Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: www.tandfonline.com/journals/terg20

The effects of wearing pilots' oxygen masks on cognitive performance

Frank Albers, Julia Maria Maier, Kevin Schudlik & Jean-Baptiste Dupuy

To cite this article: Frank Albers, Julia Maria Maier, Kevin Schudlik & Jean-Baptiste Dupuy (02 Jun 2025): The effects of wearing pilots' oxygen masks on cognitive performance, Ergonomics, DOI: 10.1080/00140139.2025.2511871

To link to this article: https://doi.org/10.1080/00140139.2025.2511871

Published online: 02 Jun 2025.

ERGONOMICS
https://doi.org/10.1080/00140139.2025.2511871

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

The effects of wearing pilots' oxygen masks on cognitive performance

Frank Albers[a], Julia Maria Maier[a], Kevin Schudlik[a] and Jean-Baptiste Dupuy[b]

[a]Department of Aerospace Psychology, German Aerospace Center (DLR), Hamburg, Germany; [b]Safran Aerosystems S.A., Plaisir, France

ABSTRACT

We analysed cognitive performance of professional pilots in two visual and one acoustical cognitive task while wearing two different kinds of oxygen masks for aircraft. The study was conducted in a simulator environment with a sample of $N = 18$ French pilots. The air composition was held constant. Compared to baseline performance, there were no performance detriments in the visual tasks when accuracy was analysed. However, wearing the masks slowed down performance speed, although marginally. Also, the detriments were much smaller than the subjective feeling of inhibition as identified in an earlier study. In the acoustical task, a decrease in performance in accuracy and speed could only be observed in multiple task and therefore high workload situations. The underlying mechanisms that might cause the diverse detriments on different stages of cognitive processing are discussed as well as the practical implications, especially for enhancing compliance to professional mask wearing.

Practitioner summary: In a simulation experiment, the effects of oxygen mask wearing on cognitive performance were shown to be marginal, although pilots might state feelings of subjective impairment. The results can help to foster mask wearing in pilots.

ARTICLE HISTORY
Received 14 January 2025
Accepted 20 May 2025

KEYWORDS
Human cognitive performance; human factors design; psychological evaluation; oxygen mask; aviation

1. Introduction

Human factors and the ergonomic evaluation of technical devices have quite a long history in engineering and psychology (Sanders and McCormick 1993; Wickens et al. 2022). Especially in aviation, the human factor in design of critical technical devices is crucial and safety-relevant (J.D. Lee et al. 2018; Martinussen and Hunter 2018) as it has possible consequences for a vast number of passengers. Oxygen systems with associated masks for pilots are a substantial part of the aviation safety system. Thus, a good and thorough evaluation of such oxygen masks serving as personal protective equipment (PPE) including a broader range of aspects and factors (Wiener and Nagel 1988) is especially important (see also Lee et al. 2013, 2018, 2021). It is known that the compliance of pilots with wearing oxygen masks can be a problem (Miller 2014) and better products might be able to enhance it: PPE devices should not impair the actual work of its users and well-evaluated devices have a higher probability to be used when it is advised or mandatory.

This study is concerned with the ergonomic evaluation of oxygen masks. The systems we evaluated had already undergone a thorough evaluation of usability and comfort (Maier et al. 2022). In the present study, we examined the effects of oxygen mask wearing itself and the differential effects of different oxygen masks on human cognitive performance. The question, if wearing of oxygen masks affects cognitive performance is very crucial because the pilot's job has very high cognitive demands (Zinn, Goerke, and Marggraf-Micheel 2020). However, the methods and results might be interesting in every-day domains, too. For example, during the Covid-19 pandemic the wearing of face masks (community masks, surgical masks, or FFP-2 masks) was a big social issue. As people were obliged to wear certain mask types in several countries at particular phases of the pandemic, also while at work, the question of evaluation was raised. Several studies have been conducted into whether well-being and cognitive, work, and scholastic performance were negatively affected by mask wearing and despite some diverse results, the conclusion can be drawn that cognitive performance was not too much affected (Schlegtendal et al. 2022; Spang and Pieper 2021; van Kampen et al. 2023).

CONTACT Frank Albers frank.albers@dlr.de DLR—Deutsches Zentrum für Luft- und Raumfahrt, Sportallee 54A, Hamburg, 22335, Germany

Previous research in the evaluation of oxygen masks regarding human performance (e.g. Sausen et al. 2001; Peacock et al. 2017) has mostly dealt with the effects of air composition/oxygen saturation. It was predominantly found that hypoxia has detrimental and additional oxygen can have enhancing effects on human performance (cf. Neuhaus and Hinkelbein 2014; Scholey, Moss, and Wesnes 1998; Scholey et al. 1999). Only a few studies focused on the question if the factor of wearing an oxygen mask in itself and without varying the amount of oxygen might have an influence on the cognitive performance of pilots. Results seem somewhat ambiguous. On the one hand, for example, Caretti, Bay-Hansen, and Kuhlmann (1995) have shown that mask wearing did not lead to detrimental effects in performance. The results of Zimmerman et al. (1991) also indicated no significant effects on the performance in certain cognitive tasks, but detrimental effects on psychomotor tasks. On the other hand, Caretti (1999) himself also found negative effects of mask wearing, which he attributed to impairments in the visual field. These inconsistencies are worthwhile to be explored further.

The present research was embedded in an evaluation study of a new kind of oxygen system with associated mask for pilots in large civilian aircraft (Maier et al. 2022). This system was evaluated very positively with regard to usability and comfort, especially in comparison to a conventional, customary oxygen system with mask. However, participants' assessments of the usability of the newly developed system revealed also that the pilots were not really satisfied with the view, which was regarded as quite obstructed (cf. Caretti 1999). The mean assessed sensation of *un*obstruction of the view was $M=3.00$ ($SD=0.88$) on a 5-point scale (1=weak to 5=strong; customary mask $M=3.37$, $SD=0.97$). The feel of the wearing was also not regarded as unfetteredly good. In their evaluation, the candidates reported that the new mask put some pressure on the face, $M=3.17$ ($SD=0.85$) on a 5-point scale (1=weak to 5=strong), albeit less than the customary mask with $M=4.52$ ($SD=0.62$). The full outcome of the usability and comfort evaluation has been published elsewhere (Maier et al. 2022). All in all, these assessments suggest that the cognitive performance might suffer while wearing the mask. Possible detriments could, for example, be due to different factors like visual impairments, cognitive or attentional resources (cf. Kahneman 1973; Wickens and Carswell 2022), or reduced alertness (Steinman, Groen, and Frings-Dresen 2023).

Derived from previous research and earlier results, our research questions were as follows:

- Does the wearing of an oxygen mask for pilots in modern civilian aircraft have direct effects on cognitive performance?
- Can possible effects on cognitive performance be observed on different dependent variables (accuracy and speed)?
- If effects are found: Do they have a size of practical significance?
- Do different masks have different effects?
- Does the wearing of oxygen masks have differential effects on different tasks (visual *vs.* acoustic, different tasks structure, e.g. demanding perceptual speed *vs.* memory)?

The data collection comprised the application of three different tasks of cognitive performance. In all three tasks the dependent variables of correct answers (hits) as a measure of accuracy and reaction time as a measure of speed were analysed. We applied the measures to all participants without mask and with two different types of oxygen masks for commercial civilian aircraft. The composition of the air that was breathed by participants without and with mask was constant, it was the composition of ambient breathing air.

By this approach, our work can be seen as a kind of linkage between the human factors evaluation and the research of (individual) differences in cognitive performance. The outcome can be a theoretical contribution to the effects of mask breathing. Also, it can generate suggestions for the practice of thorough evaluation of technical devices. Finally, if generalisable effects of mask-wearing, regardless of specific mask types, are found, it can have practical consequences for aviation, e.g. for the compliance of mask-wearing in the cockpit (Miller 2014).

2. Materials and methods

The present study was conducted in accordance with the model code of ethics of the European Federation of Psychologists' Associations (https://www.efpa.eu/model-code-ethics) and the Declaration of Helsinki. Participants were informed that their data will be evaluated anonymously. They confirmed their voluntary participation and gave their informed consent in a study participant contract.

2.1. Setting

The study was conducted in the test facilities of Safran Aerosystems S.A. in Plaisir, France. The test environment consisted of two cockpit demonstrators, each containing an oxygen mask and a human-machine interface (HMI). One cockpit demonstrator was equipped with a newly developed oxygen system and a second demonstrator provided a customary oxygen system, which can be found in the same or very similar version in cockpits of large civilian widebody aircraft.

The new mask was linked to HMI screens to exchange data via a Remote Data and Power Cabinet (RDPC) gateway. The purpose of the HMI was to control and monitor aircraft navigation, perform emergency procedures if necessary, and manage mask operation. As for the fluid part, pressurised oxygen (500 mbarg) was provided to a regulator, which was supplying the right amount of air flow through the box and the mask, based on the pilot's/participant's demand. The face-piece of the new mask was designed to meet the requirements of usability and especially of comfort better than predecessors. One important aspect was that the weight of the mask was significantly reduced, due to the cessation of the steering unit. The handling of the mask and its modes had been incorporated in the HMI.

The customary mask was working purely mechanically; a regulator electronic board and RDPC were not required here. Thus, the HMI screen's utility was about navigation and emergency procedures only. The fluid part was almost the same as with the new mask: a regulator was placed within the mask and pressurised oxygen was replaced by pressurised air, which does not make a difference for experiments on the ground (Andersson et al. 2002). This system should be known to the pilots from their usual operation. The composition of the air in our experiments was normal ambient breathing air for both masks.

A summary of the defining features and the differences between the two masks is illustrated in Figure 1. Hereafter we refer to the customary system as 'Mask A' and to the newly developed system as 'Mask B'. Mask A was a customary analogue and mechanically working system and due to its construction rather heavy. Mask B worked digitally via RDPC gateway and was operated via touch screen (HMI). Its design features have resulted in a lighter weight.

The experimental procedure included the assessment of baseline measures without wearing an oxygen mask, normal cruise flight operation—while the operator wore an oxygen mask—with an autopilot, and two emergency scenarios where the use of oxygen systems is crucial. Figure 2 shows a typical test setting.

2.2. Flight tasks (cruise and emergency scenario)

To test the masks in realistic scenarios, participants had to use the HMI that displayed aeroplane instruments during flight tasks. Three flight scenarios were embedded in the study procedure (see below): first, participants had to perform a cruise flight phase: participants received standardised instructions (similar to air traffic control communication) via speakers. They had to comply with these instructions for climb, descent, and certain headings. The instructed values were to be input in the navigation panel of the HMI. Subsequently, participants worked on two fictional emergency scenarios. The first scenario mimicked a depressurisation event, the second scenario a fire, fumes, and smoke event in the cockpit. Participants were acoustically informed about the events and then had to use specialised checklists (presented on the HMI) to deal with the emergencies. After completing the checklists, the participants were instructed to fly on for the rest of the scenario time.

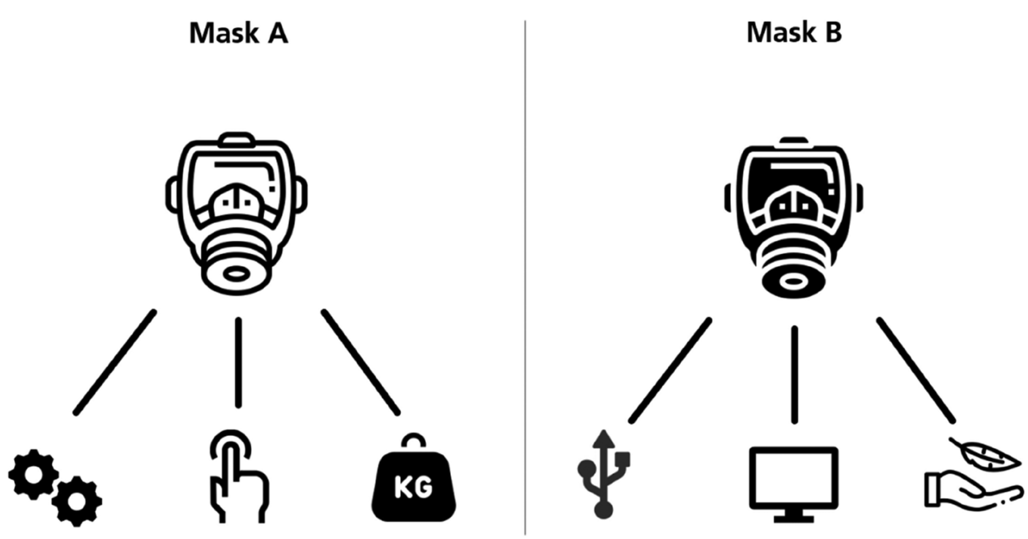

Figure 1. Schematic representation of the masks used.
Designed by Freepik (www.freepik.com).

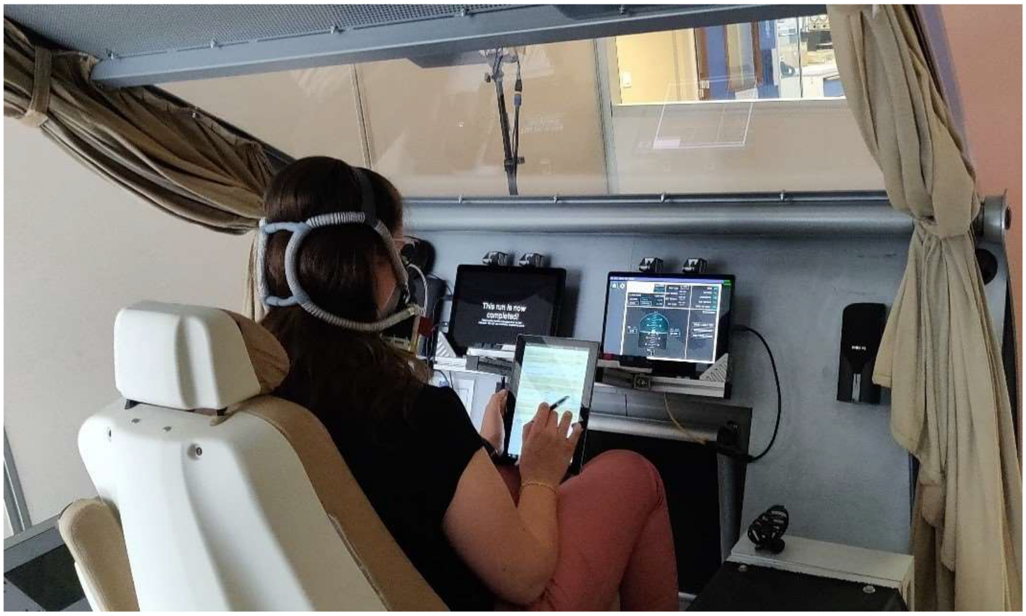

Figure 2. Rear-view of the typical test setting for the new oxygen system (participant working on a digital questionnaire).

2.3. Sample

A total of 20 French pilots were recruited as participants through contacts with French flight schools. Datasets of 18 participants (17 male and one female) were considered for the analyses. Two datasets had to be excluded because of current development works during the first experimental sessions, which led to differing conditions for the first two participants. Two-thirds ($N=12$) of the remaining participants held a commercial pilot licence (CPL) and one-third ($N=6$) held an airline transport pilot licence (ATPL). The mean age of the participants was 26.8 years ($SD=4.0$ years), ranging from 21 to 33 years.

In the analysed sample, 15 participants reported that they had experience wearing an oxygen mask in a simulator and/or an aircraft, while the remaining three participants had no previous experience with oxygen masks. The participants were paid 200€ for their participation in the study.

2.4. Measures

Three different tasks were used to assess participants' cognitive performance.

2.4.1. Deary-Liewald Task (DLT)

The Deary-Liewald reaction time task (DLT) was developed by Deary, Liewald, and Nissan (2011) and has been successfully used to measure cognitive performance in the context of pilot performance (Utamatanin and Pariwatcharakul 2022). The computerised exercise consists of two tasks: The first one is a simple reaction task, while the second is a choice reaction task. In this study only results of the choice reaction task (CRT) were used. In this task, the participant has to monitor four symmetrical, horizontally aligned white boxes. Each box is associated with a specific key (marked with different colors) on a keyboard. At random intervals (between 1 and 3 s), a black cross appears in one of the boxes and the participant has to press the corresponding key as quickly as possible. Each CRT session consisted of 20 trials.

We used two performance indicators based on the CRT: first, the percentage of correct reactions ('hits') which is an indicator of accuracy, and, second, participants' reaction times (RT, in ms) for correct reactions, which is an indicator of speed.

2.4.2. Visual Search Task (VST)

The visual search task (VST) is based on the work by Treisman and Gelade (1980) and Treisman (1982). In this conjunction search task, the participant has to decide whether a target stimulus, an upright orange letter 'T', is present among a set of distractor stimuli. Distractors are five to twenty upright blue 'T's and upside-down orange 'T's. Participants have to react by pressing the space bar of a keyboard. Half of the trials did not include the target 'T' and thus were target-absent. Each VST session consisted of 24 randomised trials.

The VST provides a measurement of accuracy and visual perceptual speed (reaction time) (cf. Wolfe 2018). Analogous to the CRT, we used the percentage of correct reactions ('hits') as an indicator of accuracy. Additionally, we used the reaction time (RT in ms) for correct reactions as the second dependent variable.

2.4.3. Acoustical Side Task (AST)

The acoustical side task (AST) is a classical acoustic memory task (Jaeggi et al. 2010) in the form of an auditory N-back task. Blundell et al. (2020) used a similar task to investigate pilots' workload under varying multi-tasking conditions during aeroplane flights.

In the AST, a continuous stream of digits between 1 and 9 is presented acoustically in English language. The presentation of each digit takes 1 s. The presentation speed is 15 digits per minute. The participant has to decide, whether the presented digit matches the one that has been presented two digits before. If yes, the participant has to press the space bar of a keyboard before the next digit is presented, if not, no reaction is to be given. The participant has a window of 3 s to respond before the next digit is presented. In 33% of the cases, the presented digits were correctly matched as described above. The number of trials was adjusted to fit the respective flight segment (e.g. 225 digits were presented during each cruise flight segment). As was the case for the DLT and VST, we used two indicators of performance for the AST: accuracy (percentage of 'hits') and speed (RT in ms).

2.4.4. Workload conditions

In the present experiment, the acoustic task was paired with the different flight tasks, which resulted in different workload demands. During the periods when the participants were working on the flight tasks, they simultaneously had to do the AST. Hereafter, we refer to this simultaneous work as *multiple task condition*. The AST can be considered a secondary task (cf. Ogden, Levine, and Eisner 1979) during these multi-tasking periods. After participants had completed the flight tasks, they performed the AST only for the remaining time within each segment. We refer to these periods as *simple task conditions* because participants were able to focus solely on the AST. Thus, the AST allows for comparisons between the conditions multiple task and simple task. Table 1 provides an overview of the measurement conditions including a categorisation regarding the resulting workload demand depending on the respective combinations.

Table 1. Overview of the measurement conditions regarding multiple task demands for the AST.

Measurement time point	Mask donned?	Flight task? (yes/no)	Multiple task condition	Workload demand
Baseline	No	No	Simple task	Low
Cruise flight	Yes	Yes	Multiple task	Medium
		No	Simple task	Low
Emergency 1	Yes	Yes	Multiple task	High
		No	Simple task	Low
Emergency 2	Yes	Yes	Multiple task	High
		No	Simple task	Low

Note. The measurement time points are listed in chronological order. The baseline was only assessed during the first part of the study.

2.5. Procedure

The experiment started with a short briefing in which participants received a brief overview, a written instruction regarding navigation, other functions of the mask interface as well as one of the two masks itself. Afterwards, the participants filled out biographical data questions on tablet computers. Next, the participants received standardised text instructions for the DLT, VST, and AST. After each instruction, a first test run (measurement time point: 'Baseline') for the respective cognitive performance task was performed. Participants then familiarised themselves with the first oxygen mask. Following this and with the mask donned, participants took the DLT and VST again (measurement time point: 'Pre-flight'). Next, they had to do the first flight task (cruise flight) while wearing the oxygen mask, and the AST was run simultaneously (measurement time point: 'Cruise'). The cruise flight phase lasted for 20 min. After the cruise flight segment ended, the participants took the DLT and VST again (measurement time point: 'After cruise flight') and then took off the mask. Next, the first emergency scenario was run for 10 min. Soon after the start, the oxygen mask had to be donned as part of the emergency procedure. Again, the AST ran during each emergency scenario (measurement time point: 'Emergency 1'). Then, the second emergency scenario with the same procedure as the first one started (measurement time point: 'Emergency 2'). Afterwards, participants had a break of ~30 min. Next, participants switched the flight simulator to undergo the cruise and emergency scenarios for a second time, this time wearing the mask not worn before. The participants did not repeat the baseline measures (i.e. participants did not take the tests without wearing a mask again). Thus, half of the participants started with the Mask A, and half of the participants started with Mask B. The assignment was randomised to balance out potential effects of sequence. Overall, the study took ~3 h per participant.

2.6. Statistical analyses

The software IBM SPSS Statistics 26 was used to analyse the data. We conducted a two-way repeated-measures ANOVA (rm ANOVA) with mask and measurement time point as within-subject factors. The dependent variables were percentage of correct hits (accuracy) and reaction time. The rm ANOVAs were used to determine differences in accuracy or speed due to the type of mask a participant had worn, due to the measurement time point, or an interaction of mask and measurement point. The associated *F*-statistic indicated the statistical significance depending on the degrees of freedom. For all statistical analyses, an alpha-error of 5% was used for significance testing.

3. Results

3.1. Visual perception tasks (DLT and VST)

Results from both visual perception tests (DLT and VST) were considered for the analyses. For each test, two dependent variables, i.e. hits and reaction times, were analysed separately. Generally, the percentage of hits was very high for both tests at all measurement time points and for both masks. It ranged from 95 to 100% for the DLT choice task and from 92 to 96% for the VST. Thus, items were rather easy to solve correctly for the participants.

For a start, an rm ANOVA was performed for hits in the DLT choice task. No meaningful effect was found for the type of mask that was worn [$F(1, 17) = 0.06$, $p = .816$]. The effect of the measurement time point was only scarcely not significant [$F(2, 34) = 3.16$, $p = .055$]. Inner subject contrasts indicated a significant performance decline between baseline and pre-flight [$F(1, 17) = 9.06$, $p = .008$, $\eta^2 = .35$] but not between baseline and after cruise flight [$F(1, 17) = 1.7$, $p = .210$], where performance with Mask B partly recovered. Performance losses when wearing a mask (considering the mean value of performance for both masks and both measurement time points with mask on) compared to the baseline without mask were 1.3%. There was no interaction between mask and measurement time point. Mean values for both masks are illustrated in Figure 3.

Performing an rm ANOVA for the hits of the second visual perception task, the VST yielded similar albeit even more coherent non-significant results for performance accuracy (Figure 4): again, there was no effect of the type of mask [$F(1, 17) = 0.14$, $p = .717$] and in this case, there was clearly no effect for the measurement time point [$F(2, 34) = 0.79$, $p = .462$] and no interaction between mask and measurement time point. Inner subject contrasts comparing the two measurement time points with the mask on and the baseline indicated no performance changes depending on wearing a mask.

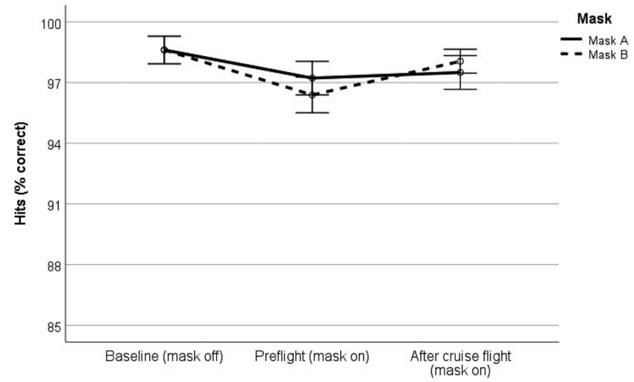

Figure 3. Hits for both masks and all measurement time points in the DLT choice task. Error bars ±1 *SD*.

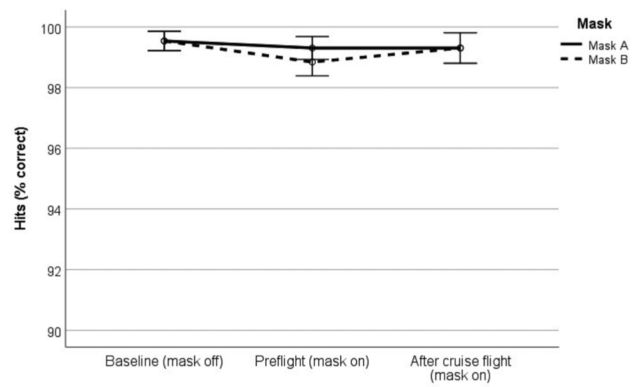

Figure 4. Hits for both masks and all measurement time points in the VST. Error bars ±1 *SD*.

For the second dependent variable, reaction time, further rm ANOVAs were performed including the factors mask × measurement time point. Regarding reaction times in the DLT choice task, like before, there was no meaningful effect for the type of mask that was worn [$F(1, 17) = 1.21$, $p = .286$] and mean values had a similar pattern for both masks. However, as illustrated in Figure 5, analyses yielded significant performance differences between the three measurement time points: Reaction times while wearing a mask, i.e. during pre-flight and after cruise flight, were significantly slower than during the baseline assessment [$F(2, 34) = 7.75$, $p = .002$, $\eta^2 = .31$]. However, performance losses when wearing a mask were rather small (4.6%). There was no meaningful interaction between mask and measurement time point.

Similar results were obtained for the second visual perception test, the VST. Rm ANOVA results indicated that reaction times did not differ depending on the type of mask [$F(1, 17) = .16$, $p = .696$], but they differed depending on the measurement time point [$F(2, 34) = 5.09$, $p = .012$, $\eta^2 = .23$]. Again, reaction times were significantly slower when a mask was worn (Figure 6) compared to the baseline, when pilots did not wear a mask. However, with performance losses of 6.9%, the slowdown was also rather small. The interaction between both variables was not significant.

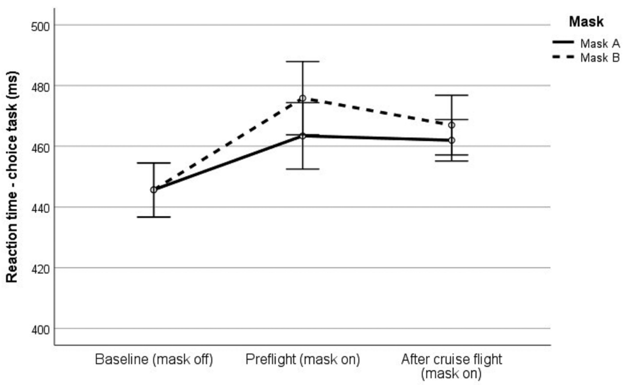

Figure 5. Reaction times (ms) for both masks and all measurement time points in the DLT choice task. Error bars ±1 *SD*.

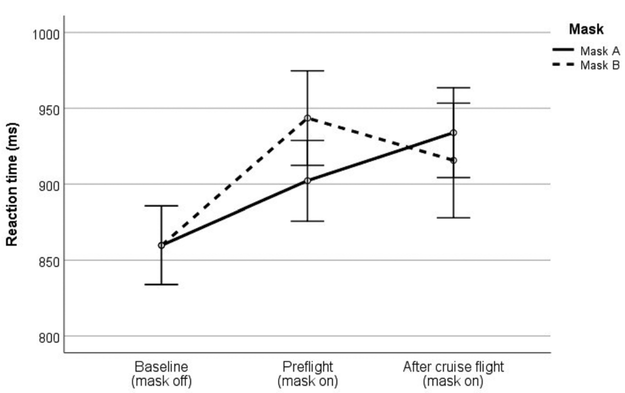

Figure 6. Reaction times (ms) for both masks and all measurement time points in the VST. Error bars ±1 *SD*.

3.2. Acoustical memory task (AST)

For the analysis of cognitive performance in the acoustical memory task, additional two-way rm ANOVAs including the factors mask x measurement time point were calculated. Analyses were performed separately for *simple* and *multiple* task conditions (see Table 1).

Like for the visual tasks, the type of mask had no effect on the pilots' acoustical performance in the *simple task* condition, neither regarding hits [$F(1, 16) = 0.06$, $p = .805$; Figure 7 left] nor regarding reaction times [$F(1, 16) = 0.55$, $p = .470$; Figure 8 left]. Moreover, the effect of the measurement time point was not significant for the percentage of hits [$F(3, 48) = 2.23$, $p = .097$] in the simple task condition and no significant interactions occurred (Figure 7 left). Looking at possible effects of the measurement time point on reaction times, also no significant effect was found [$F(3, 48) = 0.35$, $p = .786$; Figure 8 left].

Finally, two rm ANOVAs were performed for the analysis of mask effects on acoustical performance under *multiple task* conditions. There was no effect of the type of mask on the amount of hits [$F(1, 16) = 3.89$, $p = .066$, Figure 7 right], and there was no effect for the type of mask on reaction speed [$F(1, 15) = 0.94$, $p = .348$, Figure 8 right]. However, the measurement time point had a strong effect on the accuracy of the acoustical performance, and performance declined significantly with both masks over the course of time [$F(3, 48) = 48.73$, $p = .000$, $\eta^2 = .75$; Figure 7 right]. Performance losses compared to the baseline were 27.4%. Paralleling this, reaction times were significantly slower in the emergency scenarios compared to performance in the baseline and cruise flight [$F(3, 45) = 9.48$, $p = .000$, $\eta^2 = .39$; Figure 8 right]; performance losses had a height of 29%.

Moreover, a significant interaction between mask and measurement time point occurred for hits

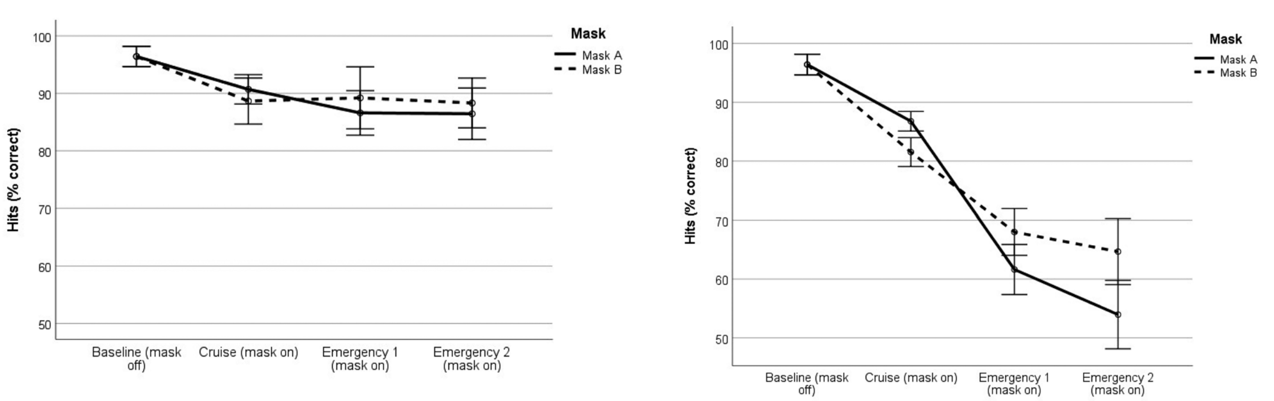

Figure 7. Hits for both masks under (left) simple task conditions and under (right) multi-task conditions in the AST. Error bars ±1 *SD*.

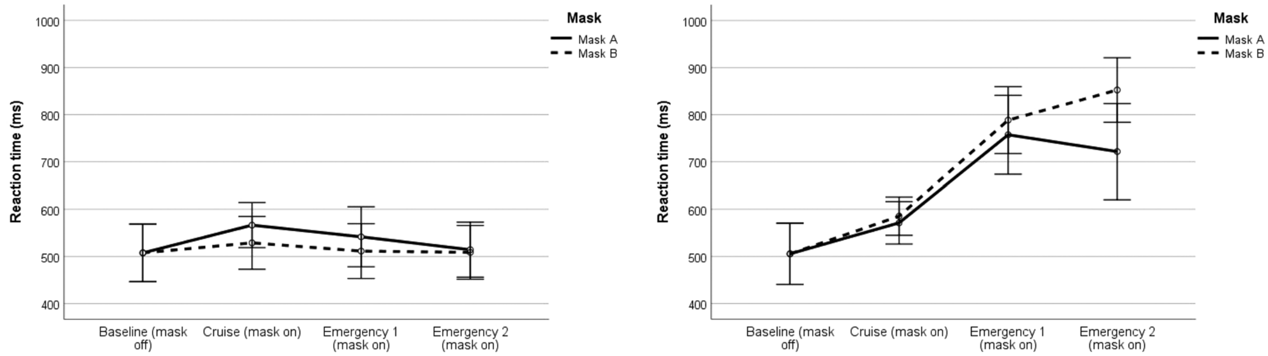

Figure 8. Reaction times (ms) for both masks and under (left) simple task conditions and under (right) multi-task conditions in the AST. Error bars ±1 *SD*.

[$F(3, 48) = 3.22$, $p = .031$, $\eta^2 = .17$]. This indicated that performance with the masks differed significantly depending on the measurement time point. Even if during cruise flight, acoustical memory performance was more accurate with Mask A, it deteriorated steeper over the time so that as a consequence, during the second emergency scenario, performance with Mask B was significantly better than with Mask A (Figure 7 right). However, there was no significant interaction between mask and measurement time point with regard to reaction times [$F(3, 45) = .92$, $p = .437$].

4. Discussion

We applied different cognitive tasks to pilot participants who either wore no mask or one of two different types of pilot's oxygen masks as PPE designed for widebody civilian aircraft. The two different masks and oxygen systems represented either the customary standard in large aircraft today (Mask A) or the prototype of a newly developed system (Mask B) which has a certain likelihood to represent a future PPE-standard in commercial aviation. By using these two systems our results have a high ecological validity and should be generalisable to today's and tomorrow's world of civil aviation. The breathing air composition in the experiments was constant and was ambient breathing air. In a previous study, a sample of pilots had evaluated the new Mask B better with regard to usability and comfort.

An important result of the current study was that, overall, there was no differential effect of the mask type in the two visual tasks we used: When effects of mask wearing occurred on human performance, then similarly with both mask types. Thus, for both masks, the performance measured as correct solutions (hits) in the DLT did not differ or differed only slightly in the respective measurement time points. One could conclude that it does not matter if a person is wearing a mask or not; or the conclusion could be that participants get used to the mask over time and quickly return to their baseline performance accuracy so that there is only a short phase of adverse effects here (measurement time point pre-flight). However, when we looked closer at the performance and examined the speed of performance, i.e. the reaction time (of hits), we identified differences between the measurement time points. These differences showed a pattern: if a participant wore an oxygen mask, the performance was weaker, i.e. reaction times were slower. This means that there were performance detriments caused by wearing oxygen masks. However, this deterioration of performance was very small with 4.6%. A similar pattern of results could be observed with the other visual task, the VST. The analyses of hits showed that there were no significant differences between measurement time points, i.e. there were no observable performance detriments in accuracy when participants wore a mask. Again, when we analysed the reaction times (of hits), performance differences became visible. The reaction times were higher in both situations when a mask was worn compared to baseline performance. Also, in this task, the detriments were minimal (with 6.9%).

Our findings regarding the accuracy correspond with other studies where the accuracy of performance was analysed (e.g. Zimmerman et al. 1991). Moreover, performance detriments on reaction times have been found in experiments with oxygen masks before (e.g. Caretti 1999). From this point of view, our study integrates in previous research despite its sometimes-contradictory findings. We suppose that the slightly longer reaction times result from a longer information processing time. The vision and sensation as the first steps in the information processing chain are affected by wearing the mask. In the model of information processing by Wickens and Carswell (2022), for example, the stages of sensory processing or the later stage of perception might be inhibited by mask wearing and thus prolong the whole information processing. The processing steps after the sensation and perception—e.g. the cognitive processing—might not be impaired and (because of this) the accuracy is not affected.

In sum, we can answer our research questions for the two visual tasks as follows: mask wearing has no effect on cognitive performance when accuracy is considered. When looking at speed in performance there are detriments. However, they are very small and far away from any practical significance. The size of the adverse effects can easily be accepted, especially when traded off with the plus of safety the wearing of oxygen masks in a cockpit generally brings.

The pattern of performance in the acoustical memory task (AST) differed in comparison to the visual tasks. Here, the accuracy as well as the speed of reaction were not impaired by wearing either of the two types of mask. This can be an indication that the small detriments in speed found in the visual tasks DLT and VST might in fact have to do with the visual sensation or perception of the stimuli (cf. Caretti 1999). As there were no visual stimuli in the AST, no detriments occurred at all. But when we analysed the performance in the AST under consideration of a higher amount of workload (Damos 1984; Xie and Salvendy 2000), i.e. memory task and flying task/dealing with emergencies at the same time, we have seen that there was a substantial detriment in cognitive performance. These detriments could be seen in accuracy (a deterioration of 27.4%) and speed (deterioration of 29%). This in sum shows that performance detriments in the AST were caused most probably in the later stages of information processing following the sensory processing and perception. Also, as shown in earlier studies these stages might be more prone to be affected by different saturations of the breathing air (e.g. Peacock et al. 2017; Varis, Parkkola, and Leino 2019).

In sum, the answer to the research questions for the acoustical task is that there are no direct effects of mask wearing on cognitive performance. The mask types do not differ in this regard. We found performance detriments of substantial quantity but they were due to effects of workload, not mask wearing.

In reference to our earlier findings (Maier et al. 2022), it has to be noted that the subjectively felt inhibition of participants, especially of the vision, by wearing a mask, is in no way mirrored by the performance in the visual cognitive tasks. This is similar to the results for subjective feelings while wearing a face mask and the (small) effects on performance during the Covid-19 pandemic (Schlegtendal et al. 2022; van Kampen et al. 2023).

On the whole, it can be stated that with the substantial advantages in usability and comfort reported previously (Maier et al. 2022), the newly developed mask system (Mask B) entails significant improvements without compromising on the pilots' cognitive performance.

Next to the theoretical insights of how human cognitive performance can or cannot be affected by wearing oxygen masks, a practical implication of our findings is that, in flight or other work situations with oxygen masks, the slight cognitive detriments that might be caused by wearing the mask could be compensated by a higher proportion of oxygen in the airflow (Moss, Scholey, and Wesnes 1998). However, this implication should be substantiated by further research. For the moment, one practical implication is that confronting pilots with results like ours, i.e. although there might be a subjective feeling of inhibition, the performance will not suffer, can help to increase pilots' low willingness to wear a mask and increase mask-wearing compliance when it is mandatory or advisable (Miller 2014). Our findings also yield a practical outcome for new developments in oxygen systems: With the changes for better usability and comfort there is no cost associated on the performance side. This leads to a promising future for the new ergonomic developments: pilots of widebody aircraft will have the same prerequisites for cognitive performance plus a lot more usability and comfort of the PPE device.

However, there should be more research, to see if our results are generalisable to an even greater variety of cognitive tasks and also to explore underlying mechanisms further. For example, AlGhamri, Murray, and Samaranayake (2013) used mental arithmetic items and have shown that mask wearing had effects on cognitive measures without a visual component. This might put another complexion on the matter. The analysis of more demanding cognitive measures (with and without visual perception) could bring new insights into the cognitive stages that are affected by mask wearing and by the interaction of task complexity/workload and oxygen mask wearing. This, again, can be fruitful to ergonomic design of new devices, that are meant to protect working people without inhibiting or maybe even enhancing their cognitive performance.

Acknowledgements

The authors thank Prof. Ralf Anken from the Institute of Aerospace Medicine (DLR) for valuable comments on an earlier version of this article.

Disclosure statement

In accordance with Taylor & Francis policy and our ethical obligation as researchers, the authors are reporting that one of the co-authors is employed at the company that manufactures oxygen masks for aircraft and at which the study was conducted. However, the author was not paid for this piece of research and our research was not affected by this involvement. We have disclosed those interests fully to Taylor & Francis.

Funding

This project has received funding from the Clean Sky 2 Joint Undertaking (JU) under grant agreement No. 945583, CS2-LPA-GAM-2020-2023-01. The JU receives support from the European Union's Horizon 2020 research and innovation program and the Clean Sky 2 JU members other than the Union.

ORCID

Frank Albers http://orcid.org/0000-0003-3043-5235
Julia Maria Maier http://orcid.org/0000-0003-1387-7939
Kevin Schudlik http://orcid.org/0000-0003-0479-6782

Data availability statement

Due to the nature of the research, the underlying data are located at the authors and are not publicly available. Data can be requested directly from the authors.

References

AlGhamri, A.A., S.L. Murray, and V.A. Samaranayake. 2013. "The Effects of Wearing Respirators on Human Fine Motor, Visual, and Cognitive Performance." *Ergonomics* 56 (5): 791–802. doi:10.1080/00140139.2013.767383.

Andersson, J., P. Berggren, M. Grönkvist, S. Magnusson, and E. Svensson. 2002. "Oxygen Saturation and Cognitive Performance." *Psychopharmacology* 162 (2): 119–128. doi:10.1007/s00213-002-1077-3.

Blundell, J., S. Scott, D. Harris, J. Huddlestone, and D. Richards. 2020. "Workload Benefits of Colour Coded Head-up Flight Symbology during High Workload Flight." *Displays* 65: 101973. doi:10.1016/j.displa.2020.101973.

Caretti, D.M. 1999. "Signal Detection Capability during Uninterrupted Full Face-Piece Respirator Wear." *Ergonomics* 42 (2): 376–384. doi:10.1080/001401399185720.

Caretti, D.M., L.A. Bay-Hansen, and W.D. Kuhlmann. 1995. "Cognitive Performance during Respirator Wear in the Absence of Other Stressors." *American Industrial Hygiene Association Journal* 56 (8): 776–781. doi:10.1080/15428119591016593.

Damos, D.L. 1984. "Individual Differences in Multiple-Task Performance and Subjective Estimates of Workload." *Perceptual and Motor Skills* 59 (2): 567–580. doi:10.2466/pms.1984.59.2.567.

Deary, I.J., D. Liewald, and J. Nissan. 2011. "A Free, Easy-to-Use, Computer-Based Simple and Four-Choice Reaction Time Programme: The Deary-Liewald Reaction Time Task." *Behavior Research Methods* 43 (1): 258–268. doi:10.3758/s13428-010-0024-1.

Jaeggi, S.M., M. Buschkuehl, W.J. Perrig, and B. Meier. 2010. "The Concurrent Validity of the N-Back Task as a Working Memory Measure." *Memory* 18 (4): 394–412. doi:10.1080/09658211003702171.

Kahneman, D. 1973. *Attention and Effort*. Prentice Hall.

Lee, J.D., C.D. Wickens, Y. Liu, and L.N. Boyle. 2018. *Designing for People: An Introduction to Human Factors Engineering*. 3rd ed. CreateSpace.

Lee, W., D. Jung, S. Park, H. Kim, and H. You. 2021. "Development of a Virtual Fit Analysis Method for an Ergonomic Design of Pilot Oxygen Mask." *Applied Sciences* 11 (12): 5332. doi:10.3390/app11125332.

Lee, W., J. Jeong, J. Park, E. Jeon, H. Kim, D. Jung, S. Park, and H. You. 2013. "Analysis of the Facial Measurements of Korean Air Force Pilots for Oxygen Mask Design." *Ergonomics* 56 (9): 1451–1464. doi:10.1080/00140139.2013.816376.

Lee, W., X. Yang, D. Jung, S. Park, H. Kim, and H. You. 2018. "Ergonomic Evaluation of Pilot Oxygen Mask Designs." *Applied Ergonomics* 67: 133–141. doi:10.1016/j.apergo.2017.10.003.

Maier, J., F. Albers, V. Oubaid, M. Fromage, and J.B. Dupuy. 2022. "Evaluation of a Next Generation Oxygen System—Assessment of Usability, Comfort and Human Performance." *Transportation Research Procedia* 66: 97–108. doi:10.1016/j.trpro.2022.12.011.

Martinussen, M., and D.R. Hunter. 2018. *Aviation Psychology and Human Factors*. CRC Press; Taylor & Francis Group.

Miller, R. 2014. *4 CFR 91.211–Associated Risks Due to Compliance. A White Paper Reviewing Risks Induced by Pilots' Routine Usage of Supplemental Oxygen Masks during High Altitude, Long Range Flights*. West Trenton.

Moss, M., A. Scholey, and K. Wesnes. 1998. "Oxygen Administration Selectively Enhances Cognitive Performance in Healthy Young Adults: A Placebo-Controlled Double-Blind Crossover Study." *Psychopharmacology* 138 (1): 27–33. doi:10.1007/s002130050641.

Neuhaus, C., and J. Hinkelbein. 2014. "Cognitive Responses to Hypobaric Hypoxia: implications for Aviation Training." *Psychology Research and Behavior Management* 7: 297–302. doi:10.2147/PRBM.S51844.

Ogden, G.D., J.M. Levine, and E.J. Eisner. 1979. "Measurement of Workload by Secondary Tasks." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 21 (5): 529–548. doi:10.1177/001872087902100502.

Peacock, C.A., R. Weber, G.J. Sanders, Y. Seo, D. Kean, B.S. Pollock, K.J. Burns, M. Cain, P. LaScola, and E.L. Glickman. 2017. "Pilot Physiology, Cognition and Flight Performance during Flight Simulation Exposed to a 3810-m Hypoxic Condition." *International Journal of Occupational Safety and Ergonomics* 23 (1): 44–49. doi:10.1080/10803548.2016.1234685.

Sanders, M.S., and E.J. McCormick. 1993. *Human Factors in Engineering and Design*. McGraw Hill.

Sausen, K P., M T. Wallick, B. Slobodnik, J M. Chimiak, E A. Bower, M E. Stiney, and J B. Clark. 2001. "The Reduced Oxygen Breathing Paradigm for Hypoxia Training: Physiological, Cognitive, and Subjective Effects." *Aviation, Space, and Environmental Medicine* 72 (6): 539–545. https://pubmed.ncbi.nlm.nih.gov/11396560/.

Schlegtendal, A., L. Eitner, M. Falkenstein, A. Hoffmann, T. Lücke, K. Sinningen, and F. Brinkmann. 2022. "To Mask or Not to Mask—Evaluation of Cognitive Performance in Children Wearing Face Masks during School Lessons (MasKids)." *Children* 9 (1): 95. doi:10.3390/children9010095.

Scholey, A.B., M. Moss, and K. Wesnes. 1998. "Oxygen and Cognitive Performance: The Temporal Relationship between Hyperoxia and Enhanced Memory." *Psychopharmacology* 140 (1): 123–126. doi:10.1007/s002130050748.

Scholey, A.B., M. Moss, N. Neave, and K. Wesnes. 1999. "Cognitive Performance, Hyperoxia, and Heart Rate Following Oxygen Administration in Healthy Young Adults." *Physiology & Behavior* 67 (5): 783–789. doi:10.1016/S0031-9384(99)00183-3.

Spang, R.P., and K. Pieper. 2021. "The Tiny Effects of Respiratory Masks on Physiological, Subjective, and Behavioral Measures under Mental Load in a Randomized Controlled Trial." *Scientific Reports* 11 (1): 19601. doi:10.1038/s41598-021-99100-7.

Steinman, Y., E. Groen, and M. Frings-Dresen. 2023. "Hypoxia Impairs Reaction Time but Not Response Accuracy in a Visual Choice Reaction Task." *Applied Ergonomics* 113: 104079. doi:10.1016/j.apergo.2023.104079.

Treisman, A. 1982. "Perceptual Grouping and Attention in Visual Search for Features and for Objects." *Journal of Experimental Psychology. Human Perception and Performance* 8 (2): 194–214. doi:10.1037/0096-1523.8.2.194.

Treisman, A., and G. Gelade. 1980. "A Feature-Integration Theory of Attention." *Cognitive Psychology* 12 (1): 97–136. doi:10.1016/0010-0285(80)90005-5.

Utamatanin, N., and P. Pariwatcharakul. 2022. "The Effect of Caffeine and Sleep Quality on Military Pilot Students' Flight Performance-Related Cognitive Function." *The International Journal of Aerospace Psychology* 32 (2–3): 152–164. doi:10.1080/24721840.2022.2034505.

van Kampen, V., E.M. Marek, K. Sucker, B. Jettkant, B. Kendzia, B. Strauß, M. Ulbrich, A. Deckert, H. Berresheim, C. Eisenhawer, F. Hoffmeyer, S. Weidhaas, T. Behrens, T. Brüning, and J. Bünger. 2023. "Influence of Face Masks on the Subjective Impairment at Different Physical Workloads." *Scientific Reports* 13 (1): 8133. doi:10.1038/s41598-023-34319-0.

Varis, N., K.I. Parkkola, and T.K. Leino. 2019. "Hypoxia Hangover and Flight Performance After Normobaric Hypoxia Exposure in a Hawk Simulator." *Aerospace Medicine and Human Performance* 90 (8): 720–724. doi:10.3357/AMHP.5289.2019.

Wickens, C.D., and C.M. Carswell. 2022. "Information Processing." In *Handbook of Human Factors*, edited by G. Salvendy and W. Karwalski. CRC Press.

Wickens, C.D., W. Helton, J.G. Hollands, and S. Banburry. 2022. *Engineering Psychology and Human Performance*. 5th ed. Routledge.

Wiener, E.L., & Nagel, D.C. (Eds.). 1988. *Human Factors in Aviation*. 2nd ed. Elsevier.

Wolfe, J.M. 2018. "Visual Search." In *Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience*, edited by J.T. Wixted, 4th ed., 569–424. Wiley & Sons. doi:10.1002/9781119170174.epcn213.

Xie, B., and G. Salvendy. 2000. "Prediction of Mental Workload in Single and Multiple Tasks Environments." *International Journal of Cognitive Ergonomics* 4 (3): 213–242. doi:10.1207/S15327566IJCE0403_3.

Zimmerman, N.J., C. Eberts, G. Salvendy, and G. McCabe. 1991. "Effects of Respirators on Performance of Physical, Psychomotor and Cognitive Tasks." *Ergonomics* 34 (3): 321–334. doi:10.1080/00140139108967316.

Zinn, F., P. Goerke, and C. Marggraf-Micheel. 2020. "Selecting for Cockpit Crew." In *Pilot Selection: Psychological Principles and Practice*, edited by R. Bor, C. Eriksen, T.L. Hubbard, and R. King, 21–34. CRC Press; Taylor & Francis Group.