Ergonomics

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/terg20

Examining flight time, cognitive reflection, workload, stress and metacognition on decision making performance for pilots during flight simulation

Aoife Mohan, Boban Simonovic, Katia C. Vione & Edward Stupple

To cite this article: Aoife Mohan, Boban Simonovic, Katia C. Vione & Edward Stupple (19 Sep 2024): Examining flight time, cognitive reflection, workload, stress and metacognition on decision making performance for pilots during flight simulation, Ergonomics, DOI: 10.1080/00140139.2024.2404642

To link to this article: https://doi.org/10.1080/00140139.2024.2404642

Published online: 19 Sep 2024.

ERGONOMICS
https://doi.org/10.1080/00140139.2024.2404642

Taylor & Francis
Taylor & Francis Group

ARTICLE

OPEN ACCESS Check for updates

Examining flight time, cognitive reflection, workload, stress and metacognition on decision making performance for pilots during flight simulation

Aoife Mohan, Boban Simonovic, Katia C. Vione and Edward Stupple

School of Psychology, University of Derby, UK

ABSTRACT

Despite technological advancements, human decision errors still contribute to civil aviation accidents. This study investigated whether flight time, cognitive reflection, task-load, metacognition, and perceived stress predicted decision-making (DM) performance during two in-flight training simulations with 104 commercial pilots at Bogota International Airport. Hierarchical regression analysis revealed that the predictors accounted for 56% of the variance. Cognitive reflection, flight time and performance task load emerged as significant positive predictors. Cognitive reflection significantly moderated the relationship between flight time and DM performance, with pilots scoring lower on cognitive reflection showing improved DM with increased flight time, while controlling for performance task load. The study did not find significant relationships between stress metacognition and DM performance. The study emphasises the significance of advanced training methods in improving pilots' DM, especially for those with low cognitive reflection. Future research should expand to multiple airlines, address gender balance, and incorporate direct measures of metacognitive monitoring.

Practitioner summary: This article examines predictors of decision-making performance among commercial pilots during in-flight simulations. Findings suggest that cognitive reflection, flight time and performance task load positively influence decision-making, while task-load has a negative impact. This insight can inform aviation training programs to enhance safety and update pilots' training in uncertain conditions.

ARTICLE HISTORY
Received 16 March 2024
Accepted 10 September 2024

KEYWORDS
Cognitive reflection; workload; metacognition; stress and decision-making performance

1. Introduction

1.1. Background and importance of decision making in aviation

In aviation, decisions that warrant examination are those that result in catastrophic events (Flight Safety Foundation [FSF] 2010). Researchers investigating accidents often focus on the 'moment of choice' when a decision is made (Strauch 2016). However, this narrow view of decision-making (DM) processes can misrepresent the situations faced by pilots in a dynamic sequential context (Brehmer 1992). Erdfelder and Buchner (1998) refer to this view of DM as hindsight bias, which is the belief that an event was more predictable than it actually was. As a result, DM can lead to over-simplification of cause and effect, where the correct choice appears obvious after the fact but not at the time the decision was made. The FSF (2010) linked several fatal accidents to this error.

Improved industry standards, systemic changes, and technological developments have made aviation safer in recent decades (International Civil Aviation Organisation [ICAO] 2019). These systemic changes include the introduction of standard operating procedures (SOPs) and flight crew checklists covering a variety of anticipated emergencies (ICAO 2019). While these systems offer benefits, problems can escalate in ways that make standard checklists or solutions ineffective. Consequently, DM errors persist and contribute to accidents due to rule fragility (Orasanu and Martin 1998). Rule fragility refers to the susceptibility of rule-based DM to failure or misapplication in certain situations (e.g. low visibility landings for pilots) (Clewley and Stupple 2015). Although rule-based cognition is

CONTACT Boban Simonovic b.simonovic1@derby.ac.uk University of Derby, Kedleston Road, Derby DE22 1GB, UK

crucial in these situations, there's a risk that rules may be applied incorrectly or may not fit the specific circumstances. Thus, it is essential to assist pilots in maintaining cognitive flexibility to adapt and make effective decisions when faced with unforeseen challenges.

1.2. Naturalistic decision making (NDM)

Naturalistic Decision Making (NDM) provides the theoretical explanation of decision-making processes in the aeronautical industry (Klein et al. 1993). NDM identifies the ways individuals use domain-specific knowledge and prior experience to make decisions in safety-critical situations in the real world (Cannon-Bowers et al. 1996). In time-sensitive scenarios, NDM results in intuitive, automatic decision-making processes that require minimal mental effort and occur with little conscious awareness (Lipshitz 1993). Pilots rely on experience and knowledge to rapidly process and execute decision-making strategies based on context-bound modelling, shifting from normative processes to a more naturalistic process based on past experiences (Klein et al. 2010). Thus, the choice of action is accepted or rejected rapidly and based on pattern recognition until the satisfying course of action is deemed reasonable.

The premises of NDM align with Simon's (1987) model of satisficing, whereby experts make quick decisions in highly critical situations using Recognition Primed Decision Making (RPD) (Klein et al. 2010). RPD can be effective in real-life scenarios, as demonstrated by the famous Hudson River landing (Eisen and Savel 2009). In this case, when the engine restart checklist failed, the captain successfully employed RPD by instinctively deciding to land the plane using available options (National Transportation Safety Board (NTSB) 2010). However, RPD can also lead to errors if the situation is incorrectly assessed, as seen in the 1989 British Midlands B737-400 Kegworth crash, where the crew mistakenly shut down the wrong engine (Cooper, 1990).

1.3. Dual process theory and workload

The RPD model aligns with Kahneman's (2011) dual process theory, categorising Type I processes (fast, automatic, associative and intuitive) in contrast to Type II processes (slow, analytic, rule-based and logical). This model posits that expert decision-makers pivot between intuitive and analytical processes based on prior learning (Klein 1998; Klein et al. 2010). The position on this continuum is predominantly dictated by external workload conditions, with optimal workload correlating with peak performance (Young and Stanton 2002). Excessive workload leads to maladaptive reactions, degrading crucial cognitive resources such as attention, working memory, and cognitive reflection (Eysenck et al. 2007). In such scenarios, operators may lack cognitive resources to recognise inaccuracies in situational assessments and DM (Di Nocera et al. 2007). Additionally, DM processes can succumb to biases (Stupple et al. 2013), as operators resist changing initial assessments and lack awareness of cues necessitating a change of plan (Orasanu et al. 2002). Clewley and Stupple (2015) demonstrated that increased system complexity raises the chance of rule-based failures, with situational dynamism and uncertainty contributing to rule fragility and failure. This highlights the need for nuanced workload management to optimise DM efficacy in complex operational contexts.

Indeed, pilots are influenced by cognitive biases and can resort to Type I, heuristic processing, forming quick deductions rather than rational responses in uncertain situations (Causse et al. 2013; Walmsley and Gilbey 2016, 2017). Heuristic processing can reduce mental workload but can also lead to riskier decisions based on incorrect assumptions (Croskerry 2014). For instance, VFR pilots making weather-related decisions tend to under-adjust to new information and make decisions based on the outcome instead of information quality (Walmsley and Gilbey 2016, 2017). Furthermore, they are likely to land during bad weather conditions, and can be biased towards risk taking (Causse et al. 2013;Orasanu et al. 2002). Thus, prioritising Type I processing may impede reflective and safer decisions. It should be noted that we acknowledge that skilled intuitive judgement can be developed with experience, and it is likely to be acquired when the individual operates in a high-validity environment (Kahneman and Klein 2009). Indeed, the decisions that pilots need to make to respond appropriately in a particular situation may be unique, and since the required DM is not rehearsed it is prone to heuristic processing and biases that require Type II, reflective correction (Dismukes et al. 2015).

1.4. Reflective processing and metacognition

Reflective processing positively correlates with DM (Lesage et al., 2013; Sirota and Juanchich 2011; Toplak et al. 2014); however, there is limited research on the impact of reflective processing in NDM models. Juanchich et al. (2016) found that reflective processing, as measured via cognitive refection test (CRT) predicted positive real-world decision outcomes.

Nevertheless, the real-world DM scenarios utilised by Juanchich et al. (2016) are very different from the dynamic scenarios faced by pilots. Therefore, research specifically into reflective processing in safety critical environments is required to clarify this relationship. This is particularly important because high cognitive reflection is associated with less risky DM behaviour (Frederick 2005), avoidance of decision biases (Toplak et al. 2014) and a greater likelihood of inhibiting and overriding 'hot' processes in stressful situations with 'cool' reflective processes (Causse et al. 2013; Simonovic et al. 2017, 2018).

An important aspect of cognitive reflection is its potential to facilitate awareness, control, and monitoring processes to help identify potential thinking errors and encourage the deployment of alternative strategies (Thompson and Johnson 2014). These metacognitive strategies aid self-regulation processes in task performance monitoring, good reasoning and DM performance (Ackerman and Thompson 2017). Metacognition refers to thinking about thinking and metacognitive processes consist of task monitoring, analysis, inferences and drawing conclusions during complex cognitive tasks (Dwyer et al. 2014). There is some evidence that metacognitive training can improve pilot awareness and regulation of cognitive processes, allowing less experienced pilots to make satisfactory decisions in a timely manner (Simpson 2001). Furthermore, strong metacognitive abilities may be pivotal in developing strong monitoring skills that enhance understanding of rules and explicit procedures, allowing for more effective and flexible DM and detection of rule failure (Orasanu et al. 1993). Moreover, metacognition can mitigate the effects of stress on cognition processes (Toneatto, 2002). However, the evidence is not conclusive as a mock air defence study found no significant impact of metacognition under high-stress conditions, possibly due to its sample of university students rather than trained pilots (Adams-White et al. 2018). This raises questions about whether experience and training enhance metacognitive skills and if these can be developed through domain-specific practice (Wojcik et al. 2021; Rosa et al. 2020).

1.5. Research motivation and objectives

Much research on aviation and NDM focuses on acute situational stress (Keinan 1987; Staal 2004) or the impact of strong negative emotions (Causse et al. 2013). Acute situational stress is believed to compromise the performance of even highly skilled pilots (Dismukes et al. 2015), and stress increases the likelihood of riskier decisions (Lehner et al. 1997). These studies highlight the detrimental effects of stress on situational assessment and DM among pilots (Strauch 1997; Wickens 2002). However, there is a lack of research on how everyday life-related stress and workload affect expert performance, despite evidence linking poor pilot performance to home stress, workload, and workload related performance (Garbóczy et al. 2021; Fiedler et al. 2000; Škare et al. 2021; Young 2008). Thus, a critical gap exists in understanding how more generalised stress and workload may significantly impact pilot's DM performance and their cognitive reflection abilities.

The current study investigated pilot's DM performance during two highly realistic scenarios taken from a large international commercial airline's recurrent simulator training programme. Correlations between stress, workload, cognitive reflection, metacognition and DM performance during the Evidence-Based Training (EBT) and its derivative, Competency-Based Training and assessment (CBTA) were examined. EBT and CBT are intended to prepare pilots for unanticipated operational risks by developing and assessing key competencies (ICAO 2013). This development in aviation training is pertinent as domain-specific training may not encourage flight crew to reflect upon their decision-making process, which could limit their ability to deal with irregular events (Rosa et al. 2020). Cultivating a finite number of competencies allows pilots to handle in-flight scenarios that are unanticipated by the industry and for which the crew has not been specifically trained. Therefore, the DM performance of the pilot sample is expected to be at a high standard across all competencies (ICAO 2013).

High-quality CBTA has a positive effect on aviation safety, regardless of accumulated flight time (FTime) experience (Smith et al. 2016). This is of concern given the projected growth of the aviation industry (Sainarayan 2018), and the prediction that pilot demand will outpace supply by up to three times (Schonland 2016). Consequently, pilots may need to be promoted to Captain positions at a younger age and with less operational FTime than in previous decades (Keller et al. 2019). The scenarios used in this study were developed by subject matter experts to test DM performance by simulating a series of high-fidelity, safety-critical events from the operational environment. The environment is considered naturalistic, as simulated tasks have been shown to elicit behaviours that are similar to those seen in an actual event but without the corresponding risk (Lipshitz et al. 2001).

The study predicts a significant combined relationship between FTime cognitive reflection, stress, workload, and metacognition on pilot's DM performance in two high-fidelity simulated inflight emergency scenarios. Based on the evidence presented above, it is predicted that metacognition and cognitive reflection will mitigate the negative effects of stress and workload, resulting in a positive overall relationship. We hypothesise a significant positive relationship between cognitive reflection ability and DM performance. Additionally, we predict significant negative relationships between task load and DM, and between perceived stress and DM, as well as a significant positive relationship between metacognition and DM. We also conducted a moderation analysis using flight time as the predictor, DM performance on different flight scenarios as the outcome, cognitive reflection as the moderator, and workload as a covariate. This approach allows us to examine how cognitive reflection influences the relationship between FTime and DM performance, while accounting for workload. Understanding these dynamics is crucial, given that high-quality CBTA positively affects aviation safety regardless of flight time experience (Smith et al. 2016), especially considering the anticipated growth in the aviation industry (Sainarayan 2018).

2. Methods

2.1. Participants

Pilots were recruited via the Competency Based and Evidence Based Training programme (CBTA and EBT) from a large international airline. A priori power analysis was conducted for the regression analysis, and for a small to medium effect size, a minimum of 105 participants would be required to have 80% power. There were 103 males and one female; 54 Captains, 50 First Officers. Participant's total FTime ranged from 3,000 to 27,000 hours, with an average of 10,847.58 hours (SD = 4891.62). Participation was voluntary and participants were current operational flight crew of wide-body jet aircraft with no known training or performance issues and no diagnosed psychiatric, affective, neurological disorders or brain injuries.

2.2. Materials

2.2.1. The 10 item perceived stress scale (PSS)

This scale was developed as a 14-item scale however the shorter 10-item version is used in the current study as it has stronger psychometric properties with higher internal consistency (Cohen et al. 1994; Lee 2012). Participants completed the Perceived Stress Scale (PSS) before the simulation to measure their stress levels accurately at that time. Pre-simulation measures are crucial because post-simulation measures may not accurately reflect the participants' initial stress state, as the simulation experience could alter their perceived stress levels. Thus, we assured that the baseline stress is not influenced by the experimental task, providing a more valid measure of participants' stress. Responses were recorded on a 5-point Likert scale, from 0 (never) to 4 (very often) except for the four positively framed items (items 4, 5, 7, and 8). which were reversely scored. The scale showed moderate reliability in our study (Cronbach's α = .595).

2.2.2. The nasa task load index (NASA-TLX)

The NASA-TLX is a subjective measure of workload across 6 dimensions including: Mental demand, physical demand, temporal demand, effort, performance and frustration level which has good internal consistency (Xiao et al. 2005; Hart and Staveland 1988). Participants rate their subjective score on an interval scale ranging from one (low) to 21 (high). Sums of scores provide an overall workload rating (Hart and Staveland 1988). The scale showed good reliability in our study (Cronbach's α = .720). We also calculated individual subscales that were used because the dimensions contribute differently to overall workload and can detect subtle variations that may be missed by the composite score alone (Louis et al. 2023). Using the subscales provided more diagnostic power, allowing us to determine which specific aspects of workload are most affected by task demands.

2.2.3. The seven-item cognitive reflection test (CRT)

The CRT comprises 7 items that measure one's ability to resist and override intuitive responses by engaging analytic ability (e.g. 'A bat and a ball cost $1.10 in total (Toplak et al. 2014). The bat costs a dollar more than the ball. How much does the ball cost?'). Here, one's intuitive response is to state that the ball costs $0.10 (totalling $1.20), when the correct answer is $0.05. Each item is rated as either correct or incorrect, with higher scores indicative of greater levels of cognitive reflection The original CRT showed high reliability (Cronbach's $\alpha = 0.720$), which was replicated in our study (Cronbach's $\alpha = 0.742$). The test was presented in the four-option multiple choice version validated by Sirota and Juanchich (2018), answer presentation was randomised for each participant.

2.2.4. Metacognition self-assessment scale (MSAS)

The MSAS measures five components of metacognition: monitoring, differentiation, integration, decentration and mastery with 18 items (Pedone et al. 2017). These are scored on a 5-point Likert scale which yields a summed score ranging between 18 and 90. Higher scores indicate better self-evaluation of metacognitive functioning. It has a good factorial validity and internal consistency and is considered a useful, rapid screening tool of functional metacognitive ability (Pedone et al., 2017). The scale showed good reliability in our study (Cronbach's $\alpha = .752$).

2.2.5. Decision making

Decision-making (DM) performance was assessed and graded by type rating examiners (TREs), who are subject matter experts with extensive airline operational and training experience. The assessment was based on observable behaviours demonstrating the relevant knowledge, skills, and attitudes required to perform job-related activities under specified conditions (ICAO 2020). Performance was evaluated according to the demonstration of observable 'on the job' behaviours. The airline's performance framework uses a five-level grading scale, integrating elements from the EBT Implementation Guide (ICAO 2013). Grades 3–5 indicate achievement of completion standards, while grades 1–2 signify performance below required standards. Thus, performance on both scenarios were graded together for on overall DM performance grade.

To ensure consistency and objectivity in assessment, TREs use a detailed list of observable behaviours and descriptive word pictures for each performance level. This approach helps characterise the level of performance and resolve any uncertainties in grading DM performance.

2.3. Procedure

The study was approved by the University of Derby's College of Health, Psychology and Social Care ethics committee. TREs were briefed by their training manager and provided with detailed facilitator instructions via email. They invited crews attending recurrent training in March, April, and May 2021 to participate voluntarily, stressing that participation would not affect their training outcomes and that results would remain anonymous and confidential. Participants accessed the study through a Qualtrics link provided by their TRE.

The study consisted of three stages. In Stage 1, permission to conduct the study was obtained from the Senior Vice President for Flight Operations at the participating airline. Participants completed the PSS before the simulation to measure pre-simulation stress levels. Demographic information, including age, sex, education level, total flight-time hours, and current rank, was recorded.

In Stage 2, the observational element was conducted in a CAE Inc. full-motion aircraft simulator. Two highly realistic scenarios, developed by subject matter experts, were used from the airline's recurrent training program. These scenarios took place at Bogota International Airport, a challenging environment due to high elevation, mountainous terrain, and prevailing wind conditions (GCAA 2021). In the first scenario, pilots experienced an aircraft technical malfunction resulting in a loss of instrumentation. In the second scenario, they managed an engine failure. Both scenarios required critical decision-making to continue the flight to Miami, divert to an alternate airfield, or return to Bogota. The TRE assessed and graded DM performance, recording grades in Qualtrics.

In Stage 3, after the simulation, participants completed the NASA TLX, CRT, and MSAS. Finally, participants were debriefed and re-consented to having their data

2.4. Procedural adjustments

Prior to the main study, a pilot test involving a small group of participants was conducted to refine the procedure and materials. Feedback from this testing led to minor adjustments in the instructions and the simulation scenarios to enhance clarity and realism. Flight Ops managed the pilot and trainer pairings based on operational factors like scheduling and training needs, making the pairings random. The disparity between male and female crew arose from the low number of female pilots, with the participating female being the only one scheduled for training during the experiment. This random pairing ensured that the DM process was not influenced by prior familiarity or specific pair dynamics. Each pair was composed of a Captain and a First Officer, reflecting real operational settings.

To control for potential confounding variables, all participants received the same instructions regarding the simulation and the completion of questionnaires. The simulation scenarios were standardised for all participants to ensure consistency. Furthermore, the TREs were aware about the purpose of the study but were blinded to the study's specific hypotheses to prevent bias in their assessments Lastly, randomisation was used in the presentation of CRT questions and in pairing pilots to mitigate order effects and pairing biases.

2.5. Analytic strategy and scoring

Prior to analysis, data were subjected to graphical and statistical screening. To address the non-normal distribution observed in DM and CRT variables, a logarithmic transformation was applied to the data and used in all analyses. All variables' scores are computed according to the original Authors instructions. We used the total PSS, CRT and MSAS scores. For the NASA-TLX scale we used 6 subscales [Mental Demand (MD), Physical Demand (PD), Temporal Demand (TD), Effort (E), Performance (P) and Frustration (F)] separately as the granular information can be valuable for identifying particular sources of workload in complex tasks such as DM task in this study Hertzum 2021). Initial analysis examined inter-correlations between all variables. Next, a hierarchical regression was conducted only with significant correlations observed. Thus, FTime, Temporal Demand, Performance, Frustration, and CRT as predictors of DM performance as an outcome. Lastly, a bootstrapped moderation analysis was conducted to examine whether the relationship between FTime and DM was moderated by cognitive reflection, while controlling for potential covariates of workload. We used the Process macro for SPSS (Hayes 2012), with 5,000 bootstrapping re-samples and bias-corrected 95% Confidence Intervals (CIs) to test the moderation effect. IBM SPSS Statistics 27 software package for Mac was used to analyse data.

3. Results

3.1. Intercorrelations between all dependent measures

Table 1 presents inter-correlations between all dependent measures. A negative correlation between DM Temporal Demand (TD) and Frustration (F) was observed. Additionally, FTime, cognitive reflection (CRT) and Performance (P) positively correlated with DM.

3.2. Hierarchical regression

The selection of predictors and their order of entry in the hierarchical regression analysis was based on theoretical considerations from prior research and the strength of previously observed significant correlations, following the recommendations of Field (2013). This approach ensures a logical and empirically supported sequence of variable entry, allowing for a more meaningful interpretation of the incremental contributions of each predictor to the model. FTime was entered first, as previous research indicated that FTime experience strongly influences pilot decision-making performance (Smith et al. 2016). Temporal Demand, Performance and Effort were entered next due to evidence suggesting that heavy task load correlates with decreased decision-making performance (Young and Stanton 2002). CRT was entered in the last block, as previous research indicated that higher-order cognitive processing may enhance pilot decision-making performance (Walmsley and Gilbey 2017). All steps within the hierarchal regression were significant (Step1, ($F(1, 102) = 6.69$, $p=0.01$; $R^2 = .06$); Step 2, ($F(4, 99) = 7.46$, $p<0.001$; $R^2 = .23$) Step 3, ($F(5, 98) = 5.50$, $p<0.001$; $R^2 = .27$). Data indicated that the five predictors combined accounted for 56% of the variability in DM performance. The Beta for CRT, FTime and Performance task load showed a positive correlation. Other predictors were not significant. The results indicated that higher cognitive reflection, longer flight time experience, and higher perceived quality and accuracy of pilot performance were associated with better decision-making (DM) performance. Other predictors were not significant (Table 2).

3.3. Moderation analysis

A moderation analysis was conducted to examine whether the relationship between FTime and DM was moderated by CRT (M), while controlling for performance task load. The main effect of FTime on DM was

Table 1. Intercorrelations between the variables ($N=104$).

Variables	DM	FTime	MD	PD	TD	E	P	F	CRT	MSAS	PSS
DM											
FTime	0.25**										
MD	−0.13	−0.12									
PD	−0.06	−0.09	0.54*								
TD	−0.32*	−0.08	0.45*	.40*							
E	−0.31	−16	0.64	0.33*	0.40*						
P	0.40*	0.13	−0.16	0.04	0.40*	−1.4					
F	−0.33*	0.20***	0.30*	0.31*	0.51*	0.30*	0.45*				
CRT	0.26**	0.14	0.05	0.08	0.05	−0.01	0.15	−0.03			
MSAS	0.01	0.01	0.17	−0.11	−0.09	0.05	0.14	0.19***	0.23***	.	
PSS	−0.03	0.01	0.13	0.03	0.11	0.18	−0.01	0.12	−0.05	0.11	

Note all *significant at $p<0.001$; all ** significant at $p=0.01$ and all *** significant at $p<0.05$.

not significant, ($\beta = .05$, $t(99) = 1.62$, $p = 0.11$). However, the conditional effect of CRT on DM was significant, ($\beta = .05$, $t(99) = 2.21$, $p = 0.03$). The overall model was significant, ($F(4,99) = 10.06$, $p < 0.001$), indicating that CRT was a significant moderator of the effect of FTime on DM (when controlling for performance task load) ($\beta = -.30$, SE = .11. $p = 0.005$. $R^2 = .29$) accounting for 28.89% added variation on DM.

The Johnson–Neyman technique identified the significant region of CRT where the effect of FTime on DM becomes significant. The effect was significant for low CRT values (Table 3). Further simple slopes analysis indicated that for pilots with lower CRT scores, increased FTime is associated with a significant increase in DM performance (Figure 1).

This suggests that pilots who perform poorly on the CRT may benefit from additional FTime in terms of their decision-making abilities.

4. Discussion

The current study aimed to investigate the relationship between flight time, cognitive reflection, workload, perceived stress, metacognition, and pilot's DM grades during two highly realistic NDM scenarios taken directly from a large international commercial airline's recurrent EBT program. The results revealed a combined positive relationship between the predictors and the DM performance. FTime, CRT and performance task load were significant independent positive predictors of DM. Other predictors were not significant. The moderation analysis revealed that the relationship between FTime and DM is moderated by CRT when controlling for performance task load.

Table 2. Summary of hierarchical regression analysis with DM Performance as an outcome variable.

DM predictors	*B*	*SE B*	β	Significance	R^2
Step 1					
FTime	0.09	0.03	0.25	$p = 0.01$*	0.06
Step 2					
FTime	0.06	0.03	0.18	$p = 0.04$*	
Temporal demand	−0.03	0.02	−0.14	$p = 0.18$	
Performance load	0.19	0.07	0.17	$p = 0.008$*	
Frustration	−0.02	0.02	−0.09	$p = 0.37$	0.23
Step 3					
FTime	0.05	0.03	0.15	$p = 0.08$	
Temporal demand	−0.04	0.02	−0.17	$p = 0.10$	
Performance load	0.17	0.07	0.24	$p = 0.02$*	
Frustration	−0.02	0.02	−0.10	$p = 0.35$	
CRT	0.05	0.02	0.21	$p = 0.02$*	0.27

*Denotes significant *p* values.

4.1. Flight time and decision making

The results of our study offer intriguing insights into the complex nature of pilot expertise and cognitive processes. While FTime is traditionally viewed as a key predictor of DM performance, our findings suggest a more nuanced relationship. Although a significant positive relationship between FTime and DM performance supports the notion that experience improves decision-making skills (Smith et al. 2016), the moderation effect of CRT scores indicates that the impact of flight experience on DM is not uniform across all pilots. This interaction between FTime and CRT implies that higher-order cognitive processes are important for effective DM (Walmsley and Gilbey 2017). This aligns with Mumford et al.'s (2015) concept of adaptive expertise suggesting that effective DM in complex environments like aviation requires not just experience, but also the cognitive flexibility to apply that experience appropriately. Furthermore, the significant contribution of performance task load to the model aligns with the work of Young and Stanton (2002), highlighting the impact of workload on cognitive performance in aviation settings. The combined influence of flight time, cognitive reflection, and task load on DM performance highlights the multifaceted nature of pilot expertise, emphasising the need for a holistic approach to pilot training and assessment. These findings suggest that while accumulating flight hours is important, developing cognitive reflection skills and managing task load effectively are equally fundamental.

4.2. Cognitive reflection and decision making

To our knowledge, this study is the first to investigate CRT and DM in a NDM context. The results concur with studies showing a positive correlation between CRT and DM (Lesage et al. 2013; Sirota and Juanchich 2011; Toplak et al. 2014). Our results are in line with previous research that reflective thinking can foster a thorough information search and heightened analytical cognition that helps good DM processes (Causse et al. 2013; Mosier and Fischer 2010; Rosa et al. 2020).

Table 3. Conditional effects of FTime on DM at different levels of CRT.

CRT	β	SE	t	*p*	LLCI	ULCI
−0.27	0.13	0.04	3.39	0.001*	0.05	0.21
0.00	0.05	0.03	1.62	0.10	−0.01	0.11
0.27	−0.03	0.04	−0.76	0.44	−0.12	0.05

*Denotes significant *p* value.

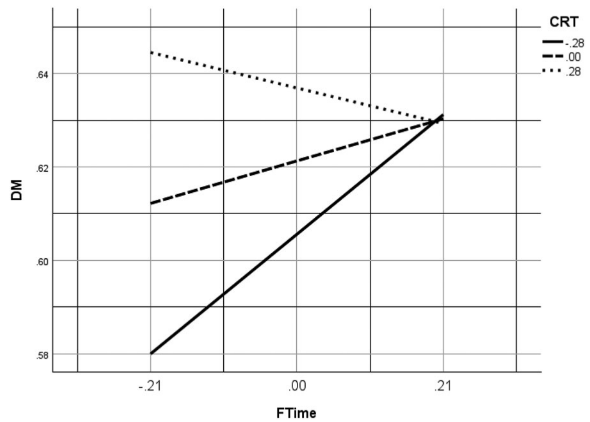

Figure 1. CRT values as moderators of flight time.

However, we do not suggest that pilots' judgement must always be based on analytical thinking. In certain scenarios, a combination of strategies may be necessary. For instance, during landing, a pilot might initially scrutinise cockpit instruments analytically, configuring mode and flight parameters to capture the glide slope, and subsequently cross-reference external cues to validate the overall assessment. Conversely, in low visibility conditions, a pilot may rely on analytical processing of cockpit information, to engage in rule-based processes (Clewley and Stupple 2015), or alternatively, opt to manually control aircraft functions during favourable visibility conditions (Walmsley and Gilbey 2017). Our results also indicate that for pilots with low cognitive reflection, increased flight time may be beneficial for DM performance. These findings highlight the necessity of considering individual cognitive differences when evaluating the impact of flight time, suggesting that personalised training and flight schedules could optimise DM outcomes (Causse et al. 2013; Taylor et al. 2005). Note that, while the results of the study are promising they may be limited to airlines that utilise effective EBT models as gaps in knowledge could lead to poorer DM performance despite high cognitive reflection.

4.3. Workload and decision making

The present study supports human factors research that demonstrates high mental workload, reduces the ability to integrate new information and increases errors (e.g. Di Nocera et al. 2007). The NASA-TLX subscales provided detailed insights into this phenomenon. For instance, temporal demand and frustration were negatively correlated, indicating that as the time pressure of a task increases, frustration levels tend to decrease, possibly due to increased focus and urgency. Conversely, performance task load showed a positive correlation with DM, suggesting that higher perceived quality and accuracy of performance are associated with better decision-making. This aligns with previous research indicating that heavy task load correlates with decreased DM performance (Young and Stanton 2002). However, it is important to note that while mental workload had a significant, albeit small (Cohen 1988), negative impact on DM, there is no clear benchmark for what constitutes high workload on the NASA-TLX scale (Hart 2006). The simulation workload might not have been high enough to demonstrate a larger effect size. Moreover, while simulator behaviours can resemble actual in-flight events (Lipshitz et al. 2001), the absence of inherent risk in simulations could attenuate the TLX results' effect size, considering the potential emotional impact on cognitive performance (Hancock and Warm 1989). Completing the NASA-TLX within 15 minutes of event completion has demonstrated strong validity (Hart 2006). However, the subjective nature of the NASA-TLX and its post-simulation completion could introduce hindsight bias (Erdfelder & Buchner 1998). The self-reported nature of the data, influenced by participants' mood during survey completion rather than the event (Paulhus and Vazire 2007) should also be acknowledged.

4.4. Perceived stress, metacognition and DM

Previous research on the effects of everyday stress on NDM is limited, although pilots have reported perceived correlations between home stress and poorer flying performance (Fiedler et al. 2000). Perceived stress was not a significant unique predictor of DM in the current study, despite the mean stress score falling into the 'very high' stress category (Lee 2012). In the present study, the anticipation of the simulation environment may not have been sufficiently stressful or prolonged to demonstrate significant impacts of stress. Indeed, the observed Cronbach's alpha value for perceived stress indicates only moderate reliability, highlighting potential issues with the scale's consistency such as test length and item interrelatedness. Thus, future research could investigate whether high levels of perceived global stress progressively degrade DM performance in experts and whether stress levels in simulation translate into real-world stress.

Metacognition did not have a significant unique effect on DM, which was unexpected given the hypothesis that pilots trained in DM should exhibit better metacognitive monitoring (Lichtenstein et al. 1977) and, consequently, greater metacognitive control over their cognitive process of DM (Kim 2018).

However, this finding is consistent with a study by Adams-White, Wheatcroft, and Jump (2018) on DM in high-stress conditions, which also found no impact of metacognition on DM. The results may be limited by the measure of metacognition used. Although the Metacognitive Skills Assessment Scale (MSAS) has good internal reliability and appears to distinguish well between different metacognitive abilities (Pedone et al. 2017), it is important to note that metacognition was self-reported and not directly observable during the task. Due to practical reasons and to maintain the realism of the simulation this study assessed metacognition as a trait rather than directly measuring monitoring during the task. Utilising a more direct measure of monitoring may offer greater insight into cognitive processes (Craig et al. 2020).

4.5. Theoretical implications

The RPD acknowledge that analytic processes come into play when situations are unfamiliar or when facets of the situation do not align with expectations (Klein 2010). Situations may necessitate analytic or constructive thought processes such as sensemaking, involving an understanding of the situation by connecting situational elements into a causal and/or temporal order. This includes placing elements into frameworks, accounting for unexpected or inconsistent elements, and engaging in planning and replanning (Klein 2008; Klein et al. 2003). Thus, it is possible that the pilots' DM strategies shift between Type I and Type II processes depending on what data or information is received. It is also possible that, framing the study as a training event with no actual risk beyond simulated emergencies may afford participants cognitive flexibility for more Type II processes. In scenarios with tangible risk, external pressure and internal uncertainty may compromise analytic cognitive performance, favouring intuitive processing (Hunter and Allen 2006). However, our results contradict the idea that intuitive cognition (Type I processes) rather than logical reflective cognition (Type II processes) guides correct DM in reaction to task settings. This implies limitations in studying NDM in simulated environments.

4.6. Practical implications

The results of the current study align with the Airline's existing policies indicating that high-quality CBTA positively impacts pilot's DM performance, regardless of flight time experience (Smith et al. 2016). Cultivating CBTA in NDM environments can enhance overall performance in unexpected emergency scenarios, a premise worth further exploration given the anticipated growth of the aviation industry and the increased demand to promote First Officers to Captains with less operational flight time experience (Keller et al. 2019; Sainarayan 2018). Additionally, for pilots with low cognitive reflection, additional flight time may help improve DM performance. Moreover, the correlation between working memory capacity and cognitive reflection (Stupple, Gale, and Richmond 2013) suggests that pilots with higher working memory capacity may better manage mental workload during flight operations. Updating pilot training manuals to enhance cognitive skills and improve performance in challenging scenarios could be beneficial. These findings also raise concerns for airlines using traditional task-based pilot training, which may not emphasise core decision-making or soft skills, posing risks even for experienced pilots who might rely on pattern recognition over comprehensive information processing. Cognitive reflection is crucial for rule-based cognition, especially in dynamic and uncertain conditions, as it supports the application and monitoring of rules, enhancing flight safety (Clewley and Stupple 2015). Without reflective processes, errors can occur in complex systems, highlighting the importance of incorporating reflective cognitive training into pilot education.

4.7. Limitations and future research

Several limitations need to be addressed. First, the study involved pilots from only one airline, limiting the generalisability of the findings. The claims should be made only concerning this airline's training approach. Second, we acknowledge that our sample reflects the gender imbalance in flight crew, which may affect the generalisability of the findings. However, this imbalance accurately represents the current, male population of pilots. This limitation should be considered, and future research should include a more balanced gender representation. Third, while simulator behaviours may resemble actual in-flight events, the absence of inherent risk in simulated events could attenuate the NASA-TLX results' effect size, particularly considering the potential emotional impact on cognitive performance in DM (Hancock and Warm 1989). Fourth, self-reported measures such as MSAS and perceived stress could generate hindsight bias and may not accurately reflect the cognitive processes during the tasks. Finally, the study assessed metacognition as a trait rather than directly measuring it during the task. Utilising more direct measures of monitoring could provide greater insight into cognitive processes (Craig et al. 2020).

Future research should address the limitations identified. Studies should include pilots from multiple airlines to enhance generalisability and explore the impact of gender balance on DM performance. Additionally, examining the relationship between global stress and DM performance over time could provide insights into how stress levels in simulations translate to real-world stress. Direct measures of metacognitive monitoring during tasks could offer a more accurate assessment of its impact on DM.

Cognitive reflection emerges as a unique and significant predictor of DM and a moderator of the flight time DM relationship in a naturalistic setting, marking the first examination of this relationship in a simulated aviation environment. Our findings reaffirm the connection between workload and DM performance. Focusing on Type II analytical processes in training sessions could enhance robust analytical cognition and promote more coherent judgement and DM strategies. The study highlights the need for further investigation into the cognitive processes underpinning pilot decision-making and the importance of developing training methodologies that enhance cognitive skills and improve performance outcomes in challenging operational scenarios.

Author's contributions

All author(s) contributed to the conceptualisation, design and writing of the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work did not receive any funding support.

Data availability statement

Data are available on request from B. Simonovic.

References

Ackerman, R., and V. A. Thompson. 2017. "Meta-Reasoning: Monitoring and Control of Thinking and Reasoning." *Trends in Cognitive Sciences* 21 (8): 607–617. doi:10.1016/j.tics.2017.05.004.

Adams-White, J. E., J. M. Wheatcroft, and M. Jump. 2018. "Measuring Decision Accuracy and Confidence of Mock Air Defence Operators." *Journal of Applied Research in Memory and Cognition 7* (1): 60–69. doi:10.1016/j.jarmac.2018.01.005.

Brehmer, B. 1992. "Dynamic Decision Making: Human Control of Complex Systems." *Acta Psychologica 81* (3): 211–241. doi:10.1016/0001-6918(92)90019-A.

Cannon-Bowers, J.A., E. Salas, and J. S. Pruitt. 1996. "Establishing the Boundaries of a Paradigm for Decision-Making Research." *Human Factors: The Journal of the Human Factors and Ergonomics Society 38* (2): 193–205. doi:10.1518/001872096779048066.

Causse, Mickaël, Frédéric Dehais, Patrice Péran, Umberto Sabatini, and Josette Pastor. 2013. "The Effects of Emotion on Pilot Decision-Making: A Neuroergonomic Approach to Aviation Safety." *Transportation Research Part C: Emerging Technologies 33*: 272–281. doi:10.1016/j.trc.2012.04.005.

Clewley, R., and E. J. N. Stupple. 2015. "The Vulnerability of Rules in Complex Work Environments: dynamism and Uncertainty Pose Problems for Cognition." *Ergonomics 58* (6): 935–941. doi:10.1080/00140139.2014.997804.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd Ed. Hillsdale, NJ: Laurence Erlbaum Associates.

Cohen, S., T. Kamarck, and R. Mermelstein. 1994. "Perceived Stress Scale." *Measuring Stress: A Guide for Health and Social Scientists 10* (2): 1–2.

Cooper, D.A., 1989. "Air Accident Investigation Branch".Department of Transport. , 1990. *Report on the accident to Boeing 737-400 G-OBME near Kegworth, Leicestershire on 8 January* . Department of Transport. Retrieved from: https://www.jesip.org.uk/uploads/media/incident_reports_and_inquiries/Kegworth%20Aircrash.pdf

Craig, K., D. Hale, C. Grainger, and M. Stewart. 2020. "Evaluating Metacognitive Self-Reports: systematic Reviews of the Value of Self-Report in Metacognitive Research." *Metacognition and Learning 15* (2): 155–213. doi:10.1007/s11409-020-09222-y.

Croskerry, P. 2014. "Bias: A Normal Operating Characteristic of the Diagnosing Brain." *Diagnosis (Berlin, Germany) 1* (1): 23–27. doi:10.1515/dx-2013-0028.

Di Nocera, F., M. Camilli, and M. Terenzi. 2007. "A Random Glance at the Flight Deck: Pilots' Scanning Strategies and the Real-Time Assessment of Mental Workload." *Journal of Cognitive Engineering and Decision Making* 1 (3): 271–285. doi:10.1518/155534307X255627.

Dismukes, R. K., T. E. Goldsmith, and J. A. Kochan. 2015. *Effects of Acute Stress on Aircrew Performance: Literature Review and Analysis of Operational Aspects: NASA Report: NASA/TM-2015–218930*. Moffett Field, CA: National Aeronautics and Space Administration.

Dwyer, C. P., M. J. Hogan, and I. Stewart. 2014. "An Integrated Critical Thinking Framework for the 21st Century." *Thinking Skills and Creativity* 12: 43–52. https://doi.org/10.1016/j.tsc.2013.12.004.

Eisen, L.A, and R.H. Savel. 2009. "What Went Right: Lessons for the Intensivist from the Crew of US Airways Flight 1549." *Chest 136* (3): 910–917.

Erdfelder, E., and A. Buchner. 1998. "Decomposing the Hindsight Bias: A Multinomial Processing Tree Model for Separating Recollection and Reconstruction in Hindsight." *Journal of Experimental Psychology: Learning, Memory, and Cognition 24* (2): 387–414. doi:10.1037/0278-7393.24.2.387.

Eysenck, M. W., N. Derakshan, R. Santos, and M. G. Calvo. 2007. "Anxiety and Cognitive Performance: Attentional Control Theory." *Emotion (Washington, D.C.) 7* (2): 336–353. doi:10.1037/1528-3542.7.2.336.

Fiedler, E. R., P. Della Rocco, D. J. Schroeder, and K. T. Nguyen. 2000. "The Relationship Between Aviators' Home-Based Stress to Work Stress and Self-Perceived Performance." *FAA Office of Aviation Medicine Reports* DOT-FAA-AM-00-32. 7.

Field, A. 2013. *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). Thousand Oaks, CA: SAGE Publications.

Flight Safety Foundation 2010, April 10. https://aviation-safety.net/database/record.php?id=20100410-0

Frederick, S. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives 19* (4): 25–42. doi:10.1257/089533005775196732.

Garbóczy, Szabolcs, Anita Szemán-Nagy, Mohamed S. Ahmad, Szilvia Harsányi, Dorottya Ocsenás, Viktor Rekenyi, Ala'a B. Al-Tammemi, and László Róbert Kolozsvári. 2021. "Health Anxiety, Perceived Stress, and Coping Styles in the Shadow of the COVID-19." *BMC Psychology 9* (1): 53. doi:10.1186/s40359-021-00560-3.

Hayes, A. F. 2012. PROCESS: a versatile computational tool for observed variable mediation, moderation and conditional process modeling [White paper]. Available online at: http://www. afhayes.com/public/process 2012.pdf

Hancock, P. A., and J. S. Warm. 1989. "A Dynamic Model of Stress and Sustained Attention." *Human Factors 31* (5): 519–537. doi:10.1177/001872088903100503.

Hammond, K. R. 1988. *Judgement and Decision Making in Dynamic Tasks*. Boulder, CO: Colorado Univ AT Boulder Center for Research on Judgment and Policy.

Hart, S. G., and L. E. Staveland. 1988. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research." In P. A. Hancock and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Amsterdam, Netherlands: North-Holland. doi:10.1016/S0166-4115(08)62386-9.

Hart, S.G. 2006. "Nasa-Task Load Index (NASA-TLX); 20 Years Later." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50* (9): 904–908. doi:10.1177/154193120605000909.

Hertzum, M. 2021. "Reference Values and Subscale Patterns for the Task Load Index (TLX): A Meta-Analytic Review." *Ergonomics* 64 (7): 869–878. Epub 2021 Feb 1. PMID: 33463402. doi:10.1080/00140139.2021.1876927.

Hoyle, R. H. 1995. "The Structural Equation Modeling Approach: Basic Concepts and Fundamental Issues.". In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (1–15). Thousand Oaks, CA: Sage Publications, Inc.

Hunter, D., and B. Allen. 2006. "Risk Perception Among General Aviation Pilots." *The International Journal of Aviation Psychology 16* (2): 135–144. doi:10.1207/s15327108ijap1602_1.

General Civil Aviation Authority 2021. [GCAA]. AMC1 ORO. FC.105(b)(2);(c) Designation as pilot-in-command/commander. Route/Area and aerodrome knowledge for commercial operations (b)(3) *CAR-AIR OPS - PART-ORO, (01).*

International Civil Aviation Organisation 2020. [ICAO]. Competency Based Training and Assessment (CBTA) ICAO DOC 9868, *Procedures for Air Navigation Services.*

International Civil Aviation Organisation [ICAO]. 2013. "Training (PANS-TRG) Third Edition, (Amendment 7) Framework and Utilise an Evidence Based Training (EBT) ICAO DOC 9995." *Manual of Evidence Based Training, 1* (1): 105–117.

International Civil Aviation Organization 2019. *ICAO Safety Report 2019 Edition* https://www.icao.int/safety/Documents/ICAO_SR_2019_29082019.pdf

Juanchich, M., C. Dewberry, M. Sirota, and S. Narendran. 2016. "Cognitive Reflection Predicts Real-Life Decision Outcomes, but Not over and above Personality and Decision-Making Styles." *Journal of Behavioral Decision Making 29* (1): 52–59. doi:10.1002/bdm.1875.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., and G. Klein. 2009. "Conditions for Intuitive Expertise: A Failure to Disagree." *The American Psychologist* 64 (6): 515–526. doi:10.1037/a0016755.

Keinan, G. 1987. "Decision Making under Stress: Scanning of Alternatives under Controllable and Uncontrollable Threats." *Journal of Personality and Social Psychology 52* (3): 639–644. doi:10.1037/0022-3514.52.3.639.

Keller, J., F. A. Mendonca, M. Suckow, and B. G. Dillman. 2019. *Transforming the professional flight program curriculum: Justification, process, and future development.* Paper presented at the Polytechnic Summit 2019, Menomonie, Wisconsin.

Kim, J. H. 2018. "The Effect of Metacognitive Monitoring Feedback on Performance in a Computer-Based Training Simulation." *Applied Ergonomics 67*: 193–202. doi:10.1016/j.apergo.2017.10.006.

Klein, G. A., Orasanu, J., Calderwood, R., and Zsambok, C. E. (Eds.). 1993. *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex Publishing.

Klein, G. A., K. G. Ross, B. M. Moon, D. E. Klein, R. R. Hoffman, and E. Hollnagel. 2003. "Macrocognition." *IEEE Intelligent Systems* 18 (3): 81–85. doi:10.1109/MIS.2003.1200735.

Klein, G. 1998. *Sources of Power: How People Make Decisions*. MIT Press, Cambridge.

Klein, G. 2008. "Naturalistic Decision Making." *Human Factors 50* (3): 456–460. doi:10.1518/001872008X288385.

Klein, G., R. Calderwood, and A. Clinton-Cirocco. 2010. "Rapid Decision Making on the Fire Ground: The Original Study Plus a Postscript." *Journal of Cognitive Engineering and Decision Making 4* (3): 186–209. doi:10.1518/155534310X12844000801203.

Lee, E. H. 2012. "Review of the Psychometric Evidence of the Perceived Stress Scale." *Asian Nursing Research* 6 (4): 121–127. doi:10.1016/j.anr.2012.08.004.

Lehner, P., M.-M. Seyed-Solorforough, M.F. O'Connor, S. Sak, and T. Mullin. 1997. "Cognitive Biases and Time Stress in Team Decision Making." *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans 27* (5): 698–703. doi:10.1109/3468.618269.

Lesage, E., G. Navarrete, and W. De Neys. 2013. "Evolutionary Modules and Bayesian Facilitation: The Role of General Cognitive Resources." *Thinking & Reasoning 19* (1): 27–53. doi:10.1080/13546783.2012.713177.

Lichtenstein, S., Fischhoff, B., and Phillips, L. D 1977. "Calibration of Probabilities: The State of the Art to 1980." In D. Kahneman P. Slovic, and A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press, in press.

Lipshitz, R. 1993. "Converging Themes in the Study of Decision Making in Realistic Settings." In G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok (Eds.), *Decision*

Making in Action: Models and Methods (103–137). Norwood, NJ: Ablex Publishing.

Lipshitz, R., G. Klein, J. Orasanu, and E. Salas. 2001. "Taking Stock of Naturalistic Decision Making." *Journal of Behavioral Decision Making 14* (5): 331–352. doi:10.1002/bdm.381.

Louis, L. L., S. Moussaoui, A. Van Langhenhove, S. Ravoux, T. Le Jan, V. Roualdes, and I. Milleville-Pennel. 2023. "Cognitive Tasks and Combined Statistical Methods to Evaluate, Model, and Predict Mental Workload." *Frontiers in Psychology* 14: 1122793. doi:10.3389/fpsyg.2023.1122793.

Means, B., E. Salas, B. Crandall, and T. O. Jacobs. 1993. "Training Decision Makers for the Real World.". In G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods* (306–326). Norwood, NJ: Ablex Publishing.

Mosier, K., and U. Fischer. 2010. "Judgment and Decision Making by Individuals and Teams: Issues, Models, and Applications." In D. H. Harris (Ed.), *Reviews of Human Factors and Ergonomics* (Vol. 6, 198–256). Santa Monica, C.A: Human Factors and Ergonomics Society. doi:10.1518/155723410X12849346788822.

Mumford, M. D., E. M. Todd, C. Higgs, and T. McIntosh. 2015. "Cognitive Skills and Leadership Performance: The Nine Critical Skills." *The Leadership Quarterly* 26 (1): 43–71.

National Transportation Safety Board (NTSB). 2010. *Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River, US Airways Flight 1549, Airbus A320-214, N106US, Weehawken, New Jersey, January 15, 2009, 10(03)*. Retrieved from: http://i2.cdn.turner.com/cnn/2016/images/09/07/flight.1548.ntsb.pdf

Orasanu, J., R. K. Dismukes, and U. Fischer. 1993. "Decision Errors in the Cockpit." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 37* (4): 363–367. doi:10.1177/154193129303700423.

Orasanu, J., and L. Martin. 1998. *Errors in Aviation Decision Making: A Factor in Accidents and Incidents. HESSD 98: 2nd Workshop on Human Error, Safety, and System Development*, 100–106. Seattle, Washington, USA: Human Error, Safety, and System Development Workshop.

Orasanu, J., L. Martin, and J. Davison. 2002. "Cognitive and Contextual Factors in Aviation Accidents.". In E. Salas, and G. Klein (Eds.), *Naturalistic Decision Making*. 343–358. Mahwah, NJ: Lawrence Erlbaum.

Orasanu-Engel, J., and K. L. Mosier. 2019. "Flight Crew Decision-Making.". In B. G. Kanki, J. Anca, and T. R. Chidester (Eds.), *Crew Resource Management* 139–183. London, UK: Elsevier Academic Press. doi:10.1016/B978-0-12-812995-1.00005-1.

Paulhus, D. L., and S. Vazire. 2007. "The Self-Report Method.". In R. W. Robins, R. C. Fraley, and R. F. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology*. 224–239. New York, NY: The Guildford Press.

Pedone, R., A. Semerari, I. Riccardi, M. Procacci, G. Nicolò, and A. Carcione. 2017. "Development of a Self-Report Measure of Metacognition: The Metacognition Self-Assessment Scale (MSAS). Instrument Description and Factor Structure." *Clinical Neuropsychiatry: Journal of Treatment Evaluation 14* (3): 185–194.

Rosa, E., N. Dahlstrom, I. Knez, R. Ljung, M. Cameron, and J. Willander. 2020. "Dynamic Decision-Making of Airline Pilots in Low-Fidelity Simulation." *Theoretical Issues in Ergonomics Science 22* (1): 83–102. doi:10.1080/1463922X.2020.1758830.

Sainarayan, A. 2018. *Aviation personnel forecasts*. Paper presented at the Second ICAO NGAP Global Summit, Shenzhen, China. Retrieved from https://www.icao.int/Meetings/ngap2018/Pages/default.aspx

Schonland, A. 2016. *The pilot shortage part 2 – The supply of commercial pilots*. Retrieved from https://airinsight.com/pilot-shortage-part-2-supply-commercial-pilots/

Simon, H. A. 1987. "Making Management Decisions: The Role of Intuition and Emotion." *Academy of Management Perspectives 1* (1): 57–64. doi:10.5465/ame.1987.4275905.

Simonovic, B., E. J. N. Stupple, M. Gale, and D. Sheffield. 2017. "Stress and Risky Decision Making: Cognitive Reflection, Emotional Learning or Both." *Journal of Behavioral Decision Making 30* (2): 658–665. doi:10.1002/bdm.1980.

Simonovic, B., E. J. N. Stupple, M. Gale, and D. Sheffield. 2018. "Performance Under Stress: An Eye-Tracking Investigation of the Iowa Gambling Task (IGT)." *Frontiers in Behavioral Neuroscience* 12: 217. doi:10.3389/fnbeh.2018.00217.

Simpson, P. A. 2001. "*Naturalistic Decision Making in Aviation Environments.*" *Defence Science and Technology Organisation Victoria (Australia) Aeronautical and Maritime Research Lab. 1* : 1–28.

Sirota, M., and M. Juanchich. 2011. "Role of Numeracy and Cognitive Reflection in Bayesian Reasoning with Natural Frequencies." *Studia Psychologica 53* (2): 151–161.

Sirota, M., and M. Juanchich. 2018. "Effect of Response Format on Cognitive Reflection: Validating a Two- and Four-Option Multiple Choice Question Version of the Cognitive Reflection Test." *Behavior Research Methods 50* (6): 2511–2522. doi:10.3758/s13428-018-1029-4.

Škare, M., D. R. Soriano, and M. Porada-Rochoń. 2021. "Impact of COVID-19 on the Travel and Tourism Industry." *Technological Forecasting and Social Change 163*: 120469. doi:10.1016/j.techfore.2020.120469.

Smith, G., E. Bjerke, M. Smith, C. Christensen, T. Q. Carney, P. Craig, and M. Niemczyk. 2016. "Pilot Source Study 2015: An Analysis of FAR Part 121 Pilots Hired after Public Law 111-216—Their Backgrounds and Subsequent Successes in US Regional Airline Training and Operating Experience." *Journal of Aviation Technology and Engineering 6* (1): 64–89. doi:10.7771/2159-6670.1140.

Smith, K., J. Brown, and L. Jones. 2016. "The Impact of Flight Experience on Pilot Decision-Making Performance." *Journal of Aviation Psychology* 26 (2): 145–159. doi:10.1080/10508414.2016.1234567.

Staal, M. A. 2004. *Stress, Cognition, and Human Performance: A Literature Review and Conceptual Framework*. Moffett Field, CA: *NASA Ames Research Center*.

Stanovich, K. E. 2009. "Distinguishing the Reflective, Algorithmic, and Autonomous Minds: Is It Time for a Tri-Process Theory?." In J. S. B. T. Evans and K. Frankish (Eds.), *In Two Minds: Dual Processes and beyond*. 55–88. Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199230167.003.0003.

Strauch, B. 1997. "Automation and Decision Making—Lessons from the Cali Accident." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 41* (1): 195–199. doi:10.1177/107118139704100145.

Strauch, B. 2016. "Decision Errors and Accidents: Applying Naturalistic Decision Making to Accident Investigations." *Journal of Cognitive Engineering and Decision Making 10* (3): 281–290. doi:10.1177/1555343416654629.

Stupple, E. J. N., L. J. Ball, and D. Ellis. 2013. "Matching Bias in Syllogistic Reasoning: Evidence for a Dual-Process account from Response Times and Confidence Ratings." *Thinking & Reasoning 19* (1): 54–77. doi:10.1080/13546783.2012.735622.

Stupple, E. J. N., M. Gale, and C. Richmond. 2013. "Working Memory.", *Cognitive Miserliness and Logic as Predictors of Performance on the Cognitive Reflection Test. The 35th Annual Conference of the Cognitive Science Society*. Berlin: Humbolt University.

Tabachnick, B. G., and L. S. Fidell. 2007. *Using Multivariate Statistics* (5th ed.). Boston, MA: Allyn & Bacon/Pearson Education.

Taylor, J. L., R. O'Hara, M. S. Mumenthaler, A. C. Rosen, and J. A. Yesavage. 2005. "Cognitive Ability, Expertise, and Age Differences in following Air-Traffic Control Instructions." *Psychology and Aging* 20 (1): 117–133. doi:10.1037/0882-7974.20.1.117.

Thompson, V. A., and S. C. Johnson. 2014. "Conflict, Metacognition, and Analytic Thinking." *Thinking & Reasoning* 20 (2): 215–244. doi:10.1080/13546783.2013.869763.

Toneatto, T. 2002. "A Metacognitive Therapy for Anxiety Disorders: Buddhist Psychology Applied. ." *Cognitive and Behavioral Practice* 9: 72–78. https://doi.org/10.1016/S1077-7229(02)80043-8.

Toplak, M. E., R. F. West, and K. E. Stanovich. 2014. "Assessing Miserly Information Processing: An Expansion of the Cognitive Reflection Test." *Thinking & Reasoning 20* (2): 147–168. doi:10.1080/13546783.2013.844729.

Walmsley, S., and A. Gilbey. 2016. "Cognitive Biases in Visual Pilots' Weather-Related Decision Making." *Applied Cognitive Psychology.30*(4): 532–543. doi:10.1002/acp.3225.

Walmsley, S., and A. Gilbey. 2017. "Debiasing Visual Pilots' Weather-Related Decision Making." *Applied Ergonomics* 65: 200–208. doi:10.1016/j.apergo.2017.06.019.

Wickens, C. D. 2002. "Multiple Resources and Performance Prediction." *Theoretical Issues in Ergonomics Science 3* (2): 159–177. doi:10.1080/14639220210123806.

Wojcik, D. Z., C. J. A. Moulin, and A. Fernandez. 2021. "Assessment of Metacognition in Aviation Pilot Students during Simulated Flight Training of a Demanding Maneuver." *Applied Ergonomics* 95: 103427. doi:10.1016/j.apergo.2021.103427.

Young, R. D. 2008. *Quality of Life Indicator Systems-Definitions, Methodologies, Uses, and Public Policy Decision Making*. Columbia: Institute for Public Service and Policy Research, University of South Carolina.

Young, M., and N. Stanton. 2002. "Malleable Attentional Resources Theory: A New Explanation for the Effects of Mental Underload on Performance." *Human Factors 44* (3): 365–375. doi:10.1518/0018720024497709.

Xiao, Y. M., Z. M. Wang, M. Z. Wang, and Y. J. Lan. 2005. "The Appraisal of Reliability and Validity of Subjective Workload Assessment Technique and NASA-Task Load Index." *Zhonghua Lao Dong Wei Sheng Zhi ye Bing za Zhi = Zhonghua Laodong Weisheng Zhiyebing Zazhi = Chinese Journal of Industrial Hygiene and Occupational Diseases 23* (3): 178–181.