




RESEARCH ARTICLE

Improving pilots' tactical decisions in air combat training using the critical decision method

H. Mansikka^{1,2}, K. Virtanen^{1,2}, T. Lipponen³ and D. Harris⁴

¹Department of Military Technology, National Defence University, Finland, ²Department of Mathematics and Systems Analysis, Systems Analysis Laboratory, Aalto University, Finland, ³Finnish Air Force, Finland and ⁴Faculty of Engineering, Environment and Computing, Coventry University, UK

Corresponding author: H. Mansikka; Email: heikki.mansikka@aalto.fi

Received: 17 May 2023; **Revised:** 26 November 2023; **Accepted:** 3 January 2024

Keywords: air combat training; critical decision method; recognition-primed decision making; situation awareness; tactical decision making

Abstract

In fighter pilot training, much of upgrade pilots' (UPs') learning takes place during mission debriefs. A debrief provides instructor pilots (IPs) the opportunity to correct situation awareness (SA) upon which the UPs base their tactical decisions. Unless the debrief is conducted with proper depth and breadth, the IPs' feedback on UPs' SA and tactical decision-making may be incomplete or false, resulting in poor, or even negative learning. In this study, a new debrief protocol based on the Critical Decision Method (CDM) is introduced. The protocol specifically addresses the SA of UPs. An evaluation was conducted to examine if a short CDM training programme to IPs would enhance their ability to provide performance feedback to UPs regarding their SA and tactical decision-making. The IPs were qualified flying instructors and the UPs were air force cadets completing their air combat training with BAe Hawk jet trainer aircraft. The impact of the training intervention was evaluated using Kirkpatrick's four-level model. The first three levels of evaluation (Reactions, Learning and Behaviour) focused on the IPs, whereas the fourth level (Results) focused on the UPs. The training intervention had a positive impact on the Reactions, Learning and debrief Behaviour of the IPs. In air combat training missions, the UPs whose debriefs were based on the CDM protocol, had superior SA and overall performance compared to a control group.

Nomenclature

ANOVA	ANalysis Of VAriance
CDM	Critical Decision Method
DP	decision point
F	Test statistics of ANOVA
IP	instructor pilot
LSD	least significant difference
LTM	long-term memory
M	mean
N	sample size
NDM	Naturalistic Decision Making
OODA	Orient, Observe, Decide, Act
OP	observer pilot
p	probability value for statistical tests
RPD	Recognition-Primed Decision Making
SA	situation awareness
SD	standard deviation
t	test statistic value in student's t-test
UP	upgrade pilot

1.0 Introduction

In air combat training, a great deal of instructor pilots' (IPs') work concerns evaluating the quality of upgrade pilots' (UPs') decision-making as they progress through fast jet training. Success in air combat is predominantly about making sound tactical decisions. Fast-paced air combat is highly challenging – even for pilots with extensive knowledge of its dynamics. Windows of opportunity open and close rapidly, and a single poor decision can quickly lead to a situation where no good decision alternatives are left.

Fighter pilots make decisions in a cycle, often described as the OODA (Orient, Observe, Decide, Act) loop [1]. An objective of air combat training is to increase the pace around this loop to achieve a dominant tactical tempo [2]. To maintain this tempo, pilots must make decisions swiftly, often with uncertain and incomplete information. The decision-making challenge is aggravated by limited ability to extract information from the tactical environment. Furthermore, as those engaged in air combat strive to hide or deceive their actual intentions, pilots' resulting awareness of the environment may deviate from its real state. As a result, tactical decisions may seem rational with respect to the subjective state of air combat, but less rational from the perspective of its objective state.

Air combat training is a form of Naturalistic Decision Making (NDM) training. NDM originated from the requirement to understand decision-making in a command-and-control context. It examines how subject matter experts use their experience to make decisions in an operational setting [3]. NDM is characterised by ill-structured decision-making problems [4], situated in uncertain, rapidly changing dynamic environments, where information is incomplete, imperfect, unreliable and/or ambiguous. Simon [5] argued that the information processing requirements underlying rational decision-making in such environments may exceed the limits of the information processing system of decision-makers. As a result, the decision-makers usually do not have complete knowledge of the consequences of all the possible alternatives. Moreover, when under time pressure, the decision-makers are likely to consider only a limited number of options and tend to adopt decision-making strategies using heuristics [6].

Early NDM studies suggested that people first try to understand the situation and establish if, from their experience, they had encountered something similar before. They use their experience to frame the situation and determine a potential course of action. They rarely consider all potential options. If a similar situation is not immediately recognised, they work sequentially until they generate a potentially satisfactory outcome. Klein et al. [7] observed that experts often pattern-matched the current situation with previous experience and determined a course of action without making a conscious decision. Klein labelled this as Recognition-Primed Decision making (RPD). RPD training approaches tend to be undertaken using simulations involving demonstration and practice with little emphasis on classroom-based instruction [8]. Zakay and Tsal [9] found that practice without stress and time pressure did not enhance decision-making when real-life decisions were made under similar conditions.

Air combat training involves all the main characteristics of NDM and RPD. The pilots' pre-existing knowledge, or mental models, about air combat is situated in their long-term memory (LTM). The content of LTM is relatively stable and is therefore referred to as static knowledge. When fighter pilots engage in air combat training, they sample the air combat environment and update their knowledge with new information. This process of information acquisition and updating, together with the short-term storage of the updated information, takes place in working memory. The resulting situation or dynamic knowledge is the pilots' subjective understanding of their tactical environment, or situation awareness (SA). Endsley [10, 11] argued that RPD requires the decision-maker to have a mental model of the decision-making environment based on past experience, but which is dynamically updated by current events.

Higher levels of SA enable decision-makers to act in a more timely and effective way. In Endsley's three-level concept, SA is "*the perception of the elements in the environment within a volume of time and space (SA-Level 1), the comprehension of their meaning (SA-Level 2), and the projection of their status in the near future (SA-Level 3)*" [10], p 36]. Unfortunately, SA and the real tactical environment are seldom perfectly aligned. Based on their SA, the pilots attempt to identify decision points (DPs). A DP is a state of the air combat environment from where the situation has potential to evolve in various directions

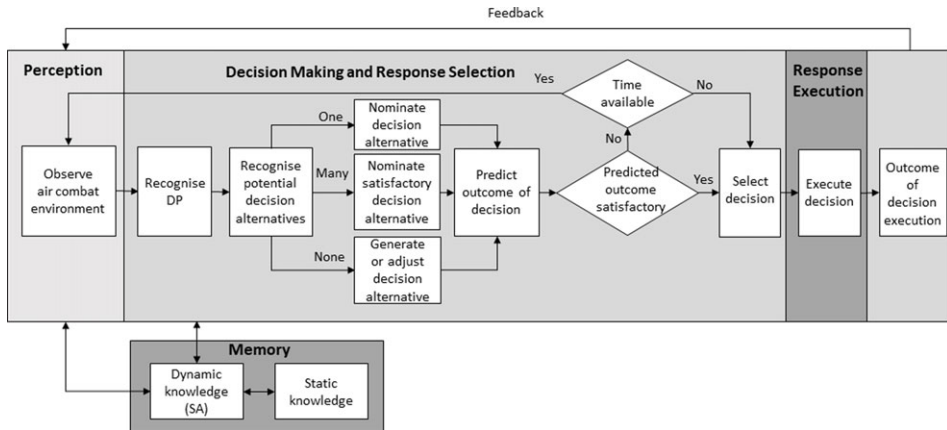


Figure 1. Pilot's decision-making and information processing in air combat.

and from which point the pilot's decision can affect how the situation will unfold. The pilots seldom have all the necessary information or time to evaluate the feasibility of all possible alternatives. Instead, they are often forced to make a rapid decision with whatever SA they have at the DP. When making a rapid decision, the pilot seeks to identify cues from the tactical environment associated with some learned decision alternative. These alternatives typically deal with the pilot's tactical and procedural options. If the identified cues are associated with just a single option, the pilot selects and executes it. If the cues are associated with several alternatives, the pilot is likely to select the first satisfactory option as it allows the fastest reaction to the tactical situation. If, however, they cannot associate the cues with any existing decision alternative, they must either adjust an existing alternative or create an ad-hoc solution. Both options are less desirable as they take time and tax the pilot's limited working memory capacity. Once the pilots have selected a decision alternative, they run a mental simulation to evaluate the expected outcome. If the expected outcome is not satisfactory and time permits, they will try to seek more information and repeat the cue-decision pairing process (see (12)). In all other cases, the pilot executes the selected decision. The executed decision changes the state of the air combat, which in turn creates a need to update SA. Such a fast-paced cyclical decision-making and information processing loop (see Fig. 1) is constantly repeated throughout the air combat mission.

In air combat, the outcome is what ultimately matters. Motivated by such a premise, an IP may put too much emphasis on the outcome of a UP's decision at the cost of omitting the analysis of the decision-making process [13]. A decision leading to a good outcome may be prematurely assessed as 'good' regardless of the rationale behind it. Similarly, an UP's decision resulting in an undesired outcome may be assessed as 'bad' irrespective of subjective reality at the time. Air combat decisions are made under uncertainty. Pilots can apply an appropriate process, but this may result in a bad outcome because of factors such as unanticipated actions of enemy aircraft. The quality of a decision cannot be solely assessed by its outcome [14]. Instead, the effectiveness of decision-making training should be gauged on underpinning processes [15–17].

When UPs' information processing is seen as a type of NDM, air combat training should emphasise recognition of patterns characteristic of domain-specific situations, i.e. RPD. To do this, IPs can use a Critical Decision Method (CDM) procedure to assist UPs to verbalise the static and dynamic knowledge underlying their decisions. CDM is a widely used semi-structured interview procedure for eliciting knowledge underlying RPD-type decisions [18, 19]. Klein [20] outlined four requirements for RPD training: engage in deliberate practice; develop a wide range of experience; obtain accurate, diagnostic and timely feedback; and review prior experiences to derive new insights and learn from mistakes. The role of IPs is critical, especially in the last two requirements. To improve UPs' decision-making by providing

Table 1. *Performance feedback judgements with respect to desired and undesired outcomes of decisions when using traditional and CDM-type performance feedback protocols*

Outcome	Traditional evaluation	CDM-type evaluation	
		Good knowledge	Poor knowledge
Desired	Good	Good	Lucky
Undesired	Bad	Unlucky	Bad

enhanced process-related performance feedback, the assessment of their decision quality should be more than simply ‘good’ or ‘bad’. Unfortunately, this has traditionally not been the standard protocol.

The quality of UP’s decisions should be evaluated against their static and dynamic knowledge relevant to the DP in question. The role of IPs in this evaluation is essential. If a pilot’s decision leading to a desired outcome is based on accurate and correct knowledge, in terms of decision quality the pilot should be considered as ‘good’. However, if the pilot’s underlying knowledge appears to be inaccurate or incorrect but s/he still achieves a desired outcome, the pilot should be regarded as ‘lucky’. If the pilot’s underlying knowledge is not evaluated at all, his/her decision-making is seen as ‘good’ in both cases. However, if the pilot’s decision leads to an undesired outcome but the decision is based on accurate and correct knowledge, the decision is simply ‘unlucky’. But if the same decision is founded on an inaccurate and incorrect knowledge, it would be reasonable to consider the pilot’s decision-making as ‘bad’. If the pilot’s knowledge is not assessed, his/her decision would be regarded as ‘bad’ in both cases. A failure to associate correctly the outcome resulting from a decision with the knowledge leading to it can have critical, and sometimes fatal, consequences in pilot training. An ‘unlucky’ pilot incorrectly assessed as ‘bad’ is likely to question what is actually correct knowledge, resulting in ineffective or even negative training. In contrast, a ‘lucky’ pilot assessed as ‘good’ is actually encouraged to strengthen his or her false representation of air combat, resulting in negative training and potentially developing inappropriate or incorrect mental models. Simply asking the UPs whether they had correct and accurate static knowledge or SA when they made a decision is not a fruitful approach. A person with a stable – but inaccurate and incorrect – SA is an unreliable assessor of his/her own SA [21]. Table 1 summarises the alternative judgements provided in the performance feedback based on either traditional or CDM-type feedback protocols.

Timely and accurate feedback has long been recognised as an essential factor for improving performance (e.g. [22]). Instructing modern air combat is primarily about building UPs’ accurate and detailed mental models of different states of air combat to support effective tactical decision-making. During mission briefs, it is the role of IPs to prepare UPs for upcoming air combat training missions by providing them with the most accurate mental models possible. During the mission, the UPs’ mental models and decision-making are tested. A great deal of learning, in terms of development of the mental models, takes place during mission debriefs, where the IPs evaluate the UPs’ decisions and give feedback on the mental models underlying those decisions.

This paper aims at improving the IPs’ ability to assess UPs’ decision quality by identifying the difference between the UPs’ subjectively perceived state and the objective state of air combat. The objective of the training programme described is to improve the IPs’ ability to identify and correctly classify the knowledge upon which the UPs base their tactical decisions during air combat training, enabling the IPs to provide enhanced feedback. The CDM-based training intervention received by the IPs is evaluated using Kirkpatrick’s four-level evaluation model [23]. At Level 1 (Reactions), the training content, materials, instructors and methods of delivery are evaluated. These evaluations help to develop the training and promote uptake. Positive reactions do not, however, suggest that the training has been effective [23]. Level 2 (Learning) typically comprises an end-of-course test to determine what knowledge and skills have been acquired. Learning does not guarantee that practices from the training will migrate to the taskwork itself, which is a fundamental objective. This is addressed at Level 3 (Behaviour), which

evaluates such transfer of Learning. Kirkpatrick & Kirkpatrick [23] suggested that evaluation at this level is important but often omitted in the assessment of training. Level 4 (Results) is concerned with organisational changes resulting from the training. It has been described as the most important and most challenging level to evaluate [24].

In this study, UPs were pilots transitioning from the BAe Hawk, an advanced jet trainer aircraft, onto fast jets such as the Lockheed Martin F-35 or Boeing F/A-18. A training intervention was developed to enhance the quality of UPs' decision-making. It was assumed that after the training the IPs would be better positioned to provide detailed and accurate feedback, ultimately resulting in UPs' better tactical decisions in air combat. A long-term field test was implemented for this purpose.

2.0 Method

The behaviour of a group of active-duty IPs using traditional performance feedback was first observed during debriefs. The objective was to evaluate the IPs' ability to identify DPs and to assess UPs' knowledge underlying their decisions. The IPs were then subject to a training intervention where they were briefed about RPD. In addition, the IPs were trained to use CDM to elicit UPs' static knowledge and SA and to use the CDM performance feedback protocol during air combat training debriefs. Four months after the training intervention, the observations were repeated.

The IPs' evaluation of the content and presentation of the training programme was assessed at the Reactions level of Kirkpatrick's model. The Learning-level evaluation examined changes in the IPs' knowledge about CDM resulting from their training. At the Behaviour-level, the IPs' debrief behaviour was compared pre- and post-training. The ultimate test of the effectiveness of the training intervention occurs at the organisational level. The Results-level impact of the IPs' CDM training was evaluated by comparing flying performance and SA scores of the UPs during their air combat training missions before and after the IPs' training.

2.1 Participants

Twenty-six BAe Hawk IPs volunteered to take part in the study. Their mean age was 37.9 years ($SD = 4.6$) and their mean flying experience was 1247.9 flight hours ($SD = 655.6$).

The IPs instructed two groups of UPs. These consisted of pilots from two back-to-back Finnish Air Force Academy flying courses. There were 12 pilots in the first group and 13 in the second. The flying experience of the UPs in both groups was similar: the first group had an average of 155.3 flight hours ($SD = 3.1$) and the second group had an average of 159.3 flight hours ($SD = 4.9$). The mean age of the UPs in the first group was 24.4 years ($SD = 0.5$) and 24.9 years ($SD = 1.2$) in the second group.

Four active-duty BAe Hawk IPs were recruited and trained as observer pilots (OPs). The OPs' task was to observe the IPs' behaviour as they debriefed the air combat training missions of the UPs. The OPs were responsible for the primary collection of debrief data relating to critical DPs. The mean age of the OPs was 33 years ($SD = 0.8$) and their mean flying experience was 530.3 flight hours ($SD = 113.9$). The OPs were somewhat less experienced than most of the IPs. However, at the time of the study, the OPs were among the group of instructors who were most intensively involved in instructing duties. All participants were male. Written informed consent was obtained from all OPs and IPs. The UPs participated in the study as a part of their normal BAe Hawk air combat flying curriculum.

2.2 Procedure

2.2.1 Training of OPs

The OPs were trained to identify DPs during IPs' debriefs. A DP was defined as an in-flight event where the UP had made (or should have made) a critical tactical decision. In addition, the OPs were trained to observe and log IPs' knowledge elicitation behaviours related to UPs' decisions concerning DPs.

Table 2. Summary of IP behaviours and levels of knowledge elicitation related to DPs

DP	IP behaviour	Level of knowledge elicitation
Not identified	None	None
Identified	IP does not elicit the UP's knowledge associated with the decision	None
	IP elicits the UP's static knowledge associated with the decision	Static knowledge
	IP elicits the UP's SA Level 1 knowledge associated with the decision	SA Level 1
	IP elicits the UP's SA Level 2 knowledge associated with the decision	SA Level 2
	IP elicits the UP's SA Level 3 knowledge associated with the decision	SA Level 3

As a part of their training, the OPs observed actual air combat training debriefs and practiced identifying and logging the knowledge elicitation behaviours of the IPs. After each debrief, the OPs' logs were discussed with subject matter experts (SMEs), who were experienced instructor pilots. This continued until the OPs were able to reach consistently similar observations with the SMEs. The OPs' training was completed before the air combat training of the first UP group commenced.

2.2.2 Observation of IPs' debriefs

Once the OPs had completed their training, they observed and categorised the elicitation behaviours of IPs as they conducted the first UP group's debriefs. At this stage, the IPs had not yet been exposed to the CDM training intervention. These IPs were considered as 'non-trained'.

If the IPs identified a DP, they may or may not have attempted to elicit the UP's underlying knowledge upon which their decision was based. The thoroughness, or 'depth', of the non-trained IP's knowledge elicitation varied from simply finding out what kind of DP-related static knowledge the UP possessed up to the elicitation of the UP's SA Level 3 (see Table 2). Data were collected over a period of seven months, during which time a total of 41 debriefs were observed.

2.2.3 IPs' CDM training intervention

The CDM interview has five phases: (1) select incident, (2) obtain unstructured incident account, (3) construct incident timeline, (4) DP identification, and (5) DP probing [19]. As a full CDM interview can be time consuming, a shortened version of CDM is sometimes more practical [18, 25]. Such a shortened version has previously been successfully used to elicit pilots' SA in air combat debriefs [26] and was also used in this study. When the CDM interview is undertaken in an air combat debrief, a computer-generated reconstruction of the air combat mission is reviewed. This reconstruction represents an objective reality, often referred to as ground truth, of a mission against which pilots can assess the correctness and accuracy of their subjective knowledge. During a review, an IP identifies a DP. As the mission reconstruction is available to all participants, there is normally no need to construct a timeline of events. Before deepening probes are used, the IP asks the UP to compare his/her SA with the ground truth. With access to the ground truth, pilots are generally of verbalising the correctness and accuracy of their relevant knowledge related to DPs [26]. Deepening probes and 'what if' questions used in Phase 5 are typically only needed to elicit how possible contingencies affected, or might have affected, pilots' decision rationale.

After the air combat training of the first UP group was complete, the SMEs provided the CDM-based training to all IPs. The content of the training included the following topics: human information

Table 3. *IPs' Reactions towards the training intervention (1 - strongly disagree to 5 - strongly agree), N = 22*

	<i>M</i>	<i>SD</i>
General arrangements of the training intervention were successful	4.77	0.43
Content of the training intervention was interesting	4.86	0.35
I learned something new during the training intervention	4.55	0.51
The training intervention was useful	4.64	0.49
I would recommend this training intervention to other instructor pilots	5.00	0.00
Overall, I am satisfied with the training intervention	4.86	0.35

processing in general; human information processing in the context of air combat; SA in general; SA in the context of air combat decision-making; DPs in air combat; identification of DPs during air combat debriefs; mental models in general; mental models in air combat decision-making; basic principles of CDM; and application of CDM in air combat debriefs.

After the training intervention, the IPs were asked for their Reactions towards the training. The IPs were given six statements regarding the content and delivery of the training intervention (see Table 3 for statements). The IPs answered using a five-point Likert scale, from 1 'strongly disagree' to 5 'strongly agree'.

Once the IPs had expressed their Reactions towards the training, they were asked to provide a proximal course evaluation. The IPs were asked how much they had learned about decision-making, SA, DPs, mental models and CDM. Two questions were asked about each topic (see Table 4). The IPs again provided their answers on a five-point Likert scale, where a score of 1 referred to 'not at all' and a score of 5 referred to 'very much'. The evaluation was repeated four months after the training intervention to provide a distal evaluation of the course. A pre-training evaluation of IPs' knowledge was not used. While control groups can provide more rigorous Learning evaluation, their use is often not practical [27] and can also sensitise the sample. It was also known that the knowledge about CDM was new to the IPs. Therefore, there was no need for pre-training evaluation [23].

2.2.4 Observations of trained IPs' debriefs

As the second group of UPs started their air combat training, the observation procedure carried out during the first group's air combat training was repeated. The IPs who delivered the second group's debriefs were now trained in the CDM approach. As before, the OPs attended the trained IPs' air combat debriefs where they observed and logged the IPs' knowledge elicitation behaviours. The second set of observations was carried out over a period of seven months, during which time a total of 56 debriefs were observed.

2.2.5 UPs' flying performance and SA

A formal procedure was in place to score the UPs' flying performance in every air combat mission. Some aspects of scoring varied from mission to mission, while some aspects remained the same. Overall performance and SA were among aspects scored for each mission. These scores were extracted from both UP groups to provide a Results measure.

3.0 Results

3.1 Analysis

The data were analysed to assess the effectiveness of the IPs' training intervention following the criteria suggested by Kirkpatrick. The IPs' Reactions towards the training were evaluated immediately after the training, followed by an end-of-course test to assess Learning. The IPs also provided a proximal (end of

Table 4. *Descriptive statistics of the IPs' proximal and distal evaluation scores for the training intervention (1 – strongly disagree to 5 – strongly agree)*

	Evaluation question	Proximal evaluation (N = 26)		Distal evaluation (N = 22)	
		M	SD	M	SD
1	How much did the training intervention increase your knowledge about the human decision-making process in general?	3.09	0.68	3.14	0.89
2	How much did the training intervention increase your knowledge about the human decision-making process in air combat context?	3.14	0.77	3.32	0.72
3	How much did the training intervention increase your knowledge about the role of SA in air combat in general?	2.73	0.77	2.68	0.57
4	How much did the training intervention increase your knowledge about the role of SA in air combat decision-making?	2.91	1.02	2.86	0.71
5	How much did the training intervention increase your knowledge about the decision points in air combat?	2.95	1.05	3.27	0.77
6	How much did the training intervention increase your knowledge about how the decision points relate to learning during debriefs?	3.50	1.10	3.50	0.74
7	How much did the training intervention increase your knowledge about the role of mental models in air combat?	3.23	1.03	3.18	1.01
8	How much did the training intervention increase your knowledge about the mental models in air combat decision-making?	3.18	1.01	3.23	0.81
9	How much did the training intervention increase your knowledge about the possible application of the Critical Decision Method during air combat debriefs?	4.00	0.93	3.91	0.68
10	How much did the training intervention increase your knowledge about the Critical Decision Method in general?	4.32	0.72	3.95	0.65

training) and distal (four months later) evaluation, collecting their opinions about how much they had learned from the intervention. The distal evaluation enabled the IPs to re-evaluate their learning after putting the content of the training into action.

The number of DPs was compared between trained and non-trained IPs, to assess if more DPs were subsequently identified. A similar comparison was undertaken to examine differences in the knowledge elicitation behaviours. A behaviour aimed at revealing the UPs' Level 3 SA relating to a DP was considered as the most in-depth form of knowledge elicitation. A situation where a DP was identified but the UPs' knowledge was not elicited at all was not considered as a knowledge elicitation. It was hypothesised that trained IPs would be able to perform more in-depth knowledge elicitation during debriefs than non-trained IPs.

The behaviour change of the IPs was considered to be the ultimate organisational result. This was evaluated from UPs' overall performance and SA scores. It was expected that the UP group instructed by the trained IPs would perform better than the group instructed by the same, but untrained, IPs.

Table 5. *P-values of the Tukey LSD post-hoc tests for the pairwise comparisons of the evaluation scores of individual survey questions*

	2	3	4	5	6	7	8	9	10
1	0.261	<0.001	0.086	1.000	0.013	0.576	0.598	<0.001	<0.001
2		<0.001	<0.001	0.365	0.062	0.862	0.853	<0.001	<0.001
3			0.104	0.009	<0.001	0.003	0.003	<0.001	<0.001
4				0.057	0.001	0.059	0.050	<0.001	<0.001
5					0.008	0.611	0.529	<0.001	<0.001
6						0.061	0.029	0.005	<0.001
7							1.000	<0.001	<0.001
8								<0.001	<0.001
9									0.042

Note: Significant differences are in bold. Numbers 2–10 in the first row and numbers 1–10 in the first column refer to the question number in Table 4.

Table 6. *Descriptive statistics for the number of DPs, CDM ratio and types of elicitation behaviours (with corresponding paired t-test results) with respect non-trained (N = 41) and trained (N = 56) IPs*

	Before training		After training		<i>t</i> (40)	<i>P</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
DP	6.46	2.35	6.20	2.02	0.57	0.5769
Static knowledge	2.39	1.79	3.90	2.92	−3.14	<0.05
SA Level 1	2.63	1.59	5.54	2.10	−6.72	<0.001
SA Level 2	0.80	0.90	2.95	1.88	−6.61	<0.001
SA Level 3	0.00	0.00	0.27	0.55	−3.13	<0.005
CDM ratio	0.95	0.46	2.05	0.45	−12.63	<0.001

3.2 Reaction measures

Table 3 summarises the descriptive statistics of IPs’ Reactions towards the training intervention.

3.3 Learning measures

Table 4 describes the IPs’ proximal and distal evaluation scores of the training intervention. Four IPs did not provide responses for the distal evaluation. A nested two-level repeated measures ANOVA showed no statistically significant difference between the overall proximal and distal evaluation scores ($F_{1,21} = 0.00$; $p > 0.05$). However, there were some statistically significant differences between responses to individual survey questions ($F_{9,13} = 9.99$; $p < 0.001$). Post hoc Tukey LSD tests showed a number of significant differences across the questions (see Table 5).

3.4 Behaviour measures

Forty-one debriefs by the IPs before training and 56 debriefs after training were observed. Table 6 presents the number of DPs identified. Additionally, the Table 6 illustrates the number of times debriefs included IPs’ elicitation behaviours targeting UPs’ static knowledge and SA Levels 1–3 with respect to the identified DPs. Table 6 also includes a CDM ratio, calculated by dividing the number of observed elicitation behaviours by the number of identified DPs. The CDM ratio reflects the ‘depth’ of the IPs’ knowledge elicitation efforts. Finally, Table 6 provides a summary of the t-test results for the number of identified DPs, the number of elicitation behaviours, and the CDM ratios before and after the training.

Table 7. Descriptive statistics of the SA and overall performance scores of the first (trained) and second (untrained) UP group. *M* = mean and *SD* = standard deviation

	Before IP training (<i>N</i> = 372)		After IP training (<i>N</i> = 402)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
SA	3.76	0.63	3.88	0.60
Overall performance	4.04	0.77	4.16	0.79

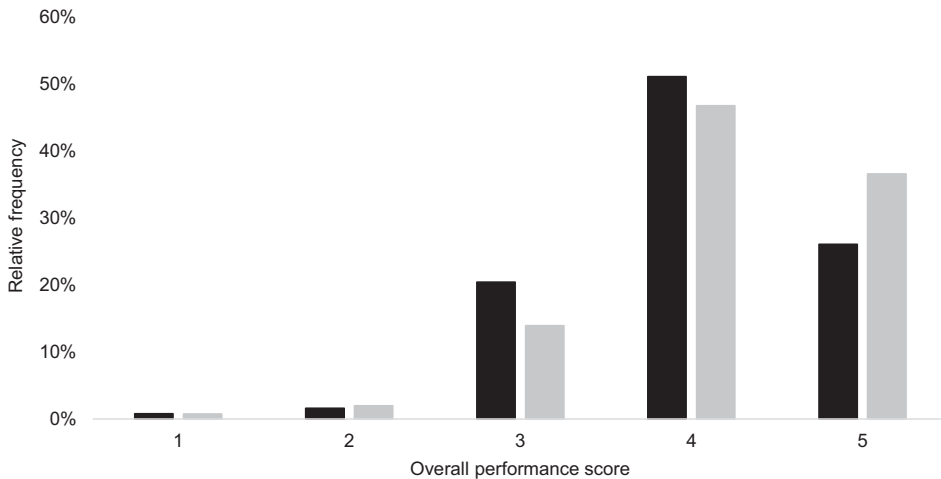


Figure 2. Relative frequency distribution of overall performance scores (1 = low, 5 = high) of the first (*N* = 372) and second (*N* = 402) group. The scores of the first group are marked with black bars and the scores of the second group with grey bars.

3.5 Results measures

The first UP group (prior to CDM training) flew a total of 372 air combat training missions and the second group (after CDM training) 402 missions. Table 7 summarises the overall performance and SA scores of both groups. Both SA and overall performance were scored on a scale 1 (low) to 5 (high). An independent samples t-test showed that the means of both the trained group's SA and overall performance scores were significantly higher, $t(772) = -2.603$, $p < 0.05$, $d = 0.20$ and $t(772) = -2.261$, $p < 0.05$, $d = 0.20$, respectively.

In the first, untrained, IP group, 26% of the pilots were given an overall performance score of 5. In the second, trained group, the percentage was 37. Figure 2 summarises the relative frequency of the overall performance scores for both groups. A similar trend was observed for SA scores. In the first group, 8% of the pilots were given a SA score of 5, whereas in the second group the percentage was 10. If both the SA scores of 4 and 5 are considered, the percentages are 70% and 80%, respectively. Figure 3 summarises the relative frequency of the SA scores for both groups.

4.0 Discussion

Endsley [10] and [11] argued that as a first step, Recognition-Primed Decision-making (RPD) requires decision-makers to have an accurate mental model upon which to build their SA. A long-term field test was conducted to determine if a CDM training intervention could enhance IPs' ability to provide

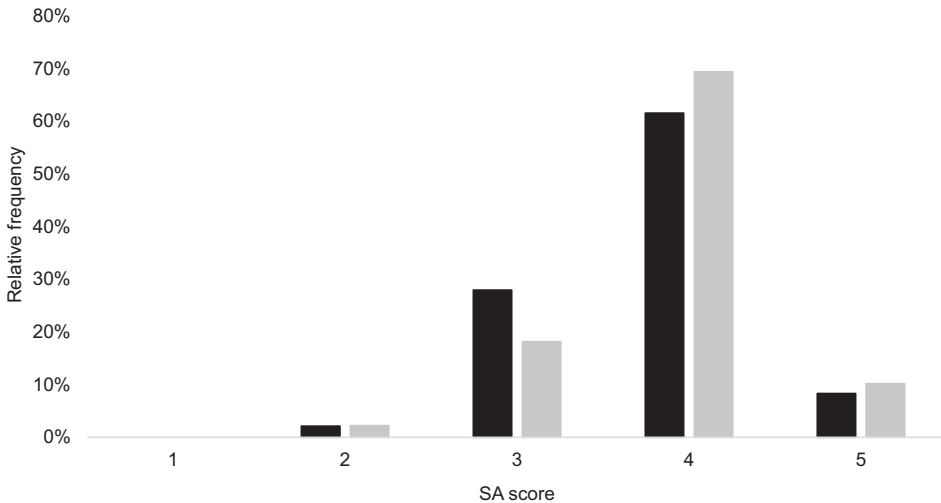


Figure 3. Relative frequency distribution of SA scores (1 = low, 5 = high) of the first ($N = 372$) and second ($N = 402$) group. The scores of the first group are marked with black bars and the scores of the second group with grey bars.

better feedback to UPs in RPD-type tactical decision-making. The training intervention was designed to improve IPs' knowledge about tactical decision-making and the use of CDM-type probes as a way to elicit the UPs' SA and static knowledge underlying those decisions. The IPs' newly acquired debrief behaviours were then used to provide enhanced feedback on the UPs' SA and decision-making performance. The impact of the intervention was evaluated using Kirkpatrick's model for training evaluation.

In contrast to many training interventions [24], this study evaluated the efficacy of the training at all four levels described by Kirkpatrick. The first level of evaluation targeted the IPs' Reactions towards the training intervention's content. The IPs' initial Reactions towards the training intervention were positive (see Table 3). Learning was evaluated by conducting the proximal and distal evaluations. These took place immediately after the training and four months later. There were no significant differences between the IPs' proximal and distal evaluation scores (see Table 4). This result demonstrated that the material was still regarded as being relevant once it had been put into practice in several hundred post-sortie debriefs. The evaluation scores reflected the fact that SA was already a well-known concept among the IPs. However, the IPs learned most about the CDM procedure and how it could be applied to air combat training (see Tables 4 and 5). In Table 5, some pairwise comparisons were non-significant as they were comparisons between closely related questions, which was the case, e.g. in the comparisons of questions 1 and 5, and 7 and 8. However, most questions addressed different aspects of the course.

Kirkpatrick and Kirkpatrick [23] have noted that an evaluation at the third level, i.e. Behaviour, is often omitted. In this study, the impact the training intervention had on the IPs' knowledge elicitation behaviours was evaluated by comparing the IPs' CDM-related debrief behaviours before and after training. There was a significant increase in the observed behaviours aimed at eliciting the UPs' static knowledge as well as their SA at Levels 1-3 (see Table 6). This result implies that the CDM training intervention was successful in improving this aspect of IPs' training evaluation.

The training intervention also resulted in a significant increase in the CDM ratio, indicating an increased 'depth' of IPs' elicitation behaviours (see Table 6). SA is a hierarchical construct, with the projection of future states (SA Level 3) being dependent upon building SA at Levels 2 (comprehension) and 1 (perception) [10, 11]. The results summarised in Table 6 are logical in that it is generally easier

to elicit what someone knows about a situation (i.e. SA Level 1) compared to eliciting how that person thinks the situation will change in the near future (SA Level 3). In summary, the increase in the CDM ratio implies that the CDM training enhanced the feedback to UPs.

A unique aspect of this paper is that it evaluated the effectiveness of the CDM training intervention by measuring the performance of UPs after IPs had received their training. In Kirkpatrick's terms, this was essentially a Level 4, Results-based, evaluation. This fourth criterion evaluated the training intervention by comparing the SA and overall performance scores of the two UP groups – one trained by IPs who had been exposed to the training intervention and the other who had not. The group of UPs which was instructed by the trained IPs had higher SA and overall performance scores than the group instructed by the non-trained IPs (see Table 7), showing evidence of enhanced combat effectiveness. In a modern air combat, even minor improvements (see Figs. 2 and 3) provide potential advantages. Also, it should be emphasised that this paper was essentially a proof-of-concept study. If changes in the curriculum are made based on the findings of this study, one can expect bigger improvements. Kirlik et al. [28] reported that providing feedback directly related to decision-making during a training exercise improved trainees' performance. Similarly, according to Li and Harris [29], even a relatively simple, short and well-planned training programme had a positive impact on pilots' decision-making. This study demonstrated that the CDM training intervention improved the IPs ability not just to identify DPs, but also to elicit the UPs' knowledge underlying their decision at DPs. Such enhanced feedback improved the performance of UPs (see Table 7). Simply immersing UPs into a complex simulation scenario without providing a structured debrief targeting specific training objectives, such as the development of SA and RPD skills, is not effective. Klein [20] emphasised that to train RPD, diagnostic and timely feedback, as well as performance review, were required. In addition, Endsley [10] and [11] argued that RPD required high levels of SA. The results of this study demonstrated that with CDM-based feedback given by IPs, UPs' decision making improved.

Training IPs to enhance their capability to make observations of UPs' performance is essential. The quality of any decision cannot be assessed by outcome alone [14]. In this study, the feedback to UPs included analysis of their decision-making processes [15–17]. The findings clearly illustrate that with the help of CDM, the IPs' understanding of the knowledge influencing the UPs' decision making improved. Similarly, as found by Crandall et al. [30], nurses' understanding of the information affecting treatment evaluation and decision-making increased significantly with CDM. A further CDM study [31] showed that nurses' comprehensive knowledge of their patients also improved patient care. Likewise, in this study, performance during air combat training improved as a result of IPs' improved understanding of UPs' knowledge.

This paper highlighted the superiority of a CDM-type performance feedback protocol compared to the traditional training approach. A short CDM training intervention to IPs resulted in a positive change in UPs' overall performance and SA scores. These observations, together with the fact that only minimal changes to the traditional debrief delivery was required, highlight the low-cost, high-benefit nature of the CDM-protocol. It should also be noted that the CDM protocol is well suited for air combat training conducted with both live and virtual as well as with live-virtual-constructive simulations [32]. When assessing the benefits of the CDM training intervention, the results were unlikely affected by cognitive bias of IPs, which could have arisen from the study's anticipated results. This is because it is implausible that the training intervention would have caused the qualified IPs to deviate from the standardised scoring principles regarding the UPs' overall performance and SA.

While this paper focused on air combat training, the CDM training approach presented can be easily applied in other military and civilian flight training contexts. Moreover, the approach is not limited to just flight training. It can be used in other training domains if those domains have a suitable debrief tradition and appropriate debrief tools to augment memory recall needed for DP identification and knowledge elicitation. This approach is not limited to just training environments as it also has potential to improve performance in operational settings.

5.0 Conclusions

The present study explored differences between traditional and CDM-type performance feedback protocols applied to air combat training. A group of IPs was given a short training intervention about the CDM protocol. The effectiveness of this training was evaluated at the four levels described by Kirkpatrick. The training intervention successfully improved the IPs' knowledge about CDM as a method of eliciting UPs' knowledge underlying their tactical decisions. As a result of this, the IPs' feedback to the UPs shifted from evaluating just the outcomes of decisions to the processes and knowledge underpinning them. The group of the UPs whose air combat training debriefs were delivered using the CDM-protocol showed superior SA and overall performance, demonstrating the effectiveness of the training intervention.

In air combat, even small improvements in UPs' performance and SA can provide a significant tactical advantage. Such positive changes were achieved with the CDM training intervention to IPs, which eventually resulted in only minimal changes to the traditional debrief delivery. In summary, the results clearly highlighted the low-cost, high-benefit nature of the CDM protocol. A large-scale implementation of the protocol into the flying curriculum should result in even greater performance gains. The objective is to replicate the study in a fighter squadron with combat-ready pilots.

Competing interests. The authors declare none.

References

- [1] Bryant, D.J. Rethinking OODA: Toward a modern cognitive framework of command decision making, *Mil. Psychol.*, 2006, **18**, (3), pp 183–206. https://doi.org/10.1207/s15327876mp1803_1
- [2] US Air Force. Air Force Doctrine Document 1, 2003.
- [3] Zsombok, C.E. and Klein, G. *Naturalistic Decision Making*, Lawrence Erlbaum Associates, Mahwah, NJ, 1997. <https://doi.org/10.4324/9781315806129>
- [4] Orasanu, J. and Connolly, T. The reinvention of decision making, In Klein, A., Orasanu, J., Calderwood, R. and Zsombok, C. (Eds.), *Decision Making in Action: Models and Methods*, Ablex, Norwood, NJ, 1993, pp. 3–20.
- [5] Simon, H.A. A behavioral model of rational choice, *Q. J. Econ.*, 1955, **69**, (1), pp 87–103. <https://doi.org/10.2307/1884852>
- [6] Payne, J.W., Bettman, J.R. and Johnson, E.J. Adaptive strategy selection in decision making, *J. Exp. Psychol. Learn. Mem. Cognit.*, 1988, **14**, (3), pp. 534–552. <https://doi.org/10.1037/0278-7393.14.3.534>
- [7] Klein, G.A., Calderwood, R. and Clinton-Cirocco, A. Rapid decision making on the fire ground: The original study plus a postscript, *J. Cogn. Eng. Decis. Mak.*, 2010, **4**, (3), pp 186–209. <https://doi.org/10.1518/155534310X12844000801203>
- [8] Cohen, M.S., Freeman, J.T. and Thompson, B.B. Critical thinking skills in tactical decision making: A model and a training strategy, In Cannon-Bowers, J. and Salas, E. (eds.), *Making Decisions Under Stress: Implications for Individual and Team Training*. American Psychological Association, Washington, DC, 1998, pp. 155–189. <https://doi.org/10.1037/10278-006>
- [9] Zakay, D. and Tsal, Y. The impact of using forced decision-making strategies on post-decisional confidence, *J. Behav. Decis. Mak.*, 1993, **6**, (1), pp 53–68. <https://doi.org/10.1002/bdm.3960060104>
- [10] Endsley, M.R. Toward a theory of situation awareness in dynamic systems, *Hum. Factors*, 1995, **37**, (1), pp 32–64. <https://doi.org/10.1518/001872095779049543>
- [11] Endsley, M.R. The role of situation awareness in naturalistic decision making, In Zsombok, C and Klein, G. (Eds.) *Naturalistic Decision Making*, Lawrence Erlbaum Associates, Mahwah, NJ, 1997, pp 269–284.
- [12] Orasanu, J. and Fischer, U. Finding decisions in natural environments: The view from the cockpit. In Zsombok, C and Klein, G. (Eds.) *Naturalistic Decision Making*, Lawrence Erlbaum Associates, Mahwah, NJ, 1997, pp 343–357.
- [13] Mansikka, H., Virtanen, K. and Harris, D. Dissociation between mental workload, performance, and task awareness in pilots of high performance aircraft, *IEEE Trans. Hum. Mach. Syst.*, 2018, **49**, (1), pp 1–9. <https://doi.org/10.1109/THMS.2018.2874186>
- [14] Brown, R.V., Kahr, A.S. and Peterson, C. *Decision Analysis for the Manager*, Holt, Rinehart and Winston, New York, NY, 1974.
- [15] Li, W.C. and Harris, D. The evaluation of the decision making processes employed by cadet pilots following a short aeronautical decision-making training program, *JAAAS*, 2006, **6**, (2), pp 315–333.
- [16] Li, W.C. and Harris, D. The evaluation of the effect of a short aeronautical decision-making training program for military pilots, *Int. J. Aviat. Psychol.*, 2008, **18**, (2), pp 135–152. <https://doi.org/10.1080/10508410801926715>
- [17] Mansikka, H., Virtanen, K., Harris, D. and Jalava, M. Measurement of team performance in air combat—have we been underperforming?, *Theor. Issues Ergon. Sci.*, 2021, **22**, (3), pp 338–359. <https://doi.org/10.1080/1463922X.2020.1779382>
- [18] Crandall, B., Klein, G. and Hoffman, R. *Working Minds: A Practitioner's Guide to Cognitive Task Analysis*, MIT Press, Cambridge, MA, 2006.
- [19] Klein, G., Calderwood, R. and MacGregor, D. Critical decision method for eliciting knowledge, *IEEE Trans. Syst. Man Cybern.*, 1989, **19**, (3), pp 462–472. <https://doi.org/10.1109/21.31053>
- [20] Klein, G. *Sources of Power: How People Make Decisions*, MIT Press, Cambridge, MA, 1998.

- [21] Endsley, M.R. The divergence of objective and subjective situation awareness: A meta-analysis. *J. Cogn. Eng. Decis. Mak.*, 2020, **4**, (1), pp 34–53. <https://doi.org/10.1177/1555343419874248>
- [22] Gagné, R.M. Educational technology and the learning process, *Educ. Res.*, 1974, **3**, (1), pp 3–8. <https://doi.org/10.3102/0013189X003001004>
- [23] Kirkpatrick, D.L. and Kirkpatrick, J.D. *Evaluating training programs: The four levels*, Berrett-Koehler Publishers, San Francisco, CA, 2006.
- [24] Reio, T.G., Rocco, T.S., Smith, D.H. and Chang, E. A critique of Kirkpatrick's evaluation model, *New Horiz. Adult Educ.*, 2017, **29**, (2), pp 35–53. <https://doi.org/10.1002/nha3.20178>
- [25] Plant, K. and Stanton, N. What is on your mind? Using the perceptual cycle model and critical decision method to understand the decision-making process in the cockpit, *Ergonomics*, 2013, **56**, (8), pp 1232–1250. <https://doi.org/10.1080/00140139.2013.809480>
- [26] Mansikka, H., Virtanen, K., Uggeldahl, V. and Harris, D. Team situation awareness accuracy measurement technique for simulated air combat - Curvilinear relationship between awareness and performance, *Appl. Ergon.*, 2021, **96**, 103473. <https://doi.org/10.1016/j.apergo.2021.103473>
- [27] Tamkin, P., Yarnall, J. and Kerrin, M. *Kirkpatrick and Beyond: A Review of Models of Training Evaluation*, Institute for Employment Studies, Brighton, UK, 2002.
- [28] Kirlik, A., Arthur, D., Walker, N. and Rothrock, L. Feedback augmentation and part-task practice in training dynamic decision-making. In Cannon-Bowers, J.A. and Salas, E. (Eds.), *Making decisions under stress: Implications for individual and team training*. American Psychological Association, Washington, DC, 1998, pp. 91–113.
- [29] Li, W.C. and Harris, D. A systems approach to training aeronautical decision making: From identifying training needs to verifying training solutions, *Aeronaut. J.*, 2007, **111**, (1118), pp 267–279. <https://doi.org/10.1017/S0001924000004516>
- [30] Crandall, B. and Getchell-Reiter, K. Critical decision method: A technique for eliciting concrete assessment indicators from the intuition of NICU nurses, *ANS Adv. Nurs. Sci.*, 1993, **16**, (1), pp 42–51. <https://doi.org/10.1097/00012272-199309000-00006>
- [31] Gazarian, P., Henneman, E. and Chandler, G. Nurse decision making in the prearrest period, *Clin. Nurs. Res.*, 2010, **19**, (1), pp 21–37. <https://doi.org/10.1177/1054773809353161>
- [32] Mansikka, H., Virtanen, K., Harris, D. and Salomäki, J. Live-virtual-constructive simulation for testing and evaluation of air combat tactics, techniques, and procedures, part I: Assessment framework, *J. Def. Model. Simul.*, 2021, **18**, (4), pp 285–293. <https://doi.org/10.1177/154851291988637>