



Selection of Norwegian police drone operators: an evaluation of selected cognitive tests from “The Vienna Test System”

Bjørn Helge Johnsen, Andreas Aalstad Nilsen, Sigurd W. Hystad, Eyvind Grytting, Jørgen Lunde Ronge, Steinar Rostad, Peter Henrik Öhman & Arne Jon Overland

To cite this article: Bjørn Helge Johnsen, Andreas Aalstad Nilsen, Sigurd W. Hystad, Eyvind Grytting, Jørgen Lunde Ronge, Steinar Rostad, Peter Henrik Öhman & Arne Jon Overland (2023): Selection of Norwegian police drone operators: an evaluation of selected cognitive tests from “The Vienna Test System”, Police Practice and Research, DOI: [10.1080/15614263.2023.2179052](https://doi.org/10.1080/15614263.2023.2179052)

To link to this article: <https://doi.org/10.1080/15614263.2023.2179052>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 15 Feb 2023.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)



Selection of Norwegian police drone operators: an evaluation of selected cognitive tests from “The Vienna Test System”

Bjørn Helge Johnsen^a, Andreas Aalstad Nilsen^b, Sigurd W. Hystad^c, Eyvind Grytting^b, Jørgen Lunde Ronge^b, Steinar Rostad^d, Peter Henrik Öhman^e and Arne Jon Overland^d

^aCentre for Crisis Psychology, University of Bergen, Bergen, Norway; ^bNorwegian Police Air Support Unit, Oslo Police District, Oslo, Norway; ^cDepartment of Psychosocial Science, University of Bergen, Bergen, Norway; ^dThe Norwegian Police University College, Oslo, Norway; ^eThe Norwegian Police University College, Stavern, Norway

ABSTRACT

A nationwide sample of 129 police officers participated in a study aimed at validating and presenting practical implications of a selection procedure for applicants to an educational program for Norwegian police drone pilots. The subjects were part of a selection program for a training and qualification course for police drone pilots. The selection program consisted of tests of spatial orientation, logical reasoning, attentional selection, sustained attention, and visual short-term memory, in addition to a performance test in a drone flight simulator. The aim of the study was to evaluate the cognitive tests used in the selection program and their relation to performance during the simulated flight. The results from the untrained applicants revealed low-to-moderate intercorrelations of the cognitive tests. Only spatial orientation, logical reasoning, and attentional selection were correlated to the performance measures of skills and proficiency. Stepwise regression analysis showed that only spatial orientation and attentional selection had unique contributions in explaining the variance in both measures of performance. Implications are discussed on both practical and scientific levels. The positive implications of using untrained respondents, the use of proficiency measures in addition to skills, the building of a clearinghouse for drone selection data, and considering both the job-analyses and the total test-performance when interpreting the applicant's test-scores are discussed.

ARTICLE HISTORY

Received 21 June 2022
Accepted 6 February 2023

KEYWORDS

Drone pilot selection;
cognitive test battery;
performance

Introduction

Both civilian and military uses of unmanned aerial systems (i.e., drones) are increasing. Civilian use of drones includes commercial use as part of an industry, use as a leisure activity by private citizens, and use in the public domain by police and research organizations. Furthermore, a proposed ‘innovation zone’ by the FAA for drone companies’ testing of technology revealed around 2,000 interested parties even before the deadline for applications to the pilot program (Ravindranath, 2017). Aydin (2019) reported that the public’s perception of drone usage was positively evaluated for public safety and research use, while commercial and private practice was not. Police drones are versatile technical means of power that could create psychological challenges for the user. For instance, Rauch and Ansari (2022) claimed that emerging technologies have a profound effect on operator’s moral values and their emotional responses to their work. More specifically, the authors report a distanced intimacy in which the sense of psychological and physical distance is broken

CONTACT Bjørn Helge Johnsen  bjorn.johnsen@uib.no  Centre for Crisis Psychology, University of Bergen, Møllendalsbakken 9, Bergen 5009, Norway

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

down. New usage of technology often provokes a change in the legal framework the technology operates in as well as a pressure for redefining old work ethics (Rauch & Ansari, 2022).

All operations in the aerial domain involve risks, and about 80% to 85% of military mishaps could be traced back to human errors (M. L. Thomas & Russo, 2007; Lei et al., 2014). The use of drones could result in conflict with other aerial operations and risk on the part of the operator, members of the public, and material on the ground. There are some unique challenges that apply when piloting a drone compared to a manned aircraft. One major difference is the information loss caused by being geographically separated from the vehicle. Lack of feedback from natural senses, such as smell and vision, kinetics, vibration, and movement, which often provide representations of weather conditions, increases the challenges faced by drone operators. The separation from the vehicle creates an increased taxing of the cognitive resources of the pilot, while the physical demands decrease (Qia et al., 2018). Another unique challenge compared to manned aviation suggests that having a drone pilot physically removed from the aircraft can lead to operational risks being perceived as less severe due to the increased distance from the situation. This has been found for both geographical and perceived distances in other domains (O'neill et al., 2016). Thus, in order to correctly level-set the operational risks, the pilot must overcome this phenomenon by relying on logical reasoning to evaluate risks (see Slovic et al., 2004, for a critical discussion of 'risk as analyses'). Taken together, the use of drones by official entities in commercial, research, and recreational practice could cause adverse effects among the public, for the drone operator, and for the equipment.

One way to potentially minimize the risks and maximize the operational effects of drones is by implementing a selection procedure for drone pilots (Armour & Ross, 2017; Fraser, 2020; Graham et al., 2021; Zanetti et al., 2022). By defining desired characteristics of drone operators and selecting subjects showing acceptable levels of the desired attribute (i.e., positive selection), the potential for adverse effects during drone operations could be lowered. A similar argument would be present if the selection program has the possibility to reveal unwanted characteristics of an applicant (i.e., negative selection). An illustration of an ill-selected drone pilot would be a pilot lacking in several cognitive abilities limiting their ability to perform tasks that are important for the drone operation to be carried out safely and efficiently. Limited spatial ability could lead to drones being inadvertently flown over densely populated areas, putting people on the ground at unnecessary risk. Otherwise, it might result in the pilot not being able to intuitively understand the directions of moving objects on the ground in relation to cardinal directions – inhibiting the drone pilot's ability to give directions and communicate effectively to other police units. Furthermore, an ill-selected drone pilot with limited ability for logical reasoning might fail to find suitable risk mitigations even though risk factors have successfully been identified – for instance if a pilot is flying in low light environments, they might have identified limited visibility as a risk factor and the fact that the drone's visual obstacle avoidance system will have poor performance – but they might not logically infer that lack of obstacle avoidance systems puts higher emphasis on the drones' other internal sensors (i.e., Global Navigation Satellite Systems/GNSS, barometer, accelerometer, gyro, and compass) to keep the drone from inadvertently drifting into buildings or other obstacles – this would mean that the pilot might not consider risk mitigating measures like checking for GNSS satellite coverage, evaluating the flight route with regard to GNSS shadows behind buildings and mountains, or checking weather forecast for solar wind activity KP index to make sure charged particles from solar activity does not impair GNSS signal quality.

A well-selected pilot, on the other hand, would be a pilot with high cognitive abilities, able to perform tasks to ensure the drone operation is carried out safely and efficiently. For instance, a drone pilot with good spatial ability would be able to keep track of the drone's position in three-dimensional space, the direction of the camera sensor and of moving cars, and people on the ground in relation to each other. Said pilots would be able to maneuver the aircraft safely and at the same time maintain an understanding of the events that occur on the ground to effectively support the police mission with relevant and precise information. A well-selected drone pilot would have good

visual memory, being able to remember visual information that is no longer available as the camera sensor pans through an area. This would enable the pilot to keep references to buildings or other points of interest that are no longer present in the video stream. Furthermore, the drone pilot would also have good logical reasoning abilities necessary to identify risk factors and logically infer the relevant mitigations. For instance, if they perform a drone operation in high wind conditions, one expects a well-selected pilot to logically infer the implications of wind direction in the operation, taking into account the need for extra battery capacity needed to fly back to the landing site in situations where the drone would face strong headwinds on the return flight path. The drone pilot would also be expected to identify the need for a lower predefined return to home (RTH) altitude at a loss of link situation, as the winds tend to be less intense at lower altitudes.

Norwegian police are currently implementing a selection program for applicants to a bachelor-level course for qualification as police drone operators. Although selection and educational programs are labor intensive, responses from potential applicants and local police districts lend support to the arguments that the role as a police drone operator is viewed as an exciting and attractive new position using a novel and tactical asset. To our knowledge, this is the first law enforcement organization in Europe to implement such an extensive selection program involving neuropsychological tests in the selection of drone pilots. Thus, the aim of this study was to evaluate the cognitive tests used in the program.

It has been suggested (Qia et al., 2018) that qualification requirements for drone pilots should encompass professional quality, medical requirements, psychological evaluation, training requirements, operating experience, and coordination (systemic cooperation training with other air and ground vehicles). Psychological evaluations are often described in terms of personality profiles and clinical disorders (Chapelle et al., 2014). However, in order to maximize the operational effect of drones, pilot selection should also focus on cognitive capacity and function. For instance, Zanetti et al. (2022) stated that performance fluctuation in drone operations was caused by limited cognitive resources.

According to Qia et al. (2018), rational thinking is one of the professional aptitudes recommended for drone pilots. It could be argued that this cognitive ability increases a pilot's ability to detect the risk factors involved in operating the drone and to apply appropriate risk mitigating measures to ensure safe operation. Drone operations also involve multitasking, where the pilot has to control the vehicle, and, at the same time, both monitor visual sensor data and communicate relevant information in relation to the aim of the mission. A typical consequence of multitasking is the curtailment of the performance of the specific tasks involved. Multitasking could also be a result of drone pilots management of uncertainty (Chérif et al., 2018). Low levels of multitasking could lead to inappropriate decisions and potentially dangerous situations. Thus, the link between the ability to multitask and the need for selection could be further strengthened. Attentional and memory resources are key components in multitasking (Chérif et al., 2018). For instance, during dual tasking, the switching of attentional resources is crucial and a limited ability to select critical and disregard unimportant information (e.g., attentional selection; Oberauer, 2019) could lead to capacity limits being exceeded (Jaeggi et al., 2007). Another type of multitasking involves task-shifting and the alternation of attention between tasks. Task-shifting involves frequent changes in the operator's cognitive framework (Chérif et al., 2018). For drone pilots, establishing a shift in the cognitive framework would involve the ability to continuously place the drone in its three-dimensional space, and a lack of adaptive spatial ability would substantially influence the drone operation in a negative way. A third type of multitasking is task interruption, which is often due to changes in critical environmental demands. This process implies the retrieval of the original status of the mission after the interruption has been handled, which taxes memory resources (Altmann & Trafton, 2002). For drone pilots, this will often be represented by visual memory ability. Furthermore, drone operations consist of sustained surveillance and targeting using highly automated systems. The monitoring of these systems draws heavily on the vigilance capacity (i.e., sustained attention) of a system operator (Zanetti et al., 2022). The link between visual cognitive

functions and piloting drones is obvious since information is mainly relayed from the drone to the operator by the visual modality. Visual attention predictions have also been found to improve drone piloting (Pfeiffer et al., 2022).

Taken together, there is a need for selection programs for drone pilots in the civilian sector (including law enforcement), and, as a consequence, there is a lack of knowledge regarding the usefulness of tests in this type of selection. This study aims to help fill this gap by testing the predictive value of visual cognitive tests taxing sustained attention, attentional selection, visual short-term memory, spatial orientation, and logical reasoning. The test battery was theoretically generated from the widely used Cattell-Horne-Carroll (CHC; Carroll, 1993) model of general cognitive abilities. The hierarchical CHC model places attentional control and visual-spatial short-term memory as facets in the same sub-dimension.

According to a recent meta-analysis (Bryan & Mayer, 2021), the C-H-C model was described as a compromise between the traditional controversy of intelligence viewed as a general factor versus a number of specific factors (multiple intelligence). Bryan and Mayer (2021) stated that the C-H-C model was criticized for encompassing too many specific factors, which led to suggestions of subsidiary groups among the broad factors to organize them. Thus, the development of such subsidiary groups took the form of crystallized and fluid intelligence and later acquired and latency for problem-solving (power and speed intelligence). The latest suggestion of grouping the abilities was people-centered versus thing-centered intelligence (Bryan & Mayer, 2021). People-centered intelligence involves personal, emotional, and social intelligence. Examples given on thing-centered intelligence are related to the understanding of specific abilities pertaining to visual patterns and movement of objects in space. Results from the meta-analyses supported the distinction between the two groups. If one should take the Bryan and Mayer (2021) approach, the cognitive tests evaluated in the present study could all be argued for belonging to the thing-centered group.

Based on the CHC model's statement of the cognitive functions as separate abilities under a common dimension, it was hypothesized that the cognitive tests would be moderately intercorrelated. It was further predicted that all cognitive tests would show a unique contribution in explaining the variance in a simulated drone flight. This was founded on the tests being grounded in a job analysis, and previous research has linked visual attention processes to drone piloting performance (Pfeiffer & Scaramuzza, 2021; Pfeiffer et al., 2022). Furthermore, Oberauer (2019) has suggested that short-term memory is a form of attentional system, further linking visual short-term memory to attentional processes.

Method

Subjects

A nationwide sample of 129 police officers (range 25–45 years) participated in the study. The sample consisted of 114 males and 15 females, and all subjects took part in the selection program in their capacity as applicants for a 10 ECTS-credit training course at the Norwegian Police University College to become certified police drone pilots. For 127 of the subjects, previous experience of flying drones was collected. Fifty-two subjects had never attempted to fly a drone prior to participating in the selection program, 53 subjects had tried flying a drone a few times, and 8 subjects had had substantial experience of flying drones or drone simulators. Applicants were drawn from nine police districts which included both rural and urban police forces.

Procedure and equipment

The testing was performed at the location of the respective subject's police district. Upon arrival, a consent form was presented and signed, and the participants were instructed that participation was voluntary and would not affect their application for the training course. They were further

informed that the data would be anonymized and that no individual data would be relayed to their superiors. They were further instructed that they could withdraw at any time during testing and during a 6-month period after testing. The subjects were tested in groups of a maximum of four subjects in a separate room at the local police station. All tests were presented digitally on a portable computer (Lenovo) with a 15.6-inch screen. The test sequence was fixed for all subjects and was anticipated to take 108 min, including two mandatory 5-min breaks. The test battery was presented after the performance test.

The performance test was conducted using a Da-Jiang Innovations (DJI) flight simulator, run on a gaming laptop (Razer Blade 17). The ground control station used was a DJI smart controller. The gaming laptop was connected to a large television set or a projector supplied by the local police district, producing a screen size ranging from 55 to 70 inches.

The test battery

The test battery was based on selected tests from the Vienna Test System (Schuhfried, 2022). A test of spatial orientation was presented as the first task. As a measure of spatial orientation, the *Adaptive Spatial Ability Test* -standard form (A3DW) was used. The test measured the ability to form mental representations and the manipulation of these representations in a three-dimensional space. The respondent's task was to correctly identify a visual target by selecting from a panel of six comparisons. The visual stimuli were represented by cubes with different patterns, and the subject had to mentally rotate the cubes in order to identify the target among the comparisons. This gives eight options when answering which include the six comparisons, an option of 'I do not know the answer' and 'No die matches'. The inclusion of the response options of 'I do not know the answer' and 'No die matches' was designed with an intention of reducing the likelihood of guessing the right answer to a minimum. It was not possible to omit an item or to go back to a preceding one.

The next task presented was the sub-test of *Logical Reasoning* (INT). This subtest was chosen from the inventory for testing cognitive capabilities. The test measured the ability to detect deviations from rules based on visual input. According to the manual, the ability entailed the aptitude of solving innovative problems through inductive thought processes. A person revealing a high manifestation of this variable possesses a great degree of ability to recognize regularities or formulas and use the rules derived by doing so. During the test, respondents were presented with nine fields with abstract figures and asked to select a field that did not follow the same rules as the remaining eight. Thus, the subjects were to identify the links between the figures and then select from a choice of different responses to show an aptitude for identifying and applying the rules underlying the matrix.

Before starting the test, the type of the task and its processing were explained on the screen, and then illustrated using an animation. This was further followed by a practice trial that had to be completed before the actual test started to ensure understanding of the task.

The *Attention and Concentration Test* (TACO) was used in order to examine the ability of attentional selection. The test measured the ability to ignore irrelevant and to mark relevant visually presented targets. The TACO consists of 50 test pages each consisting of a matrix of nine stimuli per test page with the target stimuli located above the stimulus matrix on every test page. Target stimuli always consisted of one unfilled basic shape (triangle, circle, or square) and one or multiple filled details (triangle, circle, or square). Thus, the test contained items which gives an opportunity to discriminate between relatively simple stimuli that can be clearly perceived and are not contaminated by cultural background. By using basic shape, filled details and position of the figure in the matrix, the stimuli differ in three dimensions. While two of the dimensions, basic shape and detailed shape, are relevant for the solution, the third dimension of position was irrelevant. The matrix could consist of multiple targets, and the task of the responder was to find the target stimuli in the matrix and to mark them. The respondents were given a maximum of 6 s to complete each test page.

In order to test sustained attention, the 28-min version of the *Vigilance Test* (VIGIL) was chosen. The test complied with the characteristics of vigilance tests by requiring uninterrupted attention of the subject throughout a relatively long monotonous visual monitoring task. Relevant signals appeared randomly, without prior warning, and did not arouse ‘involuntary attention’ (i.e., low-intensity signals). Furthermore, the frequency of critical signals was low encompassing 100 events during the 28 min of testing. The actual test displayed a large circular path (similar to a clock face), made up of small circles that lit up in succession (in steps of 1.5 s) which gave the impression of movement in a clockwise direction, and the participants’ task was to respond by pressing the space bar on the keyboard when the illuminated dot made a double jump. An audible signal confirmed to the respondent that the press of the spacebar was recorded.

The *Visual Memory Test* (VISGED) was the final task presented, and this assessed visual short-term memory performance. VISGED consisted of an on-screen city map on which typical locations (e.g., hospitals) are marked by symbols. The maximum number of critical symbols was eight, and the task for the responders was to memorize the location of the individual symbols and recall them by placing them correctly on another city map presented without symbols. After completion of a trial, the correct position of the symbol in question was displayed as performance feedback.

The cognitive tests of A3DW, INT, and VISGED utilized an adaptive testing method (i.e., presenting of stimulus-difficulty based on the accuracy of the subject’s previous responses). Thus, the variable used in the analyses was based on a Rasch analysis and recorded as a personal parameter (PAR). The PAR represented the capability or attainment level of tested participants, expressed in terms of a continuum of probability for correct responses (McCamey, 2014). High and positive scores, within the bounds of the item characteristic curve, represented a high ability by indicating an elevated probability for correct responses. The PAR is an indication of a latent variable and is suitable for parametric statistics (McCamey, 2014; see also Debelack, 2019, for a discussion on parametric vs. non-parametric testing of the Rasch model). Sustained attention (VIGIL) was measured as reaction time in seconds to correctly answered test-items, and TACO was recorded as the sum of correctly worked items. Detailed descriptions of the development and psychometrics of the different tests used are presented in the test-manuals (Schuhfried, 2022). This includes the norming of the selection tools as well as reasonableness, resistance to falsification and fairness of the tests.

The performance test

The performance test started with a 5-min guided free-flight training session for the participants to become accustomed to the simulator and to flying a drone. The training session was excluded from the evaluation. The next step consisted of three trials (max. 6 min each) with increased difficulty, called ‘Four directions hovering’. The performance test involved the pilot standing fixed in an open environment surrounded by four visibly marked locations on the ground: one in front, one behind, and one on each side relative to the pilot’s position. Each location was marked with a circle with an arrow inside, giving all four locations different orientations relative to the pilot’s view, within a 90-degree sector. The arrow inside the circle indicated the orientation of the drone. All trials started with the drone parked on the ground in front of the subject, and it was operated in ‘altitude mode’ (no automatic position hold, only altitude hold). When the drone is in attitude mode, the pilot needs to constantly apply pitch and roll corrections with the controller to prevent the drone from drifting away from the hovering location. The task was to attain stable hovering 1 m above the marked circle on the ground, without deviation, for 10 s. If the drone drifted away from the hovering position or was not kept in stable hover above the position, a meter indicating the progress of the task as a percentage would slowdown and eventually stop, until the drone was repositioned and stabilized above the marked location. If the subject was not able to reposition the drone, the meter would count backward. All four hovering locations had to be completed with the drone in the correct orientation to complete the trial. The pilot’s view automatically tracked towards the drone at

all times, resulting in the drone orientation above each hovering location to be 90 or 180 degrees, offset to any of the other three hovering locations seen from the pilot's perspective. All trials were at an entry level.

Two performance indicators were distilled from the performance test. *Skills* was an indicator of the quality of piloting and was measured as the mean hovering time taken to complete the three trials comprising the performance test. The recording of time, in seconds, started when the drone took off from the ground and ended after the successful completion of all four hovering tasks of the respective trial. The performance indicator of *Proficiency* was based on an expert's evaluation of the following four elements: handling of the remote control; stress while piloting; drone orientation; and progression over the three trials. Each of the elements was scored from 1 to 7, and the proficiency score represented the mean of these elements. Thus, the two performance scores focused on both the quality of flight performance and on the critical behavior underpinning the safe handling of drones.

Statistics

Pearson product moment correlations were used to test for intercorrelations, and stepwise regressions were used in order to test the predictive value of the cognitive tests on performance indicators. The variable with the smallest p-value was entered first, followed by the variable with the smallest p-value from the pool of remaining variables. Variables already entered are removed if their p-value exceeds the chosen limit due to the inclusion of new variables. The entire process stops when no more variables can be entered or removed. The chosen probabilities to enter and to remove were .05 for inclusion and .10 for removal.

Only tests correlating significantly with the performance scores were included in the regression analyses. Varying degrees of freedom in the correlation analyses were caused by missing data.

Results

Correlations

Descriptive statistics and intercorrelations are shown in Table 1.

The correlational analysis showed that spatial orientation correlated with logical reasoning and attentional selection. Logical reasoning showed significant correlations with both attentional selection and visual memory, and attentional selection correlated significantly with visual memory. All correlation coefficients were low to medium ($r(126) = .167$ to $r(126) = .324$; Cohen (1988). No correlations involving sustained attention reached a significance level.

When looking at the association between cognitive tests and the performance indicator of skills, spatial orientation revealed a negative significant correlation. Negative associations were also found for logical reasoning and attentional selection when correlated with skills. The results showed that

Table 1. Means and standard deviations (SD) for all five cognitive tests and for the performance indicators of skills (mean hover time) and proficiency (mean score of remote-control handling, stress, drone orientation, and progress) during a simulated drone flight. Sustained attention data are presented in seconds and attentional selection as the number of correct answers. All other tests are presented as personal parameters derived from Rasch analyses.

	Mean	SD	1	2	3	4	5	6
<i>Spatial orientation (1)</i>	-.44	2.24						
<i>Logical reasoning (2)</i>	-.58	1.06	.315**					
<i>Attentional selection (3)</i>	416.15	15.26	.177*	.323**				
<i>Sustained attention (4)</i>	.42	.05	-.129	-.087	-.033			
<i>Visual memory (5)</i>	2.3	1.85	.123	.167*	.324**	-.051		
<i>Skills (6)</i>	202.6	84.53	-.313**	-.261**	-.225**	.156	-.063	
<i>Proficiency (7)</i>	4.68	.81	.321**	.242*	.247**	-.159	.071	-.831**

*= $p < .05$; **= $p < .01$.

Table 2. Unstandardized (b) and standardized (β) coefficients, as well as *t*, and significance (sig.) values for the performance indicators of skills (mean hovering time) and proficiency (mean score of remote-control handling, stress, drone orientation, and progress) during a simulated drone flight. The data are separated for significant models.

	b	β	<i>t</i>	sig.
<i>Skills Model 1</i>				
Spatial orientation	-11.57	-.313	-3.67	<.001
<i>Skills Model 2</i>				
Spatial orientation	-10.51	-.284	-3.33	.001
Attentional selection	-.97	-.18	-2.1	.038
<i>Proficiency Model 1</i>				
Spatial orientation	.114	.03	3.774	.001
<i>Proficiency Model 2</i>				
Spatial orientation	.102	.03	3.411	.001
Attentional selection	.01	.004	2.372	.019

increased scores on the cognitive test are related to increased performance measured as shorter completion time of the piloting task (see Table 1).

The relationship between cognitive tests and proficiency showed that spatial orientation was significantly and positively correlated with proficiency. Furthermore, both logical reasoning and attentional selection showed significant positive correlations, with the proficiency variable exhibiting that an increased score on the cognitive tests covaried with higher scores on the SME's evaluation of proficiency. No analyses involving sustained attention revealed any significant associations (see Table 1).

Regression analyses

Based on univariate correlation analyses, spatial orientation, logical reasoning, and attentional selection were included as predictors in the stepwise regression analysis. The results with skills as dependent variable revealed a significant model for the first step including spatial orientation ($R = .313$, $F(1,124) = 13.448$, $p < .001$) and the second step adding attentional selection to the model ($R = .359$, $F(2,123) = 9.116$, $p < .001$). Both models revealed a negative association. The first step explained 9.8% of the variance ($R^2 = .098$), and the second step ($R^2 = .129$) added 3.1% of explained variance (see Table 2 for details).

When regressing proficiency on the predictor variables, a significant model occurred for spatial orientation as predictor ($R = .321$, $F(1, 124) = .14.23$, $p < .001$). The second step also revealed a significant model when adding attentional selection as an independent variable ($R = .377$, $F(2,123) = 10.201$, $p < .001$). The first model explained 10.3% ($R^2 = .103$) of the variation in proficiency, and attentional selection added 3.9% ($R^2 = .142$; see Table 2). No models including logical reasoning achieved significance.¹

Discussion

The present study showed significant intercorrelations between all cognitive tests except for sustained attention. Furthermore, only the tests for spatial orientation, logical reasoning, and attentional selection were correlated with the two performance measures used. The analyses of predictive validity indicated that both spatial orientation and attentional selection showed a unique contribution to both performance tests.

The tests chosen in the selection program were based on a job analysis of a police drone pilot's operational modus. The job analyses focused on the control of a technological system geographically separated from the pilot, and the safe direction of the system towards a mission aim. Job analyses also considered the ability to identify inhibiting and facilitating factors in order to successfully attain the mission goal. This included police tactics, topography, maps, addresses,

and other police units on the ground. At the same time, the pilots' decisions and performance have to be in line with national rules and regulations applied to aerial operations, as well as ethical and legal rules governing police conduct and operations. The pilot must orient the drone in a three-dimensional space relative to their own position and communicate with other units while considering that these units have a different visual perspective. Thus, the demands of police drone pilot tax heavily on the cognitive abilities of logical reasoning, spatial orientation, and short-term memory. Police operations involving drones are often conducted over an extended period of time and involve multitasking behavior for the drone pilot. Multitasking relies heavily on memory and the attentional processes of short-term memory, sustained attention, and attentional selection. Thus, the cognitive tests chosen in the selection program were considered relevant in order to ensure flight safety and police mission efficiency when conducting police drone operations. Cognitive tasks of multitasking are also typically used in aerial-related research (Fraser, 2020).

Four of the five cognitive tests showed significant intercorrelations. The correlations were low to medium, sharing less than 11% of the covariation. This leaves over 89% of the covariation unexplained and indicates that the tests are, to a large extent, tapping into different cognitive abilities. Since the tests are chosen based on the job analyses, this supports the argument for using the tests in the selection of drone pilots. This also ties into the discussion of domain-general or domain-specificity in attentional control (Hedge et al., 2018). The lack of association between the two measures of attention (sustained attention and attentional selection) lends support to the notion of domain-specificity in attentional control (Rey-Mermet et al., 2018). However, Draheim et al. (2020) opposed this view by claiming that a lack of shared variation in attentional tasks was caused by the use of experimental factors that were not suitable for studying individual differences. The use of reaction time measures often caused conflicting results, and they recommended the use of accuracy measures. Our study also lends support to this claim since correlational analyses involving accuracy-based scores showed significant intercorrelations, while analyses of scores using reaction time data did not.

Spatial orientation, logical reasoning, and attentional selection were correlated to both the quality of drone piloting and the underlying behavior recorded by the proficiency variable. Looking at skills, the explained covariation for significant negative associations varied, from around 5% (logical reasoning and attentional selection) to 10% (spatial orientation). Almost identical (but positive) correlations were found for proficiency. No correlations were found for sustained attention and visual short-term memory. Although the study failed to reveal associations between visual memory test, sustained attention tasks, and performance, this does not mean that these cognitive functions are irrelevant, since they are clearly linked to job analyses. One possible explanation for the lack of correlations could be the low levels of variation in responding between the subjects. This was especially the case for sustained attention which showed a very small standard deviation. Furthermore, the maximum duration of the performance test trials was only 6 min, which could be insufficient to adequately tax sustained attention. Also, the performance tests were on an entry level implicating piloting in 'visual line-of-sight'. The challenges of visual short-term memory systems are greatly reduced in this type of flight, compared to a 'beyond visual line-of-sight' operation. The latter drone operation also shows more resemblance to the memory test used, as this will take into account the transfer of information to and from other sources (e.g., maps). Thus, one explanation for the lack of correlations could be the relatively low load on vigilance and visual short-term memory created by the performance tests. Despite the lack of associations found in our study, the tests could be used in negative selection (Waters, 1998) as a tool for excluding subjects showing low levels of sustained attention and short-term memory.

The regression analyses revealed that spatial orientation and attentional selection made unique contributions to both the quality of piloting a drone and the underlying indicator of proficiency. Thus, these tests showed predictive validity in drone selection since performance scores could be estimated based on the test scores (Schober et al., 2018). The importance of visual attention was also demonstrated by Pfeiffer and Scaramuzza (2021) for expert drone pilots involved in a simulator-

based racing task. Further support was reported by Pfeiffer et al. (2022) who reported that ‘visual attentional prediction’ (gaze placement) predicted performance in a similar task. Previous studies using different test batteries have produced mixed results when predicting performance in airline pilot selection. For instance, predictive effects have been found using the WOMBAT test battery (Caponecchia et al., 2018). The WOMBAT test battery incorporates tasks that assess target tracking, spatial orientation, pattern recognition, and short-term memory skills. It could be argued that target tracking and pattern recognition could involve attentional selection. However, the use of the COMPASS test battery did not reveal any predictions of performance (Cabeza et al., 2021). In their study, a composite score of coordination between eyes and extremities, short-term memory, arithmetic ability, spatial orientation, and multitasking was used.

Practical implications

In addition to the dilemma of small correlations, Caponecchia et al. (2018) emphasize the problem that most of the studies predict training performance rather than operational flying. When it comes to the selection of drone operators, we would argue for a third approach when testing the association between tests and performance. This approach involves the use of performance tests measured in advance of training when investigating the validity of selection procedures. It could be speculated that the use of untrained subjects could, to a greater degree, reveal the potential for later performance, compared to personnel evaluated after training. Training programs often involve both self-training and a dedicated focus from instructors on trainees’ specific needs identified during training. The possibility exists that using trained subjects in studies of predictive values of tests could result in erroneous conclusions since the variation in performance is masked by the use of extended training hours and costly instructor resources. Thus, the use of results from untrained subjects in the evaluation of test batteries could enhance the benefit, relative to the cost, during selection. Furthermore, one specific confounding variable when selecting drone pilots is previous drone-related experience. The effects of previous experience with recreational drones and gaming could enhance actual piloting during a performance test. This could lead to an adverse effect, by not selecting applicants who have great potential as drone pilots but who have less experience of using drones. By using behavioral characteristics that are considered to underpin the safe handling of drones (e.g., proficiency measures) in addition to the quality of piloting (e.g., skills), the potential for performance could, to a greater degree, be investigated, regardless of experience of drone piloting. The correlation between the dependent variables of proficiency and skills used in our study was high (Schober et al., 2018), showing a close association between the variables. Detecting the potential for later performance is the goal of any selection. The possible benefit of using untrained subjects when evaluating selection procedures could give a promising direction for future research.

There are two further significant implications drawn from the present study. First, there is a need for transparency in the civilian aviation selection community with regard to documentation of pilot selection systems. Broach et al. (2019) stated that pilot selection programs appear to be largely hidden behind corporate firewalls. According to Broach and coworkers (2019), the airline companies view their selection programs as a vital source in the competition for talented aircrew. This has led to an underdeveloped knowledge base and a possibility for an emergence of selection programs lacking in scientific basis (Broach et al., 2019). Since selection programs of civilian drone pilots is in its infancy, a high potential for transparency across different civilian domains and between countries exists. This implies the generation of a system to both describe the selection program and scientific data related to the program. This is also in line with Broach and colleagues (2019) who suggested building a clearinghouse for civilian pilot selection data in order to feedback analyses to the industry.

The second implication concerns the evaluation of the validity of drone selection programs. We claim that since aviation selection programs consist of multiple tests, the validity argument for such programs should be based on the same scientific approach as for the individual tests used. When

developing such an argument, several lines of evidence have to be integrated. As stated by Reynolds and co-workers (Reynolds et al., 2021, p. 187): ‘Validity must always have a context and that context is interpretation’. When using tests in order to select drone pilots, the job analyses represent the context. By integrating both a theoretical and a practice-related proposition, a valid selection procedure could be developed which bridge the gap between a best practice and an academic perspective. Thus, a broader than the traditional approach should be taken. Analyses of the validity of tests used in selection have traditionally focused on the predictive validity of each test in order to determine the feasibility of tests involved in the program. However, the US ‘Standards for Educational and Psychological Testing’ (American Educational Research Association, 2014) define validity as ‘the degree to which evidence and theory support the interpretations of test scores for proposed uses of the test’ (p. 11). A consequence of this definition of test validity is an emphasis on the interpretation of the test results, rather than a validation of the test itself, since validity refers to the appropriateness and accuracy of the interpretation (Reynolds et al., 2021, p. 186). Thus, when evaluating the selection procedure, focus should also be put on how the scores obtained by the test taker reflect the relevant aptitude of the tested subject. According to the standards (American Educational Research Association, 2014) this interpretation should to a greater degree be validated and the test scores itself. This could influence both the practical use of selection tools and future research. In order to make inferences of how the scores reflect the latent variable, tests should be founded on sound theoretical and psychometric grounds. When using the tests as part of the selection procedure, the tests should also be closely associated with the job analyses. A sound theoretical foundation together with a close association with the demands distilled from the job-description would more clearly explain how the scores are related to the different latent variables in question. In the present study, we have outlined both the theoretical and practical foundation of all the tests used. This was the case for specific tests as well as a discussion of the relationship between multiple tests. Examples of the latter were the interpretation of attentional control and multitasking, and the theoretical foundation of several tests based on the CHC-model.

The standards (American Educational Research Association, 2014) emphasize validity as a unitary construct and follow that the division into classical types of validities is of less importance. Instead of dividing the validity construct into content, criterion and construct validity, future evaluation of tests used in selection should focus on the major threats to validity represented by construct underrepresentation and construct-irrelevant variance (Reynolds et al., 2021, p. 192). Construct underrepresentation represents a lack of validity due to tests not representing relevant knowledge, and construct-irrelevant variance is the presence of uncontrolled, extraneous variables that influence the results. This threat is also present in selection programs where a possible consequence is an increased potential for the program misrepresenting the actual characteristics of the test-taker due to the assessment of too little relevant or too much irrelevant information. Furthermore, a validity argument should identify strengths and weaknesses of the test scores. There are five different sources to such an interpretation (American Educational Research Association, 2014) that could also apply to selection programs. Test content encompasses evidence of the representativeness of the test items (or individual tests) with regard to the construct in question. Evidence based on response processes includes an evaluation of the human processes involved when taking the test. This includes not only the subjects taking the test but also the evaluation of the procedure for evaluation used by the personnel responsible for administering and interpreting the test results. This could be especially challenging when selection programs consist of multiple tests. Evidence based on internal structure evaluates the relationship between test items or tests constituting a test battery and evidence based on relations to other variables includes an analysis on how the result of a test relates to performance or a criterion (e.g., prediction on a performance test). Thus, selection programs and individual tests should be evaluated with regard to revealing the potential for later performance of applicants. This could include strengths and weaknesses of the sample used (e.g., trained vs. untrained applicants) or the definition of the variable representing the performance (e.g., skills vs. proficiency). Finally, evidence based on the consequences of testing focuses on the intended and unintended results of using a test. This could include unintended social consequences. It should be

noted that the standards (American Educational Research Association, 2014) argue for an evaluation of consequences directly related to the validity of the tests and to distinguish these analyses from evidence related to social policy. Thus, adverse analyses related to the validity of a selection program should undergo scrutiny

Limitations

Previous experience with drones, games, and flight simulators could influence the performance of the subjects. According to McKinley et al. (2011), video gamers exhibited quicker reaction times, showed higher levels of stimulus response mapping, and tracked more targets, and profit from greater spatial and psychomotor skills. The present study did not statistically control for these variables. However, levels of expertise were reported in the description of the sample, showing a low frequency of subjects being highly skilled in piloting. Furthermore, the proficiency variable focused on behavior crucial to piloting, with the aim of minimizing the effect of previous experience.

In order to maintain anonymity of the data, age data were only available for the instructors and not recorded as part of the present study. This could be a limitation since age-related reduction is found in cognitive functions relying on visual attention and memory (C. Thomas et al., 2008; Lindenberger et al., 2001; Rieck et al., 2017). Rieck et al. (2017) defined the old age as 55–69 years. However, none of our subjects were in that age category.

Conclusion

The present study investigated the usefulness and predictive validity of selected cognitive tests from the Vienna Test System (Schuhfried, 2022) on an untrained sample of applicants for a course aimed at educating and qualifying police drone pilots. This made it possible to extend previous knowledge on the evaluation of cognitive tests used in pilot selection since most studies involve performance tests conducted after a training program. Training programs aim to develop the skills of drone pilots, thereby reducing the variance in performance. By using untrained subjects, the relationships between test scores and performance scores should appear without being affected by training. The low percentage of covariation between the tests is supported in the literature of attentional control and confirms previous research. On the other hand, the present study extends previous knowledge by showing that only spatial orientation and attentional selection predicted skills and proficiency during a performance test using untrained subjects.

The present study evaluated the neuropsychological tests used in the selection part of recruitment of police drone operators. It is important to note that the selection is part of recruitment to an educational program at a college level. It is our view that selection, high-level education, and training have to be combined in order to develop a well-functioning police drone operator. Choosing applicants with sufficient cognitive abilities, educating them at a high level and increasing their skills as pilots could decrease the possibilities for adverse effects for the pilots and the public.

Although, the present study investigated selected tests involved in a selection program for drone pilots, we argue that future investigations of validity for such programs should be based on the same scientific approach as for the individual tests used. Furthermore, future research should increase its focus on how the scores obtained reflect the relevant characteristics of the tested subject whilst considering the demands drawn from the job analyses.

Note

1. We also performed standard multiple regressions with the two variables identified from the stepwise regression included as independent variables (spatial orientation and attentional selection). The results from these regressions were identical to the results from the stepwise regression in terms of the magnitude of regression coefficients.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

The work was supported by the The Norwegian Police University College, Oslo, Norway.

Notes on contributors

Bjørn Helge Johnsen (Ph.D) is a professor of personality psychology at the University of Bergen, Norway. He is heading the Centre for Crisis Psychology and has substantial experience in selection to Special Police Operative Educations in the National Police as well as research in operational psychology. His research has been published in various international journals, including *Journal of Police and Criminal Psychology*, *Military Psychology*, *Military Behavioral Health*, *Military Medicine*, *Nutrients*, *Personality and Individual Differences*, and *Frontiers in Psychology*

Andreas Aalstad Nilsen (B.S) is a police superintendent at the Norwegian National Air Support Unit. He is at present safety and security manager at Section for Unmanned Aerial Systems. Nilsen is a former air force officer, F-16 Crew Chief, and local SWAT team officer.

Sigurd W. Hystad, (Ph.D) is a professor in the Department of Psychosocial Science at the University of Bergen. His interests include research methods and statistics, individual resiliency, stress, fatigue, and other personal readiness issues. Professor Hystad's research has appeared in a variety of scholarly journals, including *Journal of Police and Criminal Psychology*, *Military Psychology*, *Military Behavioral Health*, *Personality and Individual Differences*, and *Journal of Occupational Health Psychology*

Jørgen Lunde Rogne (M.M) holds the rank of police superintendent at the Norwegian National Air Support Unit. He is at present Operational manager at Section for unmanned aerial systems

Eyvind Grytting (B.S) is a police superintendent at the Norwegian National Air Support Unit. He is at present Safety and security manager at Section for Unmanned Aerial Systems. He is a former military and police digital forensic investigator

Steinar Rostad is a police superintendent and managing the Special Police Operative Educations at the Norwegian Police University College. His previous experience covers operational experience from the local SWAT team and as instructor for several courses at the National Police level.

Peter Henrik Öhman is a police superintendent and managing the police Operative Educations at the Norwegian Police University College. His previous experience covers operational experience from the local SWAT team and as instructor for several courses at the National Police level. Öhman as previously published in *Journal of Police and Criminal Behavior*.

Arne Jon Overland is a police superintendent and managing the Special Police Operative Educations at the Norwegian Police University College. He has experience in training and operational leadership of SWAT teams as well as head of operations at the Police University College.

References

- Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation based model. *Cognitive Science a Multidisciplinary Journal*, 26(1), 39–83. [https://doi.org/10.1016/S0364-0213\(01\)00058-1](https://doi.org/10.1016/S0364-0213(01)00058-1)
- American Educational Research Association, American psychological association, & national council on measurement in education (Eds.). (2014). *Standards for educational and psychological testing*.
- Armour, C., & Ross, J. (2017). The health and well-being of military drone operators and intelligence analysts: A systematic review. *Military Psychology*, 29(2), 83–98. <https://doi.org/10.1037/mil0000149>
- Aydin, B. (2019). Public acceptance of drones: Knowledge, attitudes, and practice. *Technology in Society*, 59, 101180. <https://doi.org/10.1016/j.techsoc.2019.101180>
- Broach, D., Schroeder, D., & Gildea, K. (2019). Best practice in pilot selection. Civil Aerospace Medical Institute's publications website: <http://www.faa.gov/go/oamtechreports>
- Bryan, V. M., & Mayer, J. D. (2021). Are people-centered intelligences psychometrically distinct from thing-centered intelligences? A meta-analysis. *Journal of Intelligence*, 9(4), 48. <https://doi.org/10.3390/jintelligence9040048>

- Cabeza, I. G., Molesworth, B., Good, M., Caponecchia, C., & Steffensen, R. (2021). Investigating the predictive validity of the COMPASS pilot selection test. *The International Journal of Aerospace Psychology*, 31(3), 252–268. <https://doi.org/10.1080/24721840.2021.1885297>
- Caponecchia, C., Zhengb, W. Y., & Reganc, M. A. (2018). Selecting trainee pilots: Predictive validity of the WOMBAT situational awareness pilot selection test. *Applied Ergonomics*, 73, 100–107. <https://doi.org/10.1016/j.apergo.2018.06.004>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Chapelle, W., Swearingen, J., Goodman, T., & Thompson, W. (2014). Personality test scores that distinguish U.S. air force remotely piloted aircraft “Drone” pilot training candidates. *Tech. report AFRL-SA-WP-TR-2014-0001*
- Chérif, L., Wood, V., Marois, A., Labonté, K., & Vachon, F. (2018). Multitasking in the military: Cognitive consequences and potential solutions. *Applied Cognitive Psychology*, 32(4), 429–439. <https://doi.org/10.1002/acp.3415>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Debelack, R. (2019). An evaluation of overall goodness-of-fit testes for the Rasch model. *Frontiers in Psychology*, 9, 27. <http://dx.doi.org/10.3389/fpsyg.2018.02710>
- Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2020). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, 150(2), 242–275. <https://doi.org/10.1037/xge0000783>
- Fraser, W. D. (2020). Stress, cognition, drones, and adaptive tasks. *Aerospace Medicine and Human Performance*, 91(4), 376–378. <https://doi.org/10.3357/AMHP.5584.2020>
- Graham, A., Kutzli, H., Kulig, T. C., & Cullen, F. T. (2021). Invasion of the drones: A new frontier for victimization. *Deviant Behavior*, 42(3), 386–403. <https://doi.org/10.1080/01639625.2019.1678973>
- Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, 50(3), 1166–1186. <https://doi.org/10.3758/s13428-017-0935-1>
- Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective & Behavioral Neuroscience*, 7(2), 75–89. <https://doi.org/10.3758/CABN.7.2.75>
- Lei, G., Shuguang, Z., Peng, T., & Yi, L. (2014). An integrated graphic–taxonomic– associative approach to analyze human factors in aviation accidents. *Chinese Journal of Aeronautics*, 27(2), 226–240. <https://doi.org/10.1016/j.cja.2014.02.002>
- Lindenberger, U., Scherer, H., & Baltes, P. B. (2001). The strong connection between sensory and cognitive performance in old age: Not due to sensory acuity reductions operating during cognitive assessment. *Psychology and Aging*, 16(2), 196–205. <https://doi.org/10.1037//0882-7974.16.2.196>
- McCamey, R. (2014). A primer on the one-parameter rasch model. *American Journal of Economics and Business Administration*, 6(4), 159.163. <https://doi.org/10.3844/ajebasp.2014.159.163>
- McKinley, R. A., McIntire, L. K., & Funke, M. A. (2011). Operator selection for unmanned aerial systems: Comparing video game players and pilots. *Aviation, Space, and Environmental Medicine*, 82(6), 635–642. <https://doi.org/10.3357/ASEM.2958.2011>
- Oberauer, K. (2019). Working memory and attention – a conceptual analysis and review. *Journal of Cognition*, 2(36), 1–23. <https://doi.org/10.5334/joc.58>
- O’neill, E., Brereton, F., Shahumyan, H., & Clinch, J. P. (2016). The impact of perceived flood exposure on flood-risk perception. The role of distance. *Risk Analysis*, 36(11), 2158–2186. <https://doi.org/10.1111/risa.12597>
- Pfeiffer, C., & Scaramuzza, D. (2021). Human-piloted drone racing: Visual processing and control. *IEEE IEEE Robotics and Automation Letters*, 6(2), 3467–3474. <https://doi.org/10.1109/LRA.2021.3064282>
- Pfeiffer, C., Wengeler, S., Loquercio, A., & Scaramuzza, D. (2022). Visual attention prediction improves performance of autonomous drone racing agents. *PLoS One*, 17, e0264471. <https://doi.org/10.1371/journal.pone.0264471>
- Qia, S., Wang, F., & Jing, L. (2018). Unmanned aircraft system pilot/operator qualification requirements and training study. *MATEC Web of Conferences*, 179, 03006 <https://doi.org/10.1051/mateconf/201817903006>
- Rauch, M., & Ansari, S. (2022). Waging war from remote cubicles: how workers cope with technologies that disrupt the meaning and morality of their work. *Organization Science*, 33(1), 83–104. <https://doi.org/10.1287/orsc.2021.1555>
- Ravindranath, M. (2017). *Thousands apply to trump’s drone pilot program*. <https://www.proquest.com/magazines/thousands-apply-trumps-drone-pilotprogram/docview/1970746736/se-2?accountid=8579>
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501–526. <https://doi.org/10.1037/xlm0000450>
- Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). *Mastering modern psychological testing: Theory and methods* (2nd ed.). Springer Nature Switzerland AG. 10.1007/978-3-030-59455-8

- Rieck, J. R., Rodrigue, K. M., Boylan, M., & Kennedy, K. M. (2017). Age-related reduction of BOLD modulation to cognitive difficulty predicts poorer task accuracy and poorer fluid reasoning ability. *NeuroImage*, *147*, 262–271. <https://doi.org/10.1016/j.neuroimage.2016.12.022>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. <https://doi.org/10.1213/ANE.0000000000002864>
- Schuhfried. (2022). *Vienna Test System: Digital Testing Made Easy with SCHUHFRIED*. <https://www.shuhfried.com/vienna-test-system/>
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2004). Risk as analysis and risk as feelings: Some thoughts about affect, reason, risk, and rationality. *Risk Analysis*, *24*(2), 311–322. <https://doi.org/10.1111/j.0272-4332.2004.00433.x>
- Thomas, C., Moya, L., Avidan, G., Humphreys, K., Jung, K. J., Peterson, M. A., & Behrmann, M. (2008). Reduction in white matter connectivity, revealed by diffusion tensor imaging, may account for age-related changes in face perception. *Journal of Cognitive Neuroscience*, *20*(2), 268–284. <https://doi.org/10.1162/jocn.2008.20025>
- Thomas, M. L., & Russo, M. B. (2007). Neurocognitive monitors: Towards the prevention of cognitive performance decrements and catastrophic failures in the operational environment. *Aviation, Space, and Environmental Medicine*, *78*(5 Suppl), B144–152.
- Waters, B. (1998). Personnel selection and classification in the military. In C. C. I (Ed.), *Military psychology: An introduction* (pp. 33–49). Simon & Schuster Custom Publishing.
- Zanetti, R., Arza, A., Aminifar, A., & Atienza, D. (2022). Real-time EEG-based cognitive workload monitoring on wearable devices. *IEEE Transactions on Biomedical Engineering*, *69*(1), 265–277. <https://doi.org/10.1109/TBME.2021.3092206>