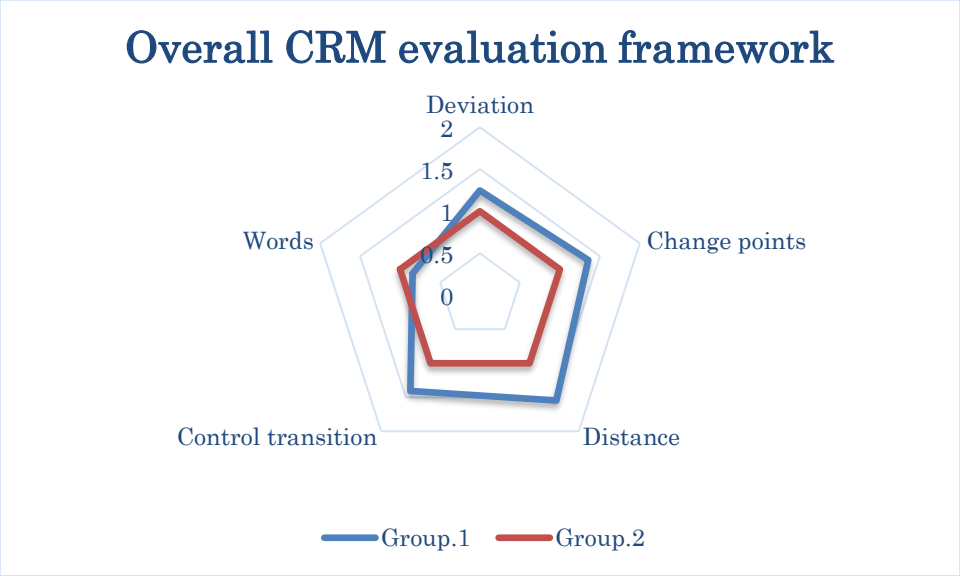

Figure 39: Overall Evaluation of framework

Judging from the results so far, the evaluation framework for CRM training developed in this study is generally considered to be useful, with the exception of some evaluation items. The above figure shows a radar chart integrating all previous evaluation items. In this figure, we plotted the normalized results of the ratings of Group 1, the experimental group, while all ratings of Group 2, the control group, were plotted on 1. Specifically, it shows the percentage of improvement in Group 1 compared with the performance of Group 2. For example, the lower the number, the better, because the indicators related to the course were based on negative criteria such as deviation. On the other hand, the higher the number of transitions in piloting operation and the number of words in communication, the better. In order to resolve the confliction on evaluation like this, we plotted the data of group 1 as a percentage of improvement compared to group 2. From this figure, it can be read that the framework clearly detects the impact of CRM training and is able to quantify that impact. However, some of the evaluation axis output results that are conflicting with other evaluation items while detecting the impact of CRM training. Therefore, while many of the axes in this evaluation framework are useful, there is still room for improvement. This is a conclusion of the verification.

	Average		Median	
	Group1	Group2	Group1	Group2
Deviation from the course	**21.3**	28.3	**17**	25
Course change points	**1.45**	2.25	**0.5**	2
Distance from the goal	**10.8**	23.9	**3.5**	20.5
Control transition	**0.55**	0.325	**0.6**	0.35
Total words count	182	**210.75**	176.5	**223**

Table 2: Average and Median of each evaluation categories

The conclusion written above can be seen from the numerical value obtained in the experiment. In the aforementioned section, only average numbers were focused on. It shall be enough to verify the effectiveness of the CRM training through the framework. However, it is also important to check the result from another perspective, the median. The table shown above is about the Average and Median of the experimental result of each group. The averages of experiment group perform better than control group as mentioned above. In addition to this, the results of median are consistent with the average. Moreover, the difference between these two groups is clearer than the averages. Judging from this result, we can see clear evidence of CRM training through the lens of this framework.

Chapter 7 - Discussion

In this study, we proposed a new 5-axis evaluation framework as a new evaluation method, replacing the previously proposed evaluation methods which have the problems of subjectivity and means to measure quantitatively. In addition, the results of comparative experiments to verify its practicality were consistent with predictions in large portion. The core idea in this evaluation method is that instead of focusing on changes in human behavior, it focuses on changes in an objective number: the course. It is fair to say that this part of the project was a success. The radar chart shown in the verification section of the experimental results clearly demonstrates this point. But one of the evaluation items produced result that were inconsistent with the other evaluation axes. This is especially true for the communication-related evaluation axes. Therefore, we are discussing about the reasons for the following two points.

1. Reason why the group of course-focused assessment items reflected the effectiveness of CRM training.
2. Reason why the evaluation based on the number of words in communication was different from what was expected.

7.1 - Successful part of this research

At first, we are discussing about reason why the course-focused assessment items reflected the effectiveness. In conclusion, it is possible that the reason for the success of this approach is that the task management training or Threat and Error Management training included in the CRM training worked well, and the results were reflected in the numerical values of the course.

Task management and Threat/Error Management training involves process of information sharing with each other in order to handle aircraft in various situations to prevent crisis situations as much as possible. This work includes a detailed process of discussion about the distribution of piloting tasks and also discussion for the conditions in case of taking over control. Thus, even before the start of the experiment, the participants in the CRM training were aware of the piloting task as "something that can be transferred and distributed among each other". On the other hand, it is just before the experiment that control group is taught that they can share the piloting task with each other or hand over the initiative. This difference in the two situations is

significant. This is because it is very difficult to discuss initiative regarding piloting without CRM training. Please recall the assumption that an artificial authority gradient is formed by false information given to the pair that is working together. If this information is received honestly, a person with high authority (perceived as high authority) possibly continues to take the initiative in piloting, while the other person remains in a support role. Although this is just a hypothetical story, it is quite possible that some subjects actually went through such a psychological situation. See the figure below.

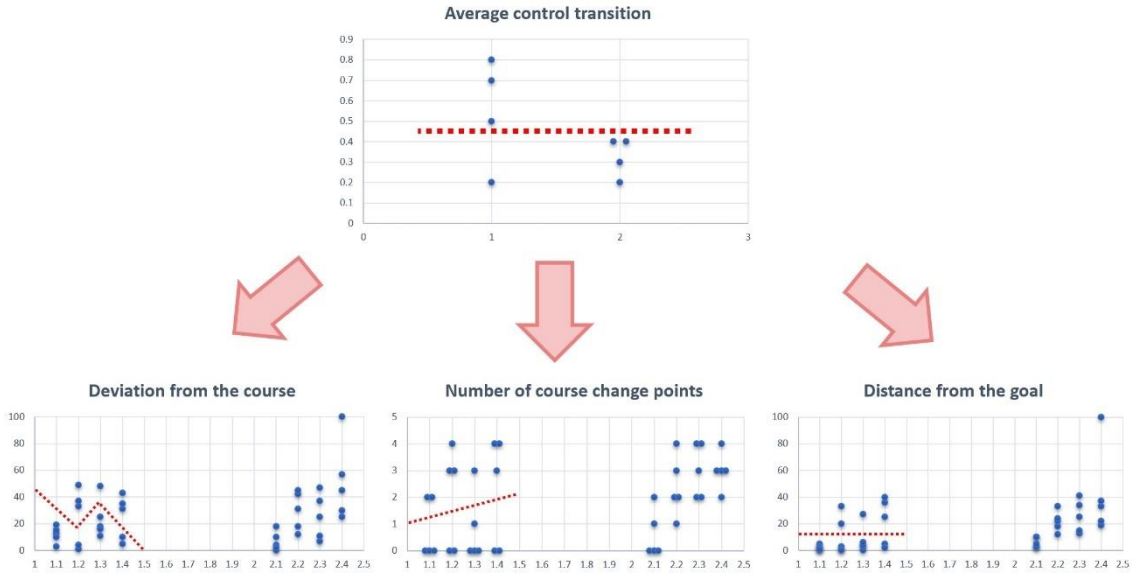

Figure 40: Results of experiment group caused by control transition

The above figure shows number of operational authority transitions and three course related data. As can be seen in this figure, 75% of Group 1, the experimental group, performed an average of 0.5 or more transfers of operating authority. Conversely, all controls are less than 0.5. In the other words, Group 1 at least distributed the piloting tasks and sometimes switches the initiative in maneuvering completely. The impact of this is evident in the course data. Deviation from the course, number of course change points, and distance. In all of these data, the Group 1 results contain polarized data: extremely good data and relatively bad data. This data is thought to be the result of switching initiative in piloting. This is because such a polarized difference does not exist in the data from Group 2, where there are fewer control transitions. Therefore, it is possible that the evaluation framework in this study was successful because the task management training or Threat and Error Management training included in the CRM

training worked well, and the results were reflected in the numerical value of the course.

7.2 - Unexpected results

In the next, we consider about the reasons why the evaluation based on the number of words in communication differed from what we expected. By nature, the word count, a measure of the amount of communication, should be proportional to the final performance measure in flight, the course outcome. This is because, as CRM is based on communication, communication cannot be avoided for CRM to be effective. However, in the data obtained, the amount of conversation in Group 1, the experimental group, was exceeded by that of Group 2, the control group. This does not make sense. So, there are two possibilities. One is the possibility that CRM training had the effect of reducing conversations between subjects. The second possibility is that there may have been some factor in the experimental setting of this study that increased the amount of conversation among members of Group 2, the control group that did not receive CRM training.

We will consider the former possibility at first. CRM training in this study includes three types of training: assertive action training, task management training, and threat/error management training. Of these, assertive action training is unlikely to be a factor in reducing the volume of conversation, since its premise is to foster the speaker's positive attitude in the first place. Therefore, if there is a cause, it would be task management training or Threat and Error Management training. Once again, we look back on each training contents to find the cause. Task management training is a scenario-based exercise in which participants make a list of necessary tasks and discuss how to divide them among themselves based on the scenario. Therefore, it is possible that the participants had already made a rough decision among themselves on how to divide up the tasks before the experiment began. If this idea were correct, the experiment would not require much conversation since it would only implement a procedure that has already been determined. This is also true for Threat and Error Management training. Since this training also shares similarities with task management training in that it involves discussing how to respond to situations, there is an undeniable risk that applying the above ideas could lead to a reduction in conversation. Therefore, it is possible that these two trainings may have caused a decrease in the amount of conversation.

However, there are contradictions in this idea. One is the results of the performance of the experimental group on the course. An artificial authority gradient should have been formed in the experiment regardless of Group 1 or Group 2. Therefore, if the task management method had been agreed upon in advance based on this information,

Group 1's performance on the course would have been equal to that of Group 2. But in practice, this was not the case. The results clearly show that Group 1 results outperform Group 2, as mentioned earlier. In addition, as noted above, the results of Group 1 are polarized. This implies a change in the way tasks are managed. In the other words, they changed the way they did to improve performance in the fluid process of flight experiment. Therefore, even if the management method was agreed upon prior to the experiment, it is essential to have a conversation when changing the management method. In light of this, the first possibility, that CRM training caused the decrease in conversations, seems unlikely.

Next, we are discussing about the second possibility. The possibility that there may have been some factor in the experimental setting of this study that increased the amount of conversation among members of Group 2, the control group that did not receive CRM training. There are two situations to be considered. One is if the selection of people in Group 2 was biased. The other is when one of the pairs speaks up more for support.

Let`s consider the first case. Group 2 has no factors to increase conversation as long as they have not received CRM training. But in this study, only four pairs of data were obtained, including eight in Group 2, which may have biased the selection of people. In other words, it is possible that the individuals who like the conversation were biased toward Group 2. If this were the case, it would be impossible to exclude the influence of individual personality from the experiment. Therefore, the only way to deal with this case would be to increase the amount of data to be sampled.

Figure 41: Gradual performance improvement of control group

Then, we are considering another case. It is a case of one of the pair speaking more for support. In the previous discussion, we noted that Group 1 significantly outperformed Group 2 in the number of transfers of piloting privileges, resulting in improved final performance. In other words, Group 2's performance was low because it could not switch operating privileges properly. However, even though Group 2 is

underperforming, this is only when compared to Group 1. When we take a look on the results of Group 2 on its own, one can see a gradual improvement in performance. Therefore, we should consider that piloting members of Group 2 received support in a different way than Group 1. This may have resulted as an increase in the number of statements and performance. Assumptions made in this experiment related to this possibility are described below.

・One pair consists of two people of varying skill levels.
・A false authority gradient is given before the experiment.
・In the experiment, a less skilled member played the role of captain and a highly skilled member played the role of co-pilot.
・Conversation is permitted during the experiment.

If we look at the experiment from the perspective of the person playing the co-pilot in light of these conditions, we can see why he chose conversation as a method of support. First, the person playing the co-pilot mistakenly believes that the person playing the captain is highly skilled. Therefore, he should not think of actively taking the initiative, even though he may support the operation. At the same time, in addition to those problems, there are relationship issues. If the person playing the role of co-pilot only assists in some of the operations, it can be presented as if the captain is the one in control. However, a proposal that takes complete control of the situation indirectly conveys that the other member is not good at piloting. This is because there is no consensus on the piloting task among the members of Group 2. If clear criteria for changing operations were established as well as the one in Group 1, the change of control could simply be viewed as a procedural task in accordance with the criteria. However, Group 2 has no such system. Therefore, the proposal to replace the pilot could hurt the captain's pride. For the co-pilot, such a situation should be avoided as much as possible. In fact, there is similar accident in the past, Korean Air Cargo Flight 8509 crashed because the co-pilot was afraid to point out captain`s mistake even though he clearly recognized the mistake. In light of these considerations, the means available to the person playing the role of co-pilot for support are very limited. As a result, it is likely that members of Group 2 chose verbal advice as a means of support. These are the possible reasons why the amount of conversation in Group 2 exceeded that of Group 1 in the experiment.

Chapter 8 - Conclusion

8.1 - Future challenges

As mentioned earlier, the 5-axis evaluation framework proposed in this study had some problems. Specifically, the total number of words item, which was employed as a measure of communication skills, did not function as expected. Therefore, it is possible that the method of counting the number of words in a conversation, as used in this study, was inappropriate as a method for measuring communication skills. Therefore, the challenge for future research on the evaluation of CRM skills will be to develop appropriate methods for measuring the quality of communication.

Below are some possible reasons why the total word count method did not work well in the previous discussion, and how to deal with them.

・Selection bias due to insufficient sample size

If this possibility is the main problem, an appropriate response would be to increase the number of samples in the experiment. As the sample size increases, the influence of factors such as individual personalities should lessen. Therefore, if bias in the selection of persons were a problem in this study, the best solution would be to expand the scale of the experiment.

・In case of the quantity of statements is out of proportion to the quality of CRM communication

The previous discussion indicated that even if the CRM training was working properly, there could be scenarios in which the control group's speech volume exceeded that of the experimental group. Therefore, we believe that the primitive method such as counting the number of words in a conversation cannot adequately measure the quality of communication in CRM skills. Therefore, it will be important to examine new and complementary methods to remedy these problems. One example would be a method of subdividing items related to communication. In the first place, this study attempted to measure communication with one simple indicator, even though the communication is a complex and varying system. This was most likely a mistake. Therefore, we believe that attention should be paid to the development of evaluation methods that focus on

the quality, rather than the quantity, of communication. For example, a frequency analysis of the words in the communication is worth considering. In the future, it is desirable to build a framework that incorporates new evaluation methods like this. These are the future challenges identified in this study.

8.2 - Summary

In the early part of this study, we introduced the current situation in the airline industry and explained the concept of CRM based on examples of accidents that have occurred in the past. At the same time, we described current and future risks by pointing out the lack of established assessment methods for CRM skills. Against this research background, this study proposed a new form of evaluation method for CRM evaluation and conducted a comparative experiment to verify its practicality. Specifically, an experimental group with CRM training and a control group for comparison were prepared, and the task was performed using a flight simulator. Although the results were generally favorable, some evaluation items showed different results than expected. Therefore, it is necessary to improve the evaluation method regarding the quality of communication that was problematic in this study. For this purpose, methods such as frequency analysis of words in conversations could be considered. We hope for further developments. This is a conclusion in this study.

References

[1] Joanna Bailey et al, “How flying today is safer than at any time in the past”, 2023
[2] NTSB aircraft accident report, “EASTERN AIR LINES, L-1011, N310EA”, 1972
[3] NTSB aircraft accident report, “UNITED AIRLINES, INC. MCDONNELL-DOUGLAS, DC-8-61, N8082U”, 1978
[4] Netherland Aviation Safety Board, “Investigation into the accident with the collision of KLM Flight 4805, Boeing747-206B, PH-BUF and PAN AMERICAN flight 1736, Boeing 747-121, N736PA at Tenerife Airport”, ICAO CIRCULAR 153-AN/56, 1977
[5] P. O'Connor, “Methods used to evaluate the effectiveness of CRM training: A literature review.”, Journal of Human Factors and Aerospace Safety, 2002
[6] R. L. Helmreich, “Cockpit management attitudes”, HUMAN FACTORS, 1984
[7] E. Salas et al, “Team Training in the Skies: Does Crew Resource Management (CRM) Training Work?”, The Journal of the Human Factors and Ergonomics Society 43(4):641-74, 2001
[8] P. O'Connor et al, “An evaluation of the effectiveness of the crew resource management programme in naval aviation”, Human Factors and Ergonomics, Vol. 1, No. 1, 2012
[9] R. W. Holt et al, “Evaluating Resource Management Training”, 2001
[10] B. Klampfer et al, “ENHANCING PERFORMANCE IN HIGH RISK ENVIRONMENTS: Recommendations for the use of Behavioural Markers”, Behavioral Markers Workshop, 2001
[11] J.A.G. van Avermaete, “NOTECHS: Non-technical skill evaluation in JAR-FCL”, NLR-TP-98518, 1998
[12] P. A. Ikomi et al, “JUMP SEAT OBSERVATIONS OF ADVANCED CREW RESOURCE MANAGEMENT (ACRM) EFFECTIVENESS”, 1999
[13] R. L. Helmreich, “System safety and threat and error management: The line operations safety audit (LOSA)”, 2001