National Defence | Défense nationale

Canadian Forces College | Collège des Forces Canadiennes

Enhanced Selection Criteria for RCAF Pilots

Major Anonymous

JCSP 49

Master of Defence Studies

Disclaimer

PCEMI n° 49

Maîtrise en études de la défense

Avertissement

Canada

CANADIAN FORCES COLLEGE - COLLÈGE DES FORCES CANADIENNES

JCSP 49 - PCEMI n° 49
2022 - 2023

Master of Defence Studies – Maîtrise en études de la défense

Enhanced Selection Criteria for RCAF Pilots

Major Anonymous

ENHANCED SELECTION CRITERIA FOR RCAF PILOTS

TABLE OF CONTENTS

ENHANCED SELECTION CRITERIA FOR RCAF PILOTS

ABSTRACT

This paper explores the evolution of pilot aptitude testing over the course of the last century and reviews the relevant literature on the cognitive psychology that underpins the most successful selection systems worldwide. The paper finds a thread of consistent outcomes from many independent studies that describe the core attributes which are the greatest enablers of success in pilot training and throughout an aviator's career. The tests by which these attributes are assessed vary in predictive validity and cost, and this research compares many different selection methods to determine which combination of tests and screening appears to be most effective. The RCAF pilot selection criteria is evaluated and the performance of candidates during basic flying training is assessed over a ten year period. Based on the results of this assessment, the paper makes recommendations to incorporate additional selection mechanisms, particularly in terms of personality screening, to reduce the cost of training failures, the burden on the training system and the likelihood of adverse incidents. The paper confirms the recommendations are in keeping with industry best practice and are in concert with stated policy objectives of the Government of Canada and the Canadian Armed Forces.

Keywords: Aptitude, Personality, Pilot, Screening, Selection, Training.

ENHANCED SELECTION CRITERIA FOR RCAF PILOTS

INTRODUCTION

The Royal Canadian Air Force will undergo a modernization of several fleets of aircraft in the years to come. The Canadian government has announced the purchase of 88 F-35 fighters from Lockheed Martin to replace the CF188;[1] the replacement of our aging A310 strategic tanker platform with the new A330 Multi-Role Tanker Transport (MRTT) aircraft from Airbus;[2] along with intended projects to introduce new Remotely Piloted Aerial Surveillance (RPAS) drones, Search and Rescue (SAR) aircraft and more.[3] Intrinsic to the Future Fighter Capability Program (FFCP) is a requirement to ensure future pilots are well prepared for operations in fifth generation fighter aircraft which necessitates the adoption of two other portfolios: Fighter Lead-in Trainer (FLIT) replacement, and the Future Aircrew Training (FAcT) portfolio which each have an unenviable task of selecting appropriate training aircraft to replace the T-6A Harvard II (CT156) and Mk 115 Hawk (CT155) aircraft.[4] As technology in the modern cockpit environment continues to evolve, the same demand for adaptation must be asked of our pilot selection system.

Pilot selection criteria have been couched in various forms of aptitude testing over the course of the last century. It is widely recognized that the training required is long and arduous and is increasingly expensive.[5] Militaries and airlines around the world understand that in an environment of finite resources, failure rates in training must be minimized to the maximum extent possible. Furthermore, follow-on training and longevity as a career aviator requires painstaking selection criteria to avoid long-term investment in candidates ill-suited, who by unsafe actions or disinterest in the profession will not remain in the cockpit for a sufficiently long period to warrant the huge capital investment.[6]

These considerations have been well studied across nearly every major aviation community since World War I (WWI) and many of the outcomes of those studies will be explored in this research paper. Eleven decades of aptitude and personality research has

---

[1] National Defence, 'Announcement Regarding the F-35 Acquisition', Government of Canada, National Defence, 9 January 2023, https://www.canada.ca/en/department-national-defence/news/2023/01/announcement-regarding-the-f-35-acquisition.html.
[2] 'Airbus Deemed Only Qualified Supplier for New RCAF Refueling and VIP Aircraft', *Ottawa Citizen*, 1 April 2021, Online edition.
[3] 'Airbus Appoints CAE to Support RCAF's FWSAR Programme', *Progressive Digital Media Defense (Incl, Airforce, Army, Navy and Homeland Security) News*, 15 February 2017.
[4] Government of Canada, 'Future Aircrew Training', Government, DGPAAPP, 1 December 2022, http://dgpaapp.forces.gc.ca/en/defence-capabilities-blueprint/project-details.asp?id=1269; Government of Canada, 'Future Aircrew Training Program', Government, TPSGC-PWGSC, 9 February 2023, www.tpsgc-pwgsc.gc.ca/app-acq/amd-dp/air/snac-nfps/ffpn-fact-eng.html; Government of Canada, 'Future Fighter Lead-in Training', Government, DGPAAPP, 18 February 2022, https://www.tpsgc-pwgsc.gc.ca/app-acq/amd-dp/air/snac-nfps/eipfc-fflt-eng.html.
[5] James F. Johnson et al., 'Predictive Validity of Spatial Ability and Perceptual Speed Tests for Aviator Training', *International Journal of Aerospace Psychology* 27, no. 3–4 (2017): 109–20, https://doi.org/10.1080/24721840.2018.1442222.
[6] In keeping with the military value of stewardship espoused in the doctrine *Trusted to Serve*.

been conducted for the purpose of designing effective pilot selection criteria by organizations that include the United States Air Force (USAF), the United States Navy (USN), the National Aeronautics and Space Administration (NASA), the Royal Air Force (RAF) and Royal Navy (RN), the Royal Canadian Air Force (RCAF) and many more.[7] The research continues to evolve as technological advances create new environments in which humans must operate. Furthermore, techniques gained from other disciplines interested in human performance, combined with a new understanding of cognitive psychology have provided a more refined lens through which we can view aptitude testing in the aviation environment.

Over the course of the twentieth century, aviation technology developed to a degree that would be unrecognizable to the progenitors of human flight. In the span of half a century, humans moved from a paradigm in which heavier than air flight was thought to be impossible, to one in which turbo-jet propelled night-fighter aircraft such as the Me 262 were equipped with airborne radar.[8] In the subsequent decade, human propelled flight exceeded twice the speed of sound with the F-104[9] and the pace of technological development of third, fourth and fifth generation aircraft have each brought with them significant challenges for human operators to contend with and overcome.

The rapid technological progression of aviation generated various iterations of aptitude tests worldwide. Each new test addressed the shortcomings of its predecessor with regards to the step changes in cognitive demands upon pilots. Hand-eye-foot coordination and other motor function tests, including reaction time and working memory (WM) tests evolved in concert with the increased speed of aircraft, while testing on divided/channelized attention and perceptual speed addressed the increased complexity of the cockpit environment.[10] In the modern era, electronic capacity has allowed engineers to present pilots with far more information than ever before, generating entire sub-fields of study into display representation for human factors.[11] This trend is a feature of a revised

---

[7] Examples of studies from each of these departments and agencies will be included in the sections that follow.

[8] Hecht, Heinrich. 'The World's First Turbojet Fighter – Messerschmitt Me 262'. Schiffer, 1990. The Me 262 night fighter variant with onboard radar was operational prior to 1945, only four decades after the Wright brothers' first flight in 1903.

[9] Lockheed Martin, 'F-104 Starfighter', Lockheed Martin, 2023.

[10] Johnson et al., 'Predictive Validity of Spatial Ability and Perceptual Speed Tests for Aviator Training'; Thomas R. Carretta and Malcolm J. Ree, 'Pilot Candidate Selection Method (PCSM): What Makes It Work' (BROOKS AFB TX: ARMSTRONG LAB, 1993); R. F. Eastman and R. L. McMullen, 'The Current Predictive Validity of the Flight Aptitude Selection Test' (ALEXANDRIA VA: ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES, 1978),

[11] Michelle Yeh, 'Attention and Trust Biases in the Design of Augmented Reality Displays' (ProQuest Dissertations Publishing, 2000); Dan Maurino Eduardo Salas, ed., *Human Factors in Aviation*, Second (UK: Academic Press, 2010); Lawrence J. Hettinger et al., 'Effects of Virtually-Augmented Fighter Cockpit Displays on Pilot Performance, Workload, and Situation Awareness', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 40, no. 2 (1996): 30–33; Patrick J. Doherty, 'Electronic Checklists on Multi-Purpose Displays: A Better Way For Fighter Pilots to Manage Information and Situational Awareness during Periods of High Workload', 1 January 2001, http://www.dtic.mil/docs/citations/ADA387109.

aptitude testing system the RCAF adopted in 2013 to replace its legacy Canadian Automated Pilot Selection System (CAPSS) testing during pilot selection.[12]

The legacy pilot selection system for the RCAF (CAPSS) included a basic simulator (in addition to a battery of written tests) which measured learning curve, hand-eye-foot coordination and reaction time to generate a pass or fail outcome for candidates following multiple assessment sessions. In Forgues' 2014 thesis dissertation comparing the RCAF CAPSS selection with the Royal Air Force Aircrew Aptitude Test (RAFAAT), the author noted the potential advantages of the RAFAAT over CAPSS which included a greater emphasis on cognitive Executive Functions (EF) that showed a greater relevance in an information-dense environment. CAPSS, by contrast, appeared better suited to selecting candidates based on the coordination required to maneuver aircraft during the more basic phases of flying training. The study highlighted numerous relevant factors for consideration in an improved pilot selection system which will be explored in greater detail. It is worth noting that the RCAF adopted the RAFAAT to replace CAPSS in 2013 and has been using it exclusively as the preferred pilot selection tool ever since. The number of candidates selected using the RAFAAT who have subsequently undergone basic pilot training is now sufficient for an analysis of its utility in comparison with the legacy system.

Significance. The expensive and time-consuming nature of pilot training necessitates that chosen applicants have a high chance of success; both in terms of the aptitude and temperament required for successful ab initio training; and the requisite motivation to complete advanced training and become career aviators whose experience is necessary at the operational level. It is necessary to evaluate the methods by which we select candidates not only in terms of aviation-specific competencies informed by decades of research into aptitude testing; but also, psychological evaluation based on what we know about the rigour of cognitively demanding pursuits and the corresponding level of emotional and psychological resiliency required in the long term. These areas will be researched in detail. The physical and medical requirements, already robust and empirically proven, will not be explored in this paper.

The research project will build upon previous research conducted by the author into selection models and situational awareness (SA), particularly with respect to SA which pervades nearly every aspect of operational flying yet can only be measured by proxy during aptitude testing[13]. The performance of student pilots will be evaluated using datasets from the NATO Flying Training in Canada program which include student flying performance as measured by GPA, and data on the number of Progress Review Boards (PRB) convened as a function of the total number of student pilot completions for Phase II (Ph II) pilot training on the T-6A Harvard II aircraft. The core reason for comparing student performance on Ph II pilot training with the selection methodology is that Ph II serves as a selection course from which students are streamed into either multi-engine, fighter/instructor, or rotary wing aircraft and will thus have the largest and most diverse

[12] Scott McPhalen, 'Executive Function Assessment', 27 January 2023.
[13] Te Koeti, 'Assessing situational awareness'; Te Koeti, 'Pilot selection methods'; Te Koeti, 'Improved aircrew selection methods: The association between executive functions, perceptual speed, and situational awareness'; Te Koeti, 'Situational awareness and astronaut selection: Assessment of methods'.

sample population. The dataset includes roughly comparable proportions of student pilots who completed legacy selection (Canadian Automated Pilot Selection System or CAPSS) whose performance on Ph II can be compared with that of candidates assessed using the new method (the Royal Air Force Aircrew Aptitude Test Pilot Battery 11 or RAFAAT PB11)[14].

METHODOLOGY

This paper will compare the performance of pilot candidates undergoing basic flying training over more than ten years to determine if a reasonable change in performance is readily apparent. The paper will also explore areas of concern regarding selection and training performance based on quantitative data analysis and qualitative assessments from those in key positions within the training environment of the RCAF. Finally, the paper will explore options to address pilot selection deficiencies informed not only by aviation best-practice, but by leading human performance experts from other domains (including CANSOFCOM). An in-depth review of selection methods in use historically and currently with a diverse range of aviation organizations is necessary to provide foundational knowledge and context for addressing the changing demands of the modern cockpit. Additionally, a review of relevant literature on human factors in aviation systems and cognitive psychology will enhance the reader's understanding of the problem this paper will address, and the solutions proposed.

Research Question

Does the RCAF pilot selection process identify the best individuals for the modern aviation environment? Scientific robustness of each study will be evaluated during literature review and data from studies low in reliability or validity will not be included. Additional hypotheses pertaining to subtests of GPA and the corresponding limitations will be provided in the body of the document where required.

Delimitations

The study was limited to peer-reviewed research published in reputable journals that specialized in cognitive psychology, aptitude selection methods, aerospace psychology, military and commercial aviation, and other military operations. Due to the qualitative nature of the comparative research, independent verification of research methodologies and replication of results was not possible.

Limitations and Assumptions

Data reported in peer-reviewed journals are assumed to be accurate. Coherence within the field of cognitive psychology is assumed when comparing complimentary cognitive models such as *inhibiting*, *updating*, and *shifting* EFs mapping onto the

---

[14] Scott McPhalen, Canadian Forces Aircrew Selection Center Personnel Selection Officer, email message to author, 27 Jan, 2023.

Baddeley model of Central Executive System (CES);[15] and acceptance of these models by other fields (such as aviation) is assumed. Analysis is limited to the years 2012-2022 due to the available dataset. The effects of the Covid-19 pandemic on student pilot production, recruiting and retention, and disruptions due to aircraft serviceability will be discussed as potential factors for consideration but cannot be fully accounted for in the context of this research.

HISTORICAL CONTEXT

The cost and timeline associated with military pilot training has produced stringent selection criteria that must identify those who are unlikely to succeed without unduly eliminating capable candidates. An effective selection system ensures that finite training resources are reserved for those with the highest chance of success by identifying and removing ill-suited candidates. Unfortunately, many aptitude tests designed to evaluate key knowledge, skills, and aptitudes (KSAs) for the aviation environment are only capable of doing so by proxy and must duplicate the challenges and complexity of the profession at a lower cost and greater flexibility than the actual environment in which candidates will be trained and evaluated. Despite considerable research on specific attributes that correlate well with success in pilot training,[16] there are numerous approaches to selection that vary in terms of predictive validity and compromise. Furthermore, different approaches in civil and military aviation each provide unique perspectives and offer complementary practices that merit consideration. While variations in testing exist, a core feature common across the field of aviation is a requirement to uphold a standard of safety and effectiveness that must continually address new challenges in this evolving domain. Sound engineering practices and technological advances have reduced the accident rate in aviation over the last half-century, but this has underscored the criticality of the human component. Both equipment failure and human error have decreased over time, but mechanical failures have been curtailed at a faster rate leaving human error as the dominant contributor to accidents: In the US in 2006, nearly 80% of general aviation accidents were due to pilot error.[17] This further emphasizes the importance of selecting the best available candidates.

---

[15] John E. Fisk and Peter Warr, ‘Age and Working Memory: The Role of Perceptual Speed, the Central Executive, and the Phonological Loop’, *Psychology and Aging* 11, no. 2 (1996): 316–23, https://doi.org/10.1037/0882-7974.11.2.316; Goldstein, B., *Cognitive Psychology Connecting Mind, Research, and Everyday Experience*, 5th ed. (Cengage, 2019).
[16] Eastman and McMullen, ‘The Current Predictive Validity of the Flight Aptitude Selection Test’; Johnson et al., ‘Predictive Validity of Spatial Ability and Perceptual Speed Tests for Aviator Training’; Martinussen, M., ‘Pilot Selection: An Overview of Aptitude and Ability Assessment’, in *Pilot Mental Health Assessment and Support* (Routledge, 2016); Claire A. Portman-Tiller, Sean Biggerstaff, and Dave Blower, ‘Relationship Between the Aviation Selection Test and a Psychomotor Battery’ (PENSACOLA FL: US NAVAL AEROSPACE MEDICAL RESEARCH LAB, 1998),
[17] Husam Kharoufah et al., ‘A Review of Human Factors Causations in Commercial Air Transport Accidents and Incidents: From to 2000–2016’, *Progress in Aerospace Sciences*, 2018, https://doi.org///doi-org.ezproxy.libproxy.db.erau.edu/10.1016/j.paerosci.2018.03.002.

Early Aviation

In aviation's infancy, the core requirements for success were not well understood though efforts were made as early as WWI to determine the attributes necessary to succeed in the strife of battle. The frequency of accidents during the landing phase and the utility of aircraft for rudimentary bombing procedures informed an early study on the assumed KSAs and personality prerequisites for aspiring aviators during WWI.[18] These included candidate estimates of both relative speed and distance; vision requirements (including peripheral vision); balance assessments; reaction time based on visual stimulus and directional commands; physical strength and endurance; and an evaluation of spatial awareness that, while crude by today's standards, actually demonstrated a keen understanding of an aptitude that remains a core part of modern testing. In its earliest form, spatial awareness testing consisted of candidates evaluating depictions (viewed through binoculars) of the intersection of parabolic curves with a flat plane.[19] Though the sample populations were small, and subjects improved their parabolic curve estimates with practice, the study concluded that spatial ability and psychomotor skill measured by reaction time were positively correlated to pilot performance.[20]

As the RAF evolved in the period following WWI, its officer corps established distinct criteria for both long- and short-term pilots. The broader selection requirements for officers included educational background, demographics (though apparently arbitrarily by today's standards), and personality profiling. From within the selected officer cadre, additional testing was to be completed by pilot candidates to meet additional medical and capability requirements.[21] The nature of aerial warfare underwent a rapid evolution during WWII, enhancing our understanding of the physiological and psychological demands placed on aircrew.

By the early 1940s, comprehensive selection criteria existed in terms of both medical and psychological requirements. The field of aviation medicine had developed into a distinct discipline with a deep understanding of the physiological needs required for successful pilot training and subsequent military operations. To minimize wastage of resources, estimated costs of pilot training were factored into selection, with a significant cost of around $22,000 estimated by Mathewson for the Royal Canadian Air Force during WWII.[22] Aeromedical requirements included limits on height and weight to ensure candidates could reach the rudder pedals while maintaining visibility above the cockpit and fitting in compact gunner turrets. Vision requirements encompassed binocular, color, and night vision, while respiratory requirements accounted for changes in atmospheric pressure. Others included middle-ear disease restrictions, and motion sickness evaluation.

---

[18] G. M. Stratton et al., 'Psychological Tests for Selecting Aviators', ed. John B. Watson, *Journal of Experimental Psychology* 3, no. 6 (1920): 405–23, https://doi.org/10.1037/h0074528.
[19] Stratton et al.
[20] Stratton et al.
[21] Tony Mansell, 'Flying Start: Educational and Social Factors in the Recruitment of Pilots of the Royal Air Force in the Interwar Years', *History of Education (Tavistock)* 26, no. 1 (1997): 71–90, https://doi.org/10.1080/0046760970260105.
[22] Mathewson, F.A.L., 'Medical Aspects of Aircrew Selection', *The Canadian Medical Association Journal*, 1942, 318–22.

A range of excluding conditions, including epilepsy, pneumothorax, and congenital heart problems that could affect oxygen saturation, were also considered. Furthermore, a range of physiological stressors such as temperature, vibration, fatigue, and decompression sickness were understood to impact candidates' physical and mental constitution, resulting in screening out of overly cautious, anxious, or “sensitive types.”[23]

The Arnprior experiment aimed to assess selection procedures for RCAF and RAF pilots using various tests,[24] including classification,[25] mechanical reasoning, code aptitude, educational achievement (in math and physics), mechanical aptitude, and visual link tests of flying aptitude. These tests were evaluated against RAF grading procedures and flight instructor grades during training. The experiment ultimately concluded that the visual link test was the most reliable predictor of success.[26] This test utilized a modified instrument procedural trainer, which was comparable to a basic simulator. The trainer's dials and gauges moved in accordance with the aircraft's attitude and position relative to navigational aids, providing a rudimentary apparatus for evaluating candidate aptitude in completing basic flying maneuvers using both external references and internal cockpit instruments (hence the term visual link). Interestingly, the visual link test's predictive validity was higher than any other contemporaneous technology used by other militaries, but its cost prevented other air forces from adopting it.

CONTEMPORARY CONTEXT

The post-WWII era saw technological and societal changes in the relevance of air travel and air power.[27] Airlines and air forces worldwide had access to foundational data from years of research such as that already discussed, enabling them to create selection programs with a higher likelihood of choosing candidates capable of operating safely and effectively in the aviation environment. The diversification of military aviation into specialized components, such as transport, bomber, high-performance fighter, rotary-wing, and surveillance aircraft, and the airline industry's adoption of jet aircraft with increased complexity and speed, drove ongoing research into various areas of human performance in aviation. This research aimed to answer questions such as why some individuals were more successful than comparable peers in specific aviation domains, and how rapidly increasing training costs and time could be reduced while maintaining high levels of pilot skill across the required spectrum. Advances in simulation and computer-based testing provided ways of screening candidates for attributes known to correlate with success in aviation. These methods offered greater fidelity and predictive validity than the limited approaches of the past.[28]

---

[23] Mathewson, F.A.L.

[24] Signori, E., ‘The Arnprior Experiment. A Study of World War II Pilot Selection Procedures in the RCAF and RAF’, *Canadian Journal of Psychology* 3, no. Journal Article (1949): 136.

[25] Mathewson, F.A.L., ‘Medical Aspects of Aircrew Selection’.

[26] Signori, E., ‘The Arnprior Experiment. A Study of World War II Pilot Selection Procedures in the RCAF and RAF’.

[27] Willingham, Frank. "Boeing 707 Group: A History." *Air Power History* 66, no. 1 (Spring, 2019): 56-57.

[28] Examples will be discussed in greater detail in the sections that follow.

Military Approach

Motivation. Military air operations have specific training requirements that differ from those of the civilian industry due to their diverse nature. Whether in the Army Air Corps, Navy Fleet Air Arm, or Air Force, military air operations involve purpose-driven mission sets from reconnaissance to air-to-air combat, tactical airlift to battlefield support and more; each with specialized aircraft and training requirements for both rotary and fixed-wing pilots. Consequently, selection systems need to consider a range of potential career paths for applicants. Early training must provide ab initio pilots with the basic skills required for advanced pilot training and select pilots for specific airframes based on undergraduate flying performance and other factors.

While certain psychological requirements may be perceived as universal for commercial multi-crewed aircraft, they may not be relevant for military initial aircrew selection since some successful candidates will be streamed into single-pilot aircraft. Crew Resource Management (CRM) training can be provided later in training for those selected for other airframes. Conversely, certain personality traits that may not be required for commercial operations could be strong predictors of long-term success in military operations. The evolution of military aircrew selection criteria includes unique features that highlight the inherent differences between military and civil aviation.[29] However, maximizing selection and training efficacy and minimizing costs in time and resources is a common feature shared with all aviation environments. In 2014, the average cost of training one undergraduate pilot was around $557,000 USD,[30] and this cost increases significantly when including advanced pilot training, Operational Training Units (OTUs), and other required training for specific squadron positions. Increasing selectivity to reduce training failures is therefore essential not only for operational effectiveness but also for the proper stewardship of government or commercial financial resources.[31]

Aptitude assessments. A particular area of military research into pilot aptitude evaluates the impact of long-term (LTM) and working memory (WM) on pilot performance. Ongoing research has revealed recurring patterns. In a study conducted for the US Air Force Human Resources Laboratory, cognitive organization of pattern recognition by fighter pilots was evaluated as a function of flying experience.[32] The study compared the performance of expert and novice fighter pilots in two classifications of pattern recognition: complex (split-plane) and simple (ground strafing) maneuvers. The goal was to determine if a model exists that can predict how similar a novice may be to an expert in terms of cognitive information retrieval.

---

[29] D. Bartram, 'The Predictive Validity of the EPI and 16PF for Military Flying Training', *Journal of Occupational and Organizational Psychology* 68, no. 3 (1995): 219–36, https://doi.org/10.1111/j.2044-8325.1995.tb00583.x.
[30] Johnson et al., 'Predictive Validity of Spatial Ability and Perceptual Speed Tests for Aviator Training'.
[31] Canada. Deptartment of National Defence, 'Canadian Forces Joint Publication 1-0: Military Personnel Management Doctrine' (Ottawa: Department of National Defence, 2008).
[32] R. W. Schvaneveldt et al., 'Cognitive Organization as a Function of Flying Experience' (Air Force Human Resources Laboratory, 1984).

The study's outcomes were applicable in predicting pilot performance and selecting pilot trainees for job placement.[33]

Decades later, a study of fighter pilot performance evaluated novice and expert fighter pilot situational awareness (SA) in terms of potential confidence bias.[34] The researchers identified features of expert performance, including the ability to "chunk information"[35] and relate it to schemas stored in long-term memory for quicker resolution of novel problems. Novices who had not yet developed these schemas devoted a greater proportion of their cognitive resources to WM to produce similar outcomes.[36] The cognitive mechanisms involved in the same situation were different for novices than for experts, and the corresponding performance in predicting future outcomes under Endsely's definition of level 3 SA were poorer for novices than for experts, even though their perception of the current environment (level 1 SA) was comparable or superior. An explanation of the three levels of SA will follow in the sections to come. Other studies have shown that experts recognize solutions in LTM rather than analyzing and interpreting each situation, which is required of novices who have not yet developed heuristics or schemas and whose (more cognitively demanding) WM resources are therefore taxed to a higher degree.[37] This emphasizes the importance of both LT memory retrieval and WM as desirable pilot aptitudes, as the role of each is dependent on experience.

A 1993 study was conducted to determine the predictive validity of the USAF Pilot Candidate Selection Method (PCSM) by analyzing the results of 678 Air Force pilot training candidates.[38] These candidates were evaluated based on their training success, class ranking, and the results of the Air Force Officer Qualifying Test (AFOQT) and the Basic Attributes Test (BAT). The AFOQT measures general cognitive ability (*g*), verbal and spatial aptitude, perceptual speed, and aircrew aptitude/interest, while the BAT evaluates psychomotor skills, information processing, and risk tolerance using a computer-administered test.

The researchers found that previous flying experience was the most predictive variable, and the AFOQT battery was a better predictor than the skills measured by the BAT. They also discovered that the information processing variable in the BAT was not a valid predictor and should be revised or discarded.[39] Interestingly, the researchers found that attitude towards risk was a valid predictor beyond the AFOQT and should be studied further. In contrast, the psychomotor component of the BAT showed

---

[33] Schvaneveldt et al., 31.
[34] Sulistyawati, K., Wickens, C.D., and Chui, Y.P., 'Prediction in Situational Awareness: Confidence Bias and Underlying Cognitive Abilities', *The International Journal of Aviation Psychology* 21, no. 2 (2011): 153–74, https://doi.org/10.1080/10508414.2011.556492.
[35] Sulistyawati, K., Wickens, C.D., and Chui, Y.P., 169.
[36] Sulistyawati, K., Wickens, C.D., and Chui, Y.P., 'Prediction in Situational Awareness: Confidence Bias and Underlying Cognitive Abilities'.
[37] Vidulich, M.A. et al., 'Information Processing in Aviation', in *Human Factors in Aviation*, 2nd ed. (Academic Press, 2010), 175–215; Randy Gibb, Rob Gray, and Lauren Scharff, *Aviation Visual Perception: Research, Misperception and Mishaps*, Book, Whole (Surrey, England;Burlington, VT; Ashgate, 2010), https://doi.org/10.4324/9781315568584.
[38] Carretta and Ree, 'Pilot Candidate Selection Method (PCSM): What Makes It Work'.
[39] Carretta and Ree, 10.

relatively low incremental validity and was considered *g*-loaded. It's worth noting that the BAT measures psychomotor ability using an alphanumeric keyboard, a monitor, and two control sticks, similar to the Royal Air Force Aircrew Aptitude Test (RAFAAT) but distinct from the apparatus used by the Canadian Automated Pilot Selection System (CAPSS). This distinction is important to consider when evaluating the validity of psychomotor aptitude and its measurement in pilot selection, especially in the context of this research.

In a dissertation released in 2014, Forgues evaluated the cognitive abilities of over 1000 pilot candidates for the Royal Canadian Air Force using the Canadian Forces Aptitude Test (CFAT), the RAFAAT, and the CAPSS. The study aimed to compare the predictive validity of the RAFAAT and CAPSS as selection tools for aircrew. Candidates completed the CFAT before attempting the more advanced test: CAPSS (which emulated the cockpit of a small single-engine aircraft) and the RAFAAT, which evaluated candidates' attentional capabilities, work rate, psychomotor ability, numerical reasoning, and spatial reasoning using computer-based testing software.

Successful candidates shared common features across testing methods, with psychomotor ability being the most dominant predictor of CAPSS candidate success. However, problem solving and spatial ability components on both the CFAT and RAFAAT were also strong contributors to success. Interestingly, those who had higher scores in RAFAAT mathematics were more likely to pass CAPSS which mirrors the results of the study into fighter pilot SA discussed earlier.[40] The high predictive validity of psychomotor performance in CAPSS underscores the importance of the apparatus used for proxy measurement of flying motor skill as the outcome stands in contrast to the earlier study on the predictive validity of the USAF BAT.[41] A feature worth further research and discussion. Ultimately, the attrition rate of CAPSS selection, especially among females,[42] was higher than the RAFAAT, which placed a lower premium on psychomotor ability than other aptitudes.[43] The RCAF now uses the RAFAAT for aircrew selection, including tests of executive function such as strategic task management, switching capabilities, cognitive updating skills, and system analysis capability, which may be necessary in advanced pilot training on more complex airframes.

The Multidimensional Aptitude Battery (MAB) is an aptitude testing battery used in the USAF, NASA's astronaut selection, and civilian airlines to broadly measure intellectual ability in various areas, such as verbal comprehension, arithmetic, digit symbol coding, picture completion, and spatial ability.[44] A 2013 study conducted by King et al. aimed to investigate the usefulness of the MAB and a neuropsychological test called MicroCog in predicting the training outcomes, flying grades, and class

---

[40] Sulistyawati, K., Wickens, C.D., and Chui, Y.P., 'Prediction in Situational Awareness: Confidence Bias and Underlying Cognitive Abilities'.
[41] Carretta and Ree, 'Pilot Candidate Selection Method (PCSM): What Makes It Work'.
[42] Susan L. Forgues, 'Aptitude Testing of Military Pilot Candidates' (ProQuest Dissertations Publishing, 2014), 47.
[43] Forgues, 73.
[44] Paul Retzlaff, J. D. Callister, J.D., and King, R.E., 'Clinical Procedures for the Neuropsychological Evaluation of U.S. Air Force Pilots', *Military Medicine* 164, no. 7 (1999): 514–19, https://doi.org/10.1093/milmed/164.7.514.

rank of USAF pilots undergoing pilot training on the T-6. The study involved 12,924 pilots assessed using the MAB and 5,582 pilots assessed separately using the MicroCog for predictive validation.[45] The MAB used Visual IQ, Performance IQ, and Full Scale IQ (FSIQ), while the MicroCog used all nine indices of the computer-administered neuropsychological test (attention control; reasoning/calculation; memory; spatial processing; information processing speed; reaction time; information processing accuracy; general cognitive proficiency; general cognitive functioning). Both tests measured general cognitive ability, and the overall IQ results were range restricted, with a higher mean IQ (120) and lower variance than normative values across both graduates and eliminees.[46] The study found that graduates had statistically significant mean score comparisons compared to eliminees in single-tailed tests at the 0.05 level, and the MicroCog had statistically significant correlations with academic results. However, the researchers noted that the discriminatory power of the tests is limited due to the range restriction, and additional measures of cognitive ability, motivation, aviation knowledge and experience, and psychomotor skills are necessary to discriminate high ability candidates.

Although the impact of *g* on aircrew performance is apparent, its limited accuracy as a sole predictor of success necessitates a nuanced approach that includes additional testing of specific aptitudes within the general category of cognitive ability. A 2017 meta-analysis of over 10,000 aircrew across multiple AFOQT evaluations revealed that the highest mean validities for USAF pilot trainees were aviation knowledge, perceptual speed, spatial ability, and (consistent with other research) "math ability also predicted pilot-related outcomes."[47] Their study also demonstrated that while spatial ability alone had minimal predictive validity compared to math or aviation knowledge, combining perceptual speed testing increased the predictive validity of spatial ability in pilot training performance.

The Direction Orientation Task (DOT) is used by the US Air Force (USAF) and US Navy (USN) to assess spatial rotation and cognitive processing speed. Keiser et al. (2019) analyzed an updated version of the DOT and found that it exceeded the predictive validity of other subtests of the Aviation Selection Test Battery used by the US Navy, Marine Corps, and Coast Guard.[48] This highlights the importance of spatial awareness and perceptual speed as predictors of pilot performance; as also noted in a meta-analysis of 50 studies across 11 nations that found specific cognitive tests, such as spatial ability and mechanical comprehension, had a higher mean validity than *g* alone, and the combination of cognitive and psychomotor abilities produced the

---

[45] Raymond E. King et al., 'Standard Cognitive Psychological Tests Predict Military Pilot Training Outcomes', *Aviation Psychology and Applied Human Factors* 3, no. 1 (2013): 28–38, https://doi.org/10.1027/2192-0923/a000040.
[46] King et al., 33.
[47] Johnson et al., 'Predictive Validity of Spatial Ability and Perceptual Speed Tests for Aviator Training', 113.
[48] Heidi N. Keiser et al., 'Updating the Direction Orientation Task: An Aviation Selection Tool', *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, no. 1 (2019): 1414–18, https://doi.org/10.1177/1071181319631451.

highest predictive validity.[49] Interestingly, the meta-analysis showed low predictive validity with personality, which stands in contrast to the results of multiple studies and underscores the importance of selecting the right personality screening test or combination of tests.

Personality. In 1998, a factor analysis of the USN's Biographical Inventory (BI) and Aviation Interest (AI) questionnaires (completed by candidates as part of the ASTB) was completed to assess their utility in pilot selection. The purpose was to measure various traits, such as military interest, adventurousness, engineering background, life experiences, and athletic orientation to address the cost of training failures and pilot attrition due to lack of motivation or interest in the flight program.[50] A 2005 study conducted by Boyd et al. used the NEO Personality Inventory-Revised (NEO-PI-R) and the MAB to investigate the relationship between personality differences and the type of aircraft flown. The study found that emotional stability and personal motivation are critical factors but are difficult to assess compared to cognitive and physical abilities. The study also found statistically significant differences in personality between pilots flying different aircraft types, with fighter pilots scoring lower in agreeableness and higher in conscientiousness than other pilots.[51]

A research study was conducted on UK Army Aircrew applicants, which aimed to compare their responses to the Eysenck Personality Inventory (EPI) and 16 Personality Factor (16PF) questionnaires with those of a sample of the general population and a sample of amateur civilian pilots.[52] The study involved collecting responses from all applicants attending selection, categorizing them into those who were selected and those who were eventually successful in training. The findings showed that there was little difference in terms of predictive validity of performance between the applicants, but significant differences existed between the applicants and the general population. Even those who were unsuccessful in selection or training had higher emotional stability, lower neuroticism, and slightly higher extraversion than both the general public and amateur aviators. This led the researchers to conclude that self-selection exists among the population, which draws certain personality types to apply for military aviation specifically, rather than aviation in general.

The USAF Personnel Research Division also conducted a study on aircrew selection, which compared different personality theories. The research ultimately rejected the 16PF theory in favor of the widely used five-factor theory (NEO PI-R).[53] The authors found that any personality theory used in aircrew selection needed to be sensitive enough to discriminate between homogeneous aircrew candidate groups. Structural theories like three-factor theory, instrumentality/expressivity, and temperament theory lacked detail, while others were not fully operationalized. 16PF was seen as

[49] Martinussen, M., ‘Pilot Selection: An Overview of Aptitude and Ability Assessment’.
[50] Sean Biggerstaff, ‘Factor Analysis of the U. S. Navy’s Aviation Interest Subtest’ (PENSACOLA FL: US NAVY ADVANCEMENT CENTER PENSACOLA FLNAVY ADVANCEMENT CENTER, 1998).
[51] Boyd, J.E., Patterson, J.C., and Thompson, B.T., ‘Psychological Test Profiles of USAF Pilots before Training vs. Type Aircraft Flown’, *Aviation, Space, and Environmental Medicine* 76, no. 5 (2005): 463–68.
[52] Bartram, ‘The Predictive Validity of the EPI and 16PF for Military Flying Training’.
[53] Larry A. Pedersen et al., ‘Personality Theory for Aircrew Selection and Classification’ (DAYTON OH: UNIVERSAL ENERGY SYSTEMS INC, 1992).

difficult to replicate and of questionable validity.[54] The airline industry has also extensively researched personality theory as it applies to airline pilot selection and CRM.

Civil Aviation Approaches

Civil aviation shares many basic aptitude requirements with military aviation, but also has unique characteristics that set it apart. Civil aviation includes corporate, commercial, private, and airline sectors, each with its own sub-categories. In contrast to private or general aviation, airline industry pilot selection incorporates various selection criteria, including aptitude testing, personality evaluation, and prior knowledge assessments.

Hoffman and Hoffman (2016) developed a proprietary method for pilot selection and assessed over 10,000 candidates for major airlines in the US. They required applicants to provide a detailed account of their understanding of airline operations, similar personality requirements to those used in military selection systems, as well as an additional aeronautical decision-making and leadership component. Unlike ab initio selection, airline selection is targeted towards specific aviation knowledge since training has already occurred. The selected applicants typically undergo a type conversion to an airframe and begin operations as a first officer or, in some cases, as direct captains. This allows for additional CRM, communication, and leadership experience and aptitude to be evaluated in aviation-specific testing.

Hoffman and Hoffman found that the big-five personality factors tested using the NEO-PI-R are valid, with particular attention given to applicants' sensitivity to criticism, depression, anxiety, and insecurity. The authors discovered that pilots in the bottom 20% of their testing averaged a six-fold greater likelihood of encountering problems in training than the top 40% and cost more money to train, lost more productivity days, and were more likely to miss work over a 12-year span post-hiring.[55] A 2016 study also explored the utility of screening for dependability, stress tolerance, and motivation using Minnesota Multiphasic Personality Inventory testing (MMPI-2).[56] However, it is essential to note that some pilot applicants may present well in psychological evaluations, and their answers may not correlate to actual underlying psychological conditions. That is to say that externalization of blame and defensiveness that would normally indicate a potential for failure to conform to regulations may be concealed by some individuals,[57] requiring a test/re-test scenario for those applicants who invalidate their MMPI-2 test with unrealistically positive or defensive responses.

The impact of test/retest scores on aptitude testing is worth discussing. A 1997 study found a consistent and significant increase in aptitude scores among RAF pilot and navigator applicants who were retested at the Officer and Aircrew Selection Centre

---

[54] Pedersen et al., 35.

[55] Hoffman, C. and Hoffman, A., ‘The Role of Assessment in Pilot Selection’, in *Pilot Mental Health Assessment and Support* (Routledge, 2016).

[56] James N. Butcher, ‘Psychological Assessment of Airline Pilot Applicants With the MMPI-2’, *Journal of Personality Assessment* 62, no. 1 (1994): 31–44, https://doi.org/10.1207/s15327752jpa6201_4.

[57] Butcher.

(OASC), regardless of the length of time between tests (up to five years).[58] This improvement in performance was observed even without additional coaching and persisted over time, contradicting the assumption of decay in positive effects. This finding is particularly relevant for administrators of selection programs dealing with applicants who have previously completed portions of aptitude or psychological testing.

Alternatively, testing methods that yield asymptotic performance over time may offer insights into innate ability and peak performance unrelated to previous exposure. O'Hare (1997) compared the performance of three groups on a computer-based stress-tolerance and situational awareness test named WOMBAT, which involved pattern recognition, spatial awareness, attention management, and working memory. The groups included a random sample from the general population, experienced pilots, and elite professional pilots. While the general population's performance plateaued by the end of the testing session, the experienced and elite pilots performed significantly better (demonstrated by chi-square test at the 0.05 level).[59]

Situational Awareness

It is worth discussing situational awareness (SA) in some detail to provide a justification for the inclusion of additional testing for pilot candidates. SA is generally categorized into three levels first suggested by Endsley in 1988. Level 1 SA refers to an individual's ability to perceive relevant elements within a finite time and space; level 2 is the accurate comprehension of their meaning; and level 3 is the more abstract ability to project their future status.[60] With the exception of experimental validation, the simple term SA refers to the combination of all three levels. In the three-dimensional, time compressed environment of limitless variables in which pilots operate, a high degree of SA is required to maintain awareness of critical parameters, prioritize effectively, and think far ahead of an aircraft which may be travelling in excess of the speed of sound. Maintaining SA over the course of a mission involves multiple cognitive processes, yet SA may be lost in moments with potentially catastrophic consequences. High levels of SA have been strongly correlated to expert pilot performance[61] while the loss of SA has been a notable contributing factor in numerous air accident investigations.[62] Understanding the core brain functions involved in developing and maintaining SA is an important component of effective aptitude testing for pilots that will operate in modern,

---

[58] Eugene F. Burke, 'A Short Note on the Persistence of Retest Effects on Aptitude Scores', *Journal of Occupational and Organizational Psychology* 70, no. 3 (1997): 295–301, https://doi.org/10.1111/j.2044-8325.1997.tb00649.x.

[59] David O'Hare, 'Cognitive Ability Determinants of Elite Pilot Performance', *Human Factors* 39, no. 4 (1997): 548, https://doi.org/10.1518/001872097778668004.

[60] Endsley, M.R., 'Design and Evaluation for Situational Awareness Enhancement.', in *Proceedings of the Human Factors Society 32nd Annual Meeting*, 1988, 97–101; Mica R. Endsley, 'Toward a Theory of Situation Awareness in Dynamic Systems', *Human Factors* 37, no. 1 (1995): 32–64, https://doi.org/10.1518/001872095779049543.

[61] Adams, M.J., Tenney, Y.J., and Pew, R.W., 'Situation Awareness and the Cognitive Management of Complex Systems', *Human Factors* 37, no. 1 (1995): 85–104.

[62] Te Koeti, T., 'ASCI 670 Research Paper: Assessing Situational Awareness' (Embry-Riddle Aeronautical University, 2020).

information-dense, high-performance aircraft, particularly given the dynamic flight profiles they will be expected to fly during military flight training and on operations.

There is a considerable body of knowledge on cognitive functions that correspond well to the resiliency of an individual's SA; particularly EF and Perceptual Speed (PS) as they pertain to divided attention tasks, fixation, WM, and perceptual load. WM is commonly considered to be a requirement for developing and maintaining SA[63] with individuals who test higher in WM also demonstrating superior performance in SA testing such as Situation Awareness Global Assessment Technique or SAGAT.[64]

WM, accepted as a contributor to SA,[65] can be evaluated in its constituent elements by measuring EF performance in individuals. Expert performance can be related to the utilization of heuristics and schema stored in the LTM and categorized by chunk or frame into key nodes that are easily retrievable and whose recollection engages related schema.[66] These may be episodic, semantic, or procedural in nature.[67] However, the novice must rely on greater WM resources while generating these schemas over time, therefore those pilots with greater WM resources may be at a cognitive advantage, especially during development or when encountering novel technologies and situations. By selecting pilots based on cognitive abilities that are likely to yield greater SA, the conduct of safe and effective operations are maximized during the period of learning required to become expert aviators,[68] while also leveraging neuroplasticity during training to ensure the most effective development of cognitive maps for recognition-primed decisions that assist in maintaining level 3 SA.

WM is comprised of lower-level cognitive processes including EFs which can be isolated for testing with greater fidelity than proxy measurements of SA. Perceptual speed (PS) is another enabler of pilot performance though the degree to which WM, EF, and PS interact has some interesting consequences for aptitude testing and performance prediction. Baddeley's widely accepted model for short-term memory storage (STS) subdivides STS processes into a visuospatial sketchpad, a phonological loop and a central executive.[69] These subcomponents contribute resources to the WM when an individual is manipulating an image or object and constitute a greater degree of cognitive demand than LTM retrieval. This is a critical requirement for pilots and many aircrew aptitude tests feature some variation of WM evaluation. Drilling further into the Baddeley model we can further subdivide the

---

[63] Adams, M.J., Tenney, Y.J., and Pew, R.W., 'Situation Awareness and the Cognitive Management of Complex Systems'; Endsley, 'Toward a Theory of Situation Awareness in Dynamic Systems'.
[64] Sulistyawati, K., Wickens, C.D., and Chui, Y.P., 'Prediction in Situational Awareness: Confidence Bias and Underlying Cognitive Abilities'.
[65] Endsley, 'Toward a Theory of Situation Awareness in Dynamic Systems'.
[66] Wongupparaj, P., Kumari, V., and Moris, R.G., 'The Relation between a Multicomponent Working Memory and Intelligence: The Roles of the Central Executive and Short-Term Storage Functions', *Intelligence* 53 (2015): 166–80, https://doi.org/10.1016/j.intell.2015.10.007; Sulistyawati, K., Wickens, C.D., and Chui, Y.P., 'Prediction in Situational Awareness: Confidence Bias and Underlying Cognitive Abilities'; Schvaneveldt et al., 'Cognitive Organization as a Function of Flying Experience'.
[67] Goldstein, B., *Cognitive Psychology Connecting Mind, Research, and Everyday Experience*.
[68] Adams, M.J., Tenney, Y.J., and Pew, R.W., 'Situation Awareness and the Cognitive Management of Complex Systems'; Forgues, 'Aptitude Testing of Military Pilot Candidates'.
[69] Goldstein, B., *Cognitive Psychology Connecting Mind, Research, and Everyday Experience*.

STS process involving the central executive system (CES) into the three EFs known as inhibiting, shifting, and updating.[70] These EFs feature prominently in various pilot aptitude tests including some of the sub-tests of the RAFAAT. Perceptual Speed (PS) by contrast refers to an individual's ability to scan, recognize, compare and analyze shapes, patterns and images.[71] Numerous studies have investigated WM as it relates to *g*, EF, and SA, while PS has been correlated to some of the processes involved in WM.[72] The research is less comprehensive on the association between each EF, PS and SA; an area addressed by the author in a previous study which suggests SA capacity may be predicted through more exact analysis of simpler cognitive processes of EF and PS. Therefore, a brief precis on EF and PS will follow.

Executive functions. A study of brain activity using Electroencephalogram (EEG) isolated the various regions associated with each WM process according to the Baddeley model with the left and right hemispheres corresponding to the phonological loop and visuospatial sketchpad respectively; while the CES involved both hemispheres frontal lobes and LTM engaged cross-hemispheric parietal lobes. WM tasks, such as manipulating recently learned shapes, resulted in greater activation of cross-hemispheric frontal lobes, and greater EF loads as compared with LTM recall which primarily activated the parietal lobes.[73] A separate study examined the relationship between the CES, the three EFs, and both crystallized general intelligence (*g*C) and fluid (*g*F) validating the model and linking all three WM processes with *g* (nondirectionally). Furthermore, an interesting relationship between verbal STS and *g*C was noted and referenced by other studies.[74] The relationship between *g* and WM is echoed in other studies that have correlated *g* with high SA individuals. A different study that also leveraged EEG to analyze brain activation further validated the CES model and examined the relationship between updating and inhibiting functions under a varying perceptual load. Common parietal and frontal lobe activation suggested a commonality to all WM functions but of particular interest was the finding that inhibition (protection against distraction) increased under greater WM load. This may suggest that controlled attention processes may activate under increased WM strain enhancing inhibition by proxy. A useful feature for pilots prioritizing which information must receive undivided attention in the presence of many distractors[75] as is neatly summarized in the passage that follows.

"Technological advancements in today's combat aircraft increase the demands on pilots, often requiring that their attention be split between multiple tasks. When divided

---

[70] Christian Scharinger et al., 'When Flanker Meets the N-Back: What EEG and Pupil Dilation Data Reveal about the Interplay between the Two Central-Executive Working Memory Functions Inhibition and Updating', *Psychophysiology* 52, no. 10 (2015): 1293–1304, https://doi.org/10.1111/psyp.12500.
[71] Jaime Arguello and Bogeum Choi, 'The Effects of Working Memory, Perceptual Speed, and Inhibition in Aggregated Search', *ACM Transactions on Information Systems* 37, no. 3 (2019): 1–34, https://doi.org/10.1145/3322128.
[72] Arguello and Choi.
[73] Sauseng, P. et al., 'Fronto-Parietal EEG Coherence in Theta and Upper Alpha Reflect Central Executive Functions of Working Memory', *International Journal of Psychophysiology* 57 (2005): 97–103.
[74] Wongupparaj, P., Kumari, V., and Moris, R.G., 'The Relation between a Multicomponent Working Memory and Intelligence: The Roles of the Central Executive and Short-Term Storage Functions'.
[75] Scharinger et al., 'When Flanker Meets the N-Back: What EEG and Pupil Dilation Data Reveal about the Interplay between the Two Central-Executive Working Memory Functions Inhibition and Updating'.

attention is coupled with stressful or mentally demanding situations, a potential for mental overload presents itself. Studies of fighter pilots show how devastating the effects of mental overload can be. These pilots can become so involved in their current situation that they forget to perform critical tasks, such as G-force straining maneuvers. As a result, some pilots have lost consciousness and their lives."[76]

Another recurrent feature of SA prediction involves the relationship between mathematical ability and level 3 SA noted by two studies cited earlier in this paper. A 2013 meta-analysis on the association of WM with mathematical ability confirmed the prevailing science linking mathematical ability with all measures of WM.[77] The cross-cultural, multiple measure study involved various age groups and further underscored the correlation of the verbal updating EF, as compared with the inhibiting and shifting EFs. There seems to be recurrent themes emerging from very different studies that highlight the peculiar contribution of verbal STS/EF, mathematical ability, *g*, and SA.

Another unintuitive finding from EF research was the influence of emotional regulation (ER) capacity of an individual on their EF effectiveness. This finding is informative for psychological evaluation as it pertains directly to key pilot aptitudes. Research shows that the amygdala (responsible for negative emotion) can suppress areas of the brain responsible for cognitive functioning. Individuals capable of regulating their emotions effectively have higher EF effectiveness, a factor evident even in young children. Since the regions of the brain responsible for ER develop before the regions responsible for EF, emotionality at a young age may be a predictor of later EF, when it becomes a more reciprocal relationship.[78] A study into the effects of emotional regulation training on EF and *g*F (measured by Stroop and digit span tests) provided further evidence on the complimentary role of ER on EF tasks. The research demonstrated an improvement in WM and *g*F of those trained to inhibit emotional stimuli as compared with the control group.[79]

Perceptual speed. The predictive validity of PS on pilot ability was a driver behind the previously mentioned DOT by the USAF and the USN which includes measures of PS and which has greater predictive validity than any other test in the ATSB used by the USN, Marines and Coast Guard.[80] A 2019 study on the connection between WM, inhibition, and PS involved a search task where participants were required to process information under time pressure. The authors found that individuals with high PS had

---

[76] Jeremy B. Noel, Kenneth W. Bauer Jr, and Jeffrey W. Lanning, 'Improving Pilot Mental Workload Classification through Feature Exploitation and Combination: A Feasibility Study', *Computers & Operations Research* 32, no. 10 (2005): 2713.

[77] Friso-van den Bos, I et al., 'Working Memory and Mathematics in Primary School Children: A Meta-Analysis', *Educational Research Review* 10 (2013): 29–44, https://doi.org/doi:10.1016/j.edurev.2013.05.003.

[78] David E. Ferrier, Hideko H. Bassett, and Susanne A. Denham, 'Relations between Executive Function and Emotionality in Preschoolers: Exploring a Transitive Cognition-Emotion Linkage', *Frontiers in Psychology* 5, no. Journal Article (2014): 487–487, https://doi.org/10.3389/fpsyg.2014.00487.

[79] Susanne Schweizer, Adam Hampshire, and Tim Dalgleish, 'Extending Brain-Training to the Affective Domain: Increasing Cognitive and Affective Executive Control through Emotional Working Memory Training', ed. Manos Tsakiris, *PloS One* 6, no. 9 (2011): e24372–e24372, https://doi.org/10.1371/journal.pone.0024372.

[80] Keiser et al., 'Updating the Direction Orientation Task: An Aviation Selection Tool'.

a faster sorting ability than those with low PS. Moreover, people with high WM and PS were able to process a larger volume of information in a given time and were less likely to engage in satisficing during decision making. Additionally, high PS individuals reported lower workloads and worked more intensively than their low PS counterparts. The study also highlighted differences in search patterns utilized by high-PS individuals, who were better at identifying relevant cues interleaved within irrelevant material. This relationship was also evident in the researchers' analysis of the inhibition executive function. The research appeared to confirm a correlation between WM and workload, and between PS and search patterns, but not between WM and PS themselves suggesting that they involve different cognitive functions.[81] This finding is confirmed by a 2012 study that showed a correlation between WM and *g*F (similar to our previous findings) but not between WM and PS.[82] Finally, if further evidence was required of the recurring importance of verbal IQ that has repeatedly emerged as a contributor to performance in the preceding sections; it was also associated with inspection time (IT), which is a quotient for perceptual speed. An interesting feature of this 2005 study into IQ and IT (which involved over 2000 twins from multiple nations) was a conclusion that there may be common genetic factors to IT and IQ (particularly verbal IQ).[83]

RCAF MODEL

A rigorous study completed by Forgues in 2014 outlined in detail the differences between the legacy CAPSS (which was the chosen method during her study) and the replacement RAFAAT selection methods. In her study, Forgues analyzed the specific aptitudes that each method examined and contrasted each with the other in terms of their predictive validity for pilot success during basic and advanced training. At the time of her study, the RAFAAT version under assessment was the PB06 which was not as comprehensive as its current form (PH11) and she astutely noted its various deficiencies. Regarding WM, she noted:

"tests of other aptitudes considered important for pilots, including WM, situational awareness, and decision making (Wickens, 2007), are missing in the RAFAAT battery . . . The RAF considers Digit Recognition in the Attentional Capability domain to be a test of WM, but testing candidates on their ability to remember how many times a specific digit appeared in a previously viewed number string is a low-level WM task."[84]

Of situational awareness:

"There are no RAFAAT tests that specifically assess situational awareness. The Comprehension 2 subtest, part of the Spatial Reasoning domain, is similar to the test

[81] Arguello and Choi, 'The Effects of Working Memory, Perceptual Speed, and Inhibition in Aggregated Search'.
[82] Thomas S. Redick et al., 'Faster, Smarter? Working Memory Capacity and Perceptual Speed in Relation to Fluid Intelligence', *Journal of Cognitive Psychology (Hove, England)* 24, no. 7 (2012): 844–54, https://doi.org/10.1080/20445911.2012.704359.
[83] Luciano, H.N. et al., 'Perceptual Speed Does Not Cause Intelligence and Intelligence Does Not Cause Perceptual Speed', *Biological Psychology* 70 (n.d.): 1–8, https://doi.org/10.1016/j.biopsycho.2004.11.011.
[84] Forgues, 'Aptitude Testing of Military Pilot Candidates', 70.

Sohn and Doane (2004) used in their situational awareness study, however, the Instrument Comprehension subtest is missing the critical temporal component. As such, Instrument Comprehension is included in the Spatial Reasoning domain leaving situational awareness largely untested by the RAFAAT battery."[85]

The staff at the Canadian Forces Aircrew Selection Center (CFASC) responded to inquiries about potential shortcomings to the RAFAAT with the clarification that the version of the RAFAAT under assessment during the study was denoted PB06, and in the current PB11 designation SA and WM are tested.[86] The degree to which EF is evaluated remains unclear but the following was discussed in Forgues thesis:

"Causse et al. (2011) identified EF as a critical component of the complex and constantly changing air environment in which a pilot operates, providing support for its inclusion in pilot selection batteries. While the subtests of the CFAT and RAFAAT do not specifically identify EF as one of the cognitive constructs being assessed, its components as described by Diamond (2013) and Miyake et al. (2000), appear to be present. For example, the RAFAAT subtest Colours, Letters, and Numbers in the Attentional Capability domain assesses the EF components of inhibition, WM, and shifting. Although this subtest was not statistically significant in any of the analyses completed for this research, the development of ability tests that focus on situational awareness, selective search, and switching attention between tasks should be a priority for future pilot selection research."[87]

The PB11 RAFAAT adopted by the CAF assesses seven ability domains: Strategic Task Management; Perceptual Processing; Short Term Memory and Capacity; Symbolic Reasoning; Central Information Processing; Spatial Reasoning; and Psychomotor Ability domains. There are also subtests that assess switching capabilities, cognitive updating skills, and system analysis capacity. This synopsis was confirmed by the CFASC in response to the concerns raised following Forgues' analysis.[88] Further study may be warranted to determine the depth and effectiveness of the PB11 WM, SA and EF subtests to ensure that they are sufficiently comprehensive and that the cut-off criteria for success is set to the appropriate level. This includes the requirement for combined spatial awareness and perceptual speed tests, and combined psychomotor/cognitive evaluation. CFASC also confirmed that no personality testing is completed at CFASC; that any personality testing during the pilot selection process is completed at the recruiting centres.

Performance

There have been enough candidates selected using the RAFAAT (PB11) since 2013 to make a performance comparison with those selected using CAPSS. However, the comparison is confounded by numerous variables. Notably, delays in training and multiple possible career paths and entry plans mean that candidates selected under the CAPSS model were peppered throughout training courses alongside their RAFAAT course mates for several years. This introduces problems for the strict

[85] Forgues, 71.
[86] Scott McPhalen, 'Executive Function Assessment', 27 January 2023.
[87] Forgues, 'Aptitude Testing of Military Pilot Candidates', 71.
[88] Scott McPhalen, 'Executive Function Assessment', 27 January 2023.

comparison of course average GPA, or of the measurement of performance averaged over timelines that are useful for analysis (e.g. all student performance over the span of each training year). The author acknowledges this problem and was unable to retrieve personal data on all individual candidates over the previous decade to determine exactly which selection method applied to each student pilot. A compromise that still yields useful information is to widen the experimental timeline aperture to include candidates from 2012 till 2022.

Those selected in 2013 under the RAFAAT did not arrive at Ph II pilot training in large numbers till approximately 2016 in and many case much later. The student pilots who dominated (by proportion) the Ph II pilot training courses at 15 Wing before 2016 were those selected prior to 2013. There will be an unknown period anchored around 2016 where candidates from both selection models were under training simultaneously, and an even longer period for which stragglers from the legacy selection may have rejoined later courses (having been delayed due to injury or other reasons) that were predominantly comprised of RAFAAT-selected student pilots. Despite this feature, an analysis of the number of successful completions per year yields useful information. Between 2012 and 2016 (inclusive) there were 523 completions of Ph II pilot training and between 2017 and 2022 (inclusive) there were 389.[89] The overlap of some students will be insufficient when comparing the size of the sample populations to account for any large differences noted, though small differences will be more difficult to discern and justify.

There are other confounding factors that affect our ability to accurately compare student performance over time. These include changes to the application of Integrated Training Plans (ITPs) of both the Flight Instructor Course (FIC) and the assessment of Ph II pilot training;[90] the variability of instruction according to the experience level of the cadre posted to NFTC; serviceability of aircraft and its second order effects on flying continuity and student performance; A propensity towards grade inflation in recent years (which will be discussed in greater detail in the sections that follow); and the deleterious effect of Covid restrictions to the training cycle.

During the author's time as a Qualified Flight Instructor (QFI) at NFTC there were several trends noted for which little supporting documentation exists. The author instructed on both the CT156 Harvard II and CT155 Hawk aircraft and taught all phases of flying (Ph II, Ph III, Ph IV Trans (Hawk), and the international Ph III Hawk course for Royal Singaporean Air Force student pilots). The author also conducted numerous Progress Review Boards (PRBs) of students who had met course fail criteria on each of the aforementioned training courses and conducted the selection courses which streamed student pilots from Ph II into either jet/instructor,

[89] From the NFTC dataset.

[90] In some cases, ITPs were not rewritten, but the way they were interpreted to apply changed over time. Instructional prerequisites were relaxed to allow instructor pilots to qualify as junior "C Category" QFIs without having met the standard required to teach Ph III (a prior requirement) and would then complete the other upgrades after gaining experience teaching Ph II. In this way, the requirements were split into two separate courses which allowed QFIs to teach Ph II as a means of gaining experience. The transition to this model occurred circa 2016.

multi-engine, or helicopter career paths. The author was posted to 2 CFFTS (NFTC) from 2012 to 2019.

It was noted circa 2016 that a propensity was building to debrief items to the level required if the level was not achieved on the flying sortie and to apply an overall grade to students that exceeded the performance actually demonstrated out of an aversion to fail students on Ph II.[91] Errors that would have resulted in failure only a couple of years prior were increasing in frequency but were not assessed to the previous level of scrutiny and were not accurately captured by appropriate grading according to the FIC guidelines and the course ITP.[92] Particularly egregious cases were captured in other documentation, but many were not appropriately managed and the end of course evaluation and GPA was not reflective of the demonstrated performance. Other changes include the removal of some sequence items from the Ph II syllabus over time (such as Practice Forced Landings or PFLs, and assessment of student pilots' ability to lead formations) due to the difficulty level and the number of failures that were associated with specific sequence items. This is problematic from a statistical analysis perspective as comparing average GPA of more recent courses with those a decade earlier is not a peer-to-peer comparison.

There is an insidious justification for the gerrymandering of student evaluation in recent years. There has traditionally always been an attrition rate during pilot training which is skewed towards earlier phases (predominantly Ph I and Ph II) for which even the best selection systems were unable to account. Obviously, the goal of an effective selection system is to reduce the attrition rate to the maximum extent so that training resources are not wasted. The process by which a student is removed from training includes a formal PRB during which all aspects of training are evaluated, including the quality of instruction, degree to which delays may have affected student proficiency, weather considerations, flying ability, learning curve and so on. Among these considerations is an assessment of the student pilot's KSAs based on many of the same proven concepts already discussed (DM, SA, hand-eye-foot coordination, etc.).[93] After a comprehensive and impartial review is completed (which can range from days to weeks in complex cases) a recommendation based on the assessment of all involved is made to the unit's Commanding Officer who has the authority to support or overturn the decision which can be either Cease Train (CT), re-course, or request additional time in excess of the allotted course flying hours. If a decision to CT is supported, the student still has an opportunity to grieve the result to an independent board convened external to the Wing.

In recent years a disturbing trend evolved whereby a significant proportion of grievances (for which staff at the independent review stage re-evaluated the unit's recommendation for CT and agreed that it was the right choice) were nevertheless

[91] Debriefing to the level was not a new concept and could be reasonably completed for minor procedural errors uncharacteristic of the overall student performance. But should not be applied in situations where there are clear safety implications.

[92] 2 CAD, 'NATO Flying Training in Canada Integrated Training Plan' (2 CDN Air Div HQ AF Trg, 2014); CFS, *Flight Instructor's Handbook*, A-PD-050-001/PF-001 (Central Flying School, 2005).

[93] CFS, *Flight Instructor's Handbook*.

overturned and the training system was directed to recourse those individuals.[94] This significantly increased the cost of training these candidates in comparison with their peers; yet even after recourse, and the preferential use of only the most senior instructors, the preferential scheduling for consistency over their peers and the addition of many more flight hours of training in comparison to the assessed learning curve of their course mates, met course fail criteria a second time and, in some cases, a third time. The resources consumed during the conduct of a single PRB cannot be understated and come at the expense of other students and of the training unit whose resources are already stretched to the brink in the conduct of regular operations. The net result was an inflation of grades at the unit level in the knowledge that in all but the most egregious cases, students would likely be re-coursed and the resources expended during PRB were too precious to use unless the most iron-clad case could be made. The burden of proof became almost too heavy to bear and it became necessary to ensure some students only flew with the most senior instructors in the best possible conditions because a future PRB would inevitably disregard the assessment of junior instructors or place undue emphasis on the weather conditions or other factors. To be fair, the PRB process exists for a reason, and there have been many cases where students met course fail criteria, and the PRB uncovered training deficiencies which when corrected allowed the students to successfully complete their courses and undertake their flying careers. The process itself is designed to uncover deficiencies in the training system, but the expert assessment of the many individuals involved in the process must not be disregarded along the way.

Assessment

With all the complicating factors that muddy the waters around a true assessment of student performance, one possible way forward is to compare the frequency of PRBs year over year with the number of course completions over the same period. This eliminates GPA as a stand-alone variable, and it yields a return on investment in terms of the resources expended to produce a given number of successful Ph II students who will go on to more advanced training. Using a dataset provided by NFTC the frequency of PRBs is fairly consistent over time, but the number of course completions is trending downwards (significantly) over time. Therefore, the relative proportion of PRBs to course completion is increasing, indicating a much larger resource expenditure per student pilot completion.

The data on course completions shows a strong negative correlation over time (r = -0.75), with a particularly significant decline in completions after 2016. It should be noted that the capacity of NFTC to absorb new students has steadily diminished for many of the confounding factors already described, so the negative trend is not purely an indicator of student performance. The raw number of PRBs over time has remained fairly constant with a mean (and median) of 28 and a *SD* of 7.90 over the assessment period (2012-2022).[95] Therefore, the PRB rate as a proportion of total completions is

[94] During discussions with a 17 Wing member whose role was to evaluate grievances from unsuccessful 15 Wing Ph II pilot training candidates, it was confirmed that during the mbr's posting to that position all of the PRB recommendations for CT that were supported by the grievance analysts were overturned under the direction that 2 CFFTS would recourse the pilots in question.

[95] Raw dataset is available upon request.

increasing over time with a positive correlation of $r = 0.63$. For example, in 2022, there were 24 PRBs and only 33 completions. If we remove the Covid years from the analysis, the correlation between PRB rate over time is insignificant, although there is still a positive trend in the relative proportion.

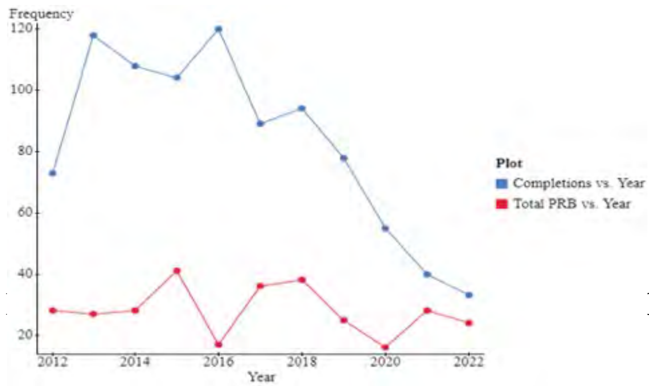

Figure 1. …)22.

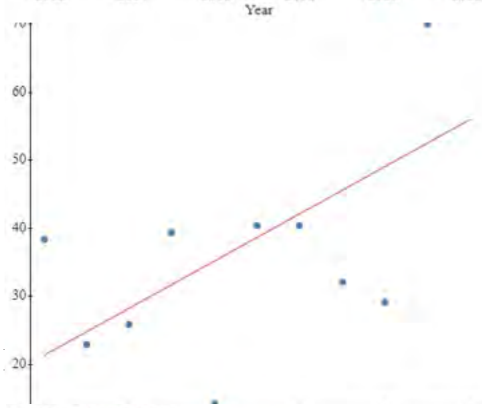

Figure 2. … tal Completions, 2012-2020.

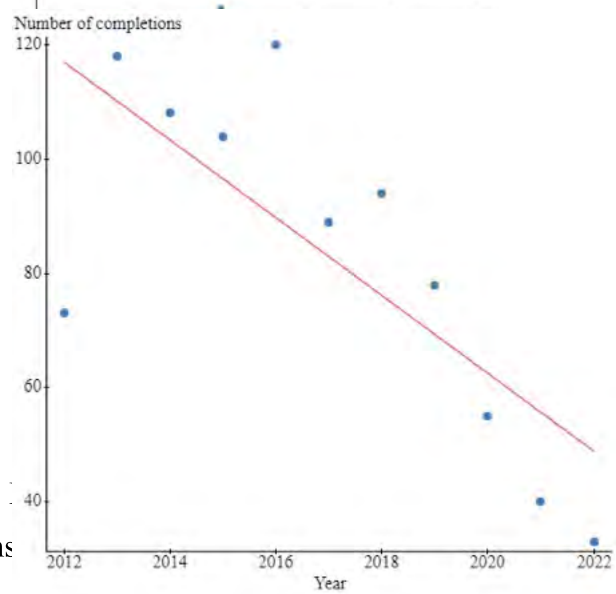

Figure 3. …

GPA Cons…

It is worth considering the subgroup of pilots from our sample who attempted the CF188 Operational Training Unit (also referred to as Fighter Pilot's Course or FPC) prior to the Covid pandemic. Those who were streamed into single-pilot high-performance cockpits based on their Ph II flying performance either became QFIs or completed additional training at 419 Squadron (Fighter Lead-In Training or FLIT) before attending FPC. Some pilots became QFIs following FLIT before beginning FPC, other QFIs completed an instructional tour before proceeding through FLIT and on to FPC. Since there are multiple routes to FPC, the common denominator is Ph II flying performance. Success rates on FPC ought to be investigated as a function of Ph II flying performance (and to a lesser extent, QFI experience). The expense of training pilots for fifth generation aircraft under FFCP requires a clear understanding of the standards evaluated during the common phase of training (pre-streaming) and this is informative of what selection methods are used when providing candidates to the training system. The dataset was narrowed to the period from 2012 to 2018 and it does not fully capture all those selected under the RAFAAT battery. It is informative from another perspective; chiefly the role of GPA in predicting future success in more complex flying environments.

Data was analysed on the performance of pilots selected to fly fighter aircraft to determine if a significant association exists between flying GPA prior to FPC and the frequency of success on FPC from 2012-2018. Additionally, all QFI were also evaluated on a pass/fail basis for FPC to determine if an association exists between QFI selection and likelihood of success on FPC. Failure at 419 Sqn was considered a failure at 4 Wing. GPA of candidates who successfully completed 419 Sqn were considered in the overall GPA assessment prior to FPC. To conduct this subtest, the following hypotheses were tested.

- Research hypothesis 1: There will be an association between flying GPA of pilots prior to FPC and frequency of success.
- Null hypothesis 1: There will be no association between the flying GPA prior to FPC and the frequency of success.
- Research hypothesis 2: There will be an association between candidates who previously qualified as QFI and frequency of successful completions of FPC, as compared to students who had no QFI background.
- Null hypothesis 2: There will be no association between candidates who previously qualified as QFI and frequency of successful completions of FPC, as compared to students who had no QFI background
- Assumption 1: Data cell counts will be above 5 in the contingency table
- Assumption 2: Two categorical variables
- Assumption 3: Independence of variables
- Assumption 4: Mutually exclusive
- Limitation 1: Many candidates completed a portion of training (T-6 or T-38) at ENJJPT and those grades are unavailable.
- Limitation 2: Many candidates who attended 419 Sqn and 410 Sqn between 2012 and 2018 competed syllabi which differs from the current status quo (e.g. Trial serials of FLIT, PH III Hawk vice Ph III Hvd/Ph IV(T) Hawk, Ph IIB Hvd vice Ph III Hvd).
- Limitation 3: Some candidates were originally selected as multi-engine or rotary-wing pilots from Ph IIA/Ph II Hvd, completed an operational tour on another airframe and later returned to 15 Wing as QFI. These candidates are not “pipeline

QFI" nor were they fighter-tracked candidates and therefore did not complete Wings standard on a single-pilot high-performance ITP.

- Limitation 4: Staff turnover at 15 Wing has generated varying GPA through cultural shift (e.g. potential tendency to over grade as discussed); leverage of a block program to evaluate learning curve, increased proportion of low-hour QFIs in comparison to historical levels and other confounding variables that limit the absolute objectivity of GPA as a standalone predictor.

- Limitation 5: The sample population of pipeline QFIs is too small to generate reliable statistics as a single group. This limitation will undermine the power of findings under hypothesis 2.

Results. An a priori power test was conducted to determine the minimum required sample size for a medium effect size (0.05), an alpha-level of 0.05 and a power of 0.8. The total required sample size is a minimum of *N* = 88 (with a critical Chi-square of 3.841). Our total sample size of 98 for hypothesis 1 exceeds these requirements and meets the assumptions. However, the sample size cannot be met for hypothesis 2 due to the low number of pipeline QFIs who have attended FLIT (pass or fail).

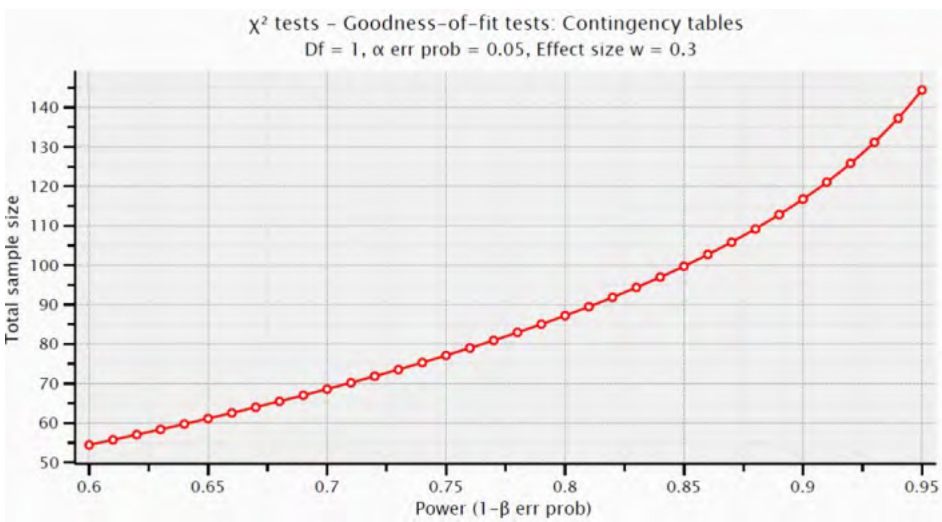

Figure 4. Required sample size as a function of power.

<u>Assessment of assumptions underlying bivariate correlation</u>

Normality

Table 1

Descriptive statistics for GPA of all pilots selected for training at 4 Wing up to and including 419 Sqn (N=98)

	N	Mean	Median	Mode	SD	Skewness	Kurtosis
Income	98	3.773	3.756	-	0.326	-0.181	0.081

Grades for pilots up to and including 419 Sqn but prior to FPC show small kurtosis and skew (inside the limits for parametric statistics). The median and mean are close, which is indicative of normally distributed data. Visually, the data are normally distributed and a Shapiro-Wilk test returns a value of 0.99 indicating normality. Quantile plot deviates from normal at the outliers.

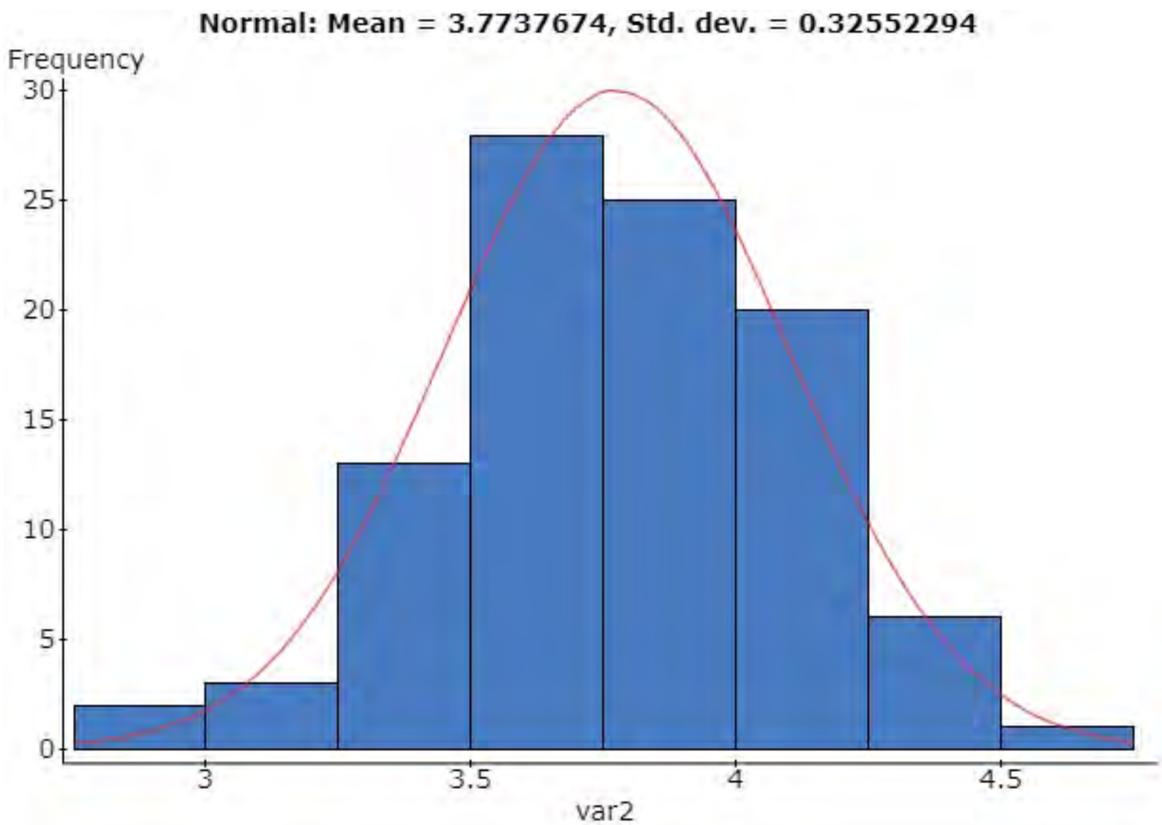

Figure 5. Histogram of GPA for all pilots up to and including 419 Sqn (including candidates who were unsuccessful at 419 Sqn) with normal curve overlaid.

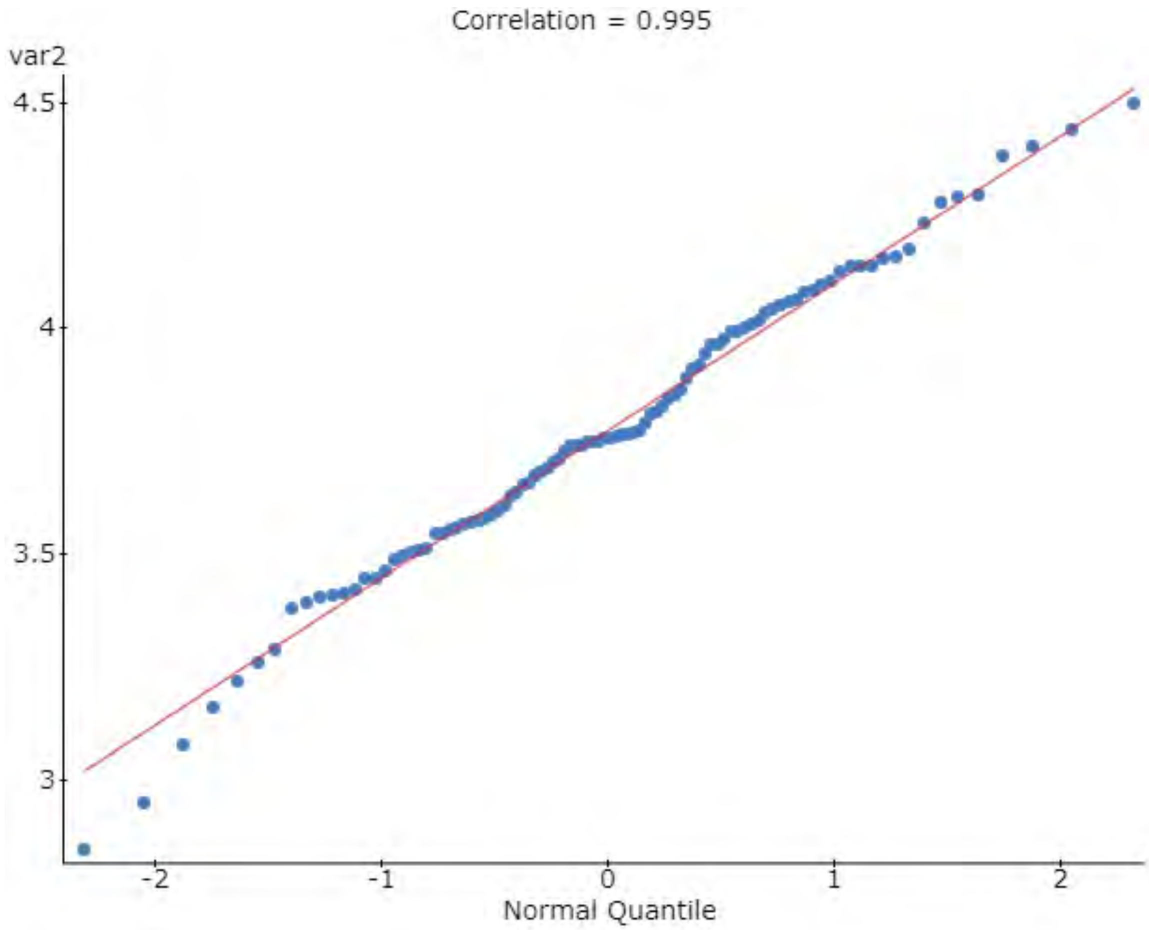

Figure 6. Quantile-quantile plot showing a simple linear regression line for Ph II GPA of the subgroup of single-pilot high-performance candidates up to and including 419 Sqn (including unsuccessful candidates at 419 Sqn).

The data show a strong positive linear regression (r = 0.995) and are normally distributed. Using the average GPA (*M* = 3.773) as a cut-off, the following nominal variables are defined:

Independent variable: GPA (above or below 3.773)

Dependent variable: Successful completion of FPC

Table 2

Contingency table of totals and expected values for success frequency on FPC among all pilots sent to 419 Sqn for training according to GPA (N = 98)

	Successful	Not successful	Row Total
GPA above 3.773	39 (32.6)	6 (12.4)	45
GPA below 3.773	32 (38.4)	21 (14.6)	53
Column Total	71	27	98

Note. Expected values in parentheses.

Table 3

Results of Chi-square test and descriptive statistics for success frequency on FPC among all pilots sent to 419 Sqn for training according to GPA

GPA	FPC completion	
	Successful	Not successful
Above 3.773	39 (39.8%)	6 (6.1%)
Below 3.773	32 (32.7%)	21 (21.4%)

Note. χ^2= 8.43, *df* = 1. Numbers in parentheses indicate total percentages.

Analysis produced a chi-square statistic of $\chi^2(1, N = 98) = 8.43$, $p = 0.0037$. The cell count in the contingency table meets the assumptions (minimum above 5), however, due to the low cell count in the unsuccessful above-average GPA category, a Fisher’s exact test was performed. The Fisher’s exact test returned a value of 0.0059 which is significant at the 0.05 alpha level. The null hypothesis can be rejected.

Table 4

Contingency table of totals and expected values for success frequency on FPC among all pilots sent to 419 Sqn for training according to whether or not the candidate was a QFI (N = 98)

	Successful	Not successful	Row Total
QFI	6 (8.0)	5 (3.0)	11
Not a QFI	65 (63.0)	22 (24.0)	87
Column Total	71	27	98

Note. Expected values and percentage of total in parentheses.

Table 5

Results of Chi-square test and descriptive statistics for success frequency on FPC among all pilots sent to 419 Sqn for training according to whether the candidate was a QFI

Type	FPC completion	
	Successful	Not successful
QFI	6 (6.1%)	5 (5.1%)
Not a QFI	65 (66.3%)	22 (22.5%)

Note. χ^2= 1.99, *df* = 1. Numbers in parentheses indicate total percentages.

Analysis produced a chi-square statistic of $\chi^2(1, N = 98) = 1.99$, $p = 0.1584$. The cell count in the contingency table does not meet the assumptions (minimum above 5). The Fisher's exact test returned a value of 0.1694 which is not significant at the 0.05 alpha level. The null hypothesis that there is no difference between the populations cannot be rejected.

Findings and Interpretations

Hypothesis 1. There is a significant association between GPA and successful FPC completion. Null hypothesis 1 can be rejected. Of the 27 candidates who were unsuccessful, 21 had a GPA below 3.773 (78%). It should be noted, however, that of the 71 successful candidates, 32 had a GPA less than 3.773 (45% of successful completions). Therefore, lowering the failure rate at 410 Sqn below 27.4% will likely reduce also the success frequency (since 45% off successful completions were candidates whose GPA was less than the 3.773 average). This suggests there may be other factors that could be considered in fighter-flow selection.

Hypothesis 2. There is no significant association at the 0.05 alpha level between the independent variable (QFI) and the dependent variable (success on FPC) with p = 0.1694 (Fisher's). The Chi-square value of $\chi^2(1, N = 98) = 1.99$ is well below the critical cut-off of 3.481, therefore the null hypothesis cannot be rejected. To reduce the chance of a type II error the sample size of QFI (N = 9) must be increased.

There is a caveat that there have been 12 QFI qualified candidates who have attended training at 4 Wing (419/410) from 2012-2018. Of these, there was one voluntary withdrawal. Of the remaining 11 candidates, several had been selected as multi-engine or rotary-wing candidates during Ph II and returned to 15 Wing after completing operational tours on other airframes. These are not pipeline QFI as they did not complete Wings standard as a single-pilot high-performance candidate. They have been included in the calculations nevertheless, due to the small sample size. Finally, 4 out of 5 of the QFI-qualified candidates who were unsuccessful had a GPA below 3.77.

Conclusion. There is a significant association between GPA and success on FPC, but no significant correlation between QFI status and likelihood of success. The latter conclusion is not statistically robust due to the small sample size. The average GPA of successful candidates on FPC is 3.77 with 78% of failures (N = 21) occurring among the population with less than 3.77 GPA. However, 45% of successful completions on FPC were candidates with a GPA lower than 3.77. Very few pipeline QFI have attended 410 Sqn. Most have laterally moved from another platform, or completed 419 Sqn as a student, before returning to 15 Wing for a QFI tour. The current paradigm of selecting pipeline QFI from Wings standard is relatively untested through to the conclusion of FPC. Data would indicate that candidates (student or QFI) who have obtained a minimum GPA of 3.77 have a low likelihood of failure at 410 Sqn.

Clearly, GPA is a predictor, though it has been more recently conflated with other factors and its fidelity as a sole determinant may be invalidated by a recent tendency to introduce systemic grading errors. A methodology could be employed to account for this, which may yield useful information about any shortcomings of our selection methods to act as an appropriate filter prior to Ph II. By looking at the GPA spread between the top and bottom thirds over time we may be able to determine if the current filter is too relaxed. It may also provide a correction factor to the average GPA, as the performance of the top third over time should be normally distributed about a mean that remains consistent with the syllabus. The bottom third, by contrast, will fluctuate according to the stringency of the selection mechanisms. A very selective system will see a narrow spread between the performance of the top and bottom thirds; while casting a wide net will capture a larger number of individuals whose performance will be drastically different from those at the top of their class.

TOWARDS A COMPREHENSIVE SOLUTION

The literature has isolated numerous aptitudes and personality attributes that are associated with pilot performance to varying degrees. Key takeaways relevant to our own case are summarized here. Psychomotor predictive validity depends on the testing mechanism involved and is increased when combined with other cognitive processes, particularly those measuring spatial ability and perceptual speed. Fluid intelligence (as a subcategory of general intelligence), perceptual speed and working memory all show strong predictive validity; and separating executive function tests to target WM aspects of inhibiting, updating and shifting are highly predictive, especially when combined with testing on emotional regulation. SA too may be tested by proxy using EF tests and is influenced by both LTM and WM. General intelligence (*g*) is a predictor, but the discriminatory factor is limited, and it should be combined with aviation knowledge, experience, and motivation (in addition to the aforementioned).

Mathematical ability repeatedly predicted pilot performance in multiple studies cross-culturally and is a predictor of level 3 SA, particularly when combined with PS. Verbal STS is also a factor supported by multiple independent studies. Personality assessments from both civilian and military sources, in multiple countries all showed that emotional stability and personal motivation (intrinsic motivation) are critically important and those high in conscientiousness, low in neuroticism, slightly higher in extraversion, and slightly lower in agreeableness fared best in pilot training. The 16PF personality test was insufficiently sensitive as a selection tool as compared with NEO-PI-R. Furthermore, an understanding of candidate sensitivity to criticism, depression, anxiety, and insecurity is necessary. MMPI2 testing demonstrated the utility of testing for dependability, stress tolerance, motivation, openness to experience, attitudes towards risk, and resilience. In many cases, previous flying experience itself was the most incrementally valid predictor.

Given the RAFAAT PB11 addresses many of the EF, WM, and spatial awareness requirements one must consider what exactly is missing from our selection that may be contributing to a reduction in performance during Ph II. It will be necessary to conduct a comprehensive analysis of the degree to which subtests truly test high-level WM tasks and an analysis of the extent to which combined tests of spatial ability and PS are leveraged.

From a psychomotor perspective, the method by which psychomotor ability is determined is too low-resolution in comparison with CAPSS. The USAF discovered a similar problem with its BAT, which uses a similar interface for evaluating psychomotor ability and for which the outcomes were comparable. The apparatus itself is a factor in the predictive validity of combined psychomotor and cognitive tests. Another missing element is a rigorous approach to personality-based selection criteria. Nearly all other aviation organizations and militaries around the world utilize comprehensive personality testing, yet the RCAF approach is limited in this regard. MMPI2 and NEO-PI-R (or the more recent NEO-PI3[96]) are possible candidates for

[96] The NEO-PI3 was developed in 2005 but little research has been conducted in the aviation context. It is worthy of future study.

employment within the existing selection system, as they are in use with our allies with some success. Darr (2009) completed a study on Non-Commissioned Members (NCMs) in the CAF to assess Trait Self Descriptive (TSD) method in comparison with the NEO and her results are worth considering as they apply to pilot selection. Regarding conscientiousness (specifically achievement-striving), she noted: "that high order individuals tend to plan, organize, manage time, and impose structure on one's environment which is essential to success during early phases of a job. On the other hand, high achievement striving individuals are diligent, hardworking, and internally committed towards goals that are self-set or assigned, traits that motivate on-going performance once the transition period has ended."[97] Building on historical studies involving CAF members, it is worth considering the approach taken by those with experience and success in the CAF in terms of personality screening.

CANSOFCOM

CANSOFCOM use Hogan testing[98] in conjunction with IQ testing and a 360 assessment during selection and training for operators.[99] 360 assessments have been noted to have some shortcomings based on how the information is gathered but still serve both a baseline selection tool and (perhaps more usefully) as a development tool. Prolonged exposure to operational psychologists continues over one year to improve or deselect candidates as required. This is not from a clinical perspective but rather to consider the institutional requirements, mental robustness, mental agility and more. Interestingly, over six years of data shows that cadre intuitive assessment of an individual during selection correlates approximately 50% to success on course (while other tests are in the region of 10% correlation). The psychologists employed by CANSOFCOM can articulate better what the cadre sees through intuition and their contribution continues far beyond selection and initial training. In fact, retesting occurs for leadership selection at the section and detachment commander level and there are few (if any) examples in recent memory of any PRB occurring without an operational psychologist on the board.[100]

The Instructional Environment

Reflections from the author's time in the training environment may offer anecdotal support for the empirical findings of studies described in this research. It is of the author's opinion that selection should be based on aptitude, and training should be based on requirements. There must be a curve that forces individuals to improve daily rather than allowing students to plateau. This will result in an attrition rate (which is not unwarranted and may be inescapable), but this attrition rate must be minimized through effective leveraging of appropriate selection methods. Imperfection is tolerated (humans are fallible, and mistakes will happen); therefore, it is important to create a culture that reinforces open sharing of ways to improve (rather than shaming errors, which will lead to people hiding their mistakes).

---

[97] Wendy Darr, 'The Trait Self Descriptive (TSD) Inventory: A Facet-Level Examination', Technical Memorandum (Director General Military Personnel Research & Analysis, February 2009), 11.
[98] A commercial product that assists with talent acquisition and development which is also used in many other sectors including US SOF communities.
[99] Curtis Chow, CANSOFCOM, Discussion on personality testing, Telephone, 27 March 2023.
[100] Curtis Chow, CANSOFCOM.

Likewise, individuals must be accountable for errors, humble in accepting feedback and proud of improving.

Conscientiousness in this regard is critical, and conscientiousness is an aptitude that can be tested during selection and further improved by training in those who are high in this trait. Those successful in training will have adopted a culture that underscores their place in the team and reinforces the required cultural mindset (striving for improvement). This must be reinforced constantly, or it will be eroded over time. Cultural adaptation should also prepare individuals for the sustained pressure of high expectations; both their own expectations and the external stressor of demonstrated performance. Human Performance in Military Aviation (HPMA) training and the Flight Safety system feed into a "just" reporting culture and the idea of sharing improvements.[101] Competitiveness against others is motivating from a physical sense, but not useful in the aviation context (it can become quite toxic). So, competing against oneself to improve is a better methodology. This requires constant feedback so that the individual can see their own improvement (and where they need to improve). This must be measured against a standard. An oft used colloquialism "drinking from the fire hose" refers to constantly feeling as though there is a requirement to learn a new skill before having mastered the old skill which forces individuals to push themselves to achieve the learning curve set by the instructors—not to relax into the comfort of setting their own training schedule. This is how to ensure people are studying, training, and preparing to the maximum extent possible while instructors are not present. It also highlights the fastest learners and the most conscientious.

Inability to properly measure one's own performance against peers or to receive the amount of praise they are hoping for can be upsetting for some people, and those are individuals may be more likely to voluntarily withdraw (VW). Of the 308 PRBs conducted over the sample period, 41 were due to VW or for disciplinary Officer Development deficiencies (OD).[102] Since NFTC distributes the total program costs of Ph II training over the annual graduates of the program, every withdrawal from the course increases the per person cost of a completion.[103] Even at a conservative estimate, tens of millions has been spent over the course of a decade solely on students that have quit (VW) or been removed from training for disciplinary reasons (OD).[104]

---

[101] CDS, 'HPMA Handbook' (1 Canadian Air Division Air Force Standards Advanced Training Centre, 2015).

[102] Raw data from NFTC available on request.

[103] CAE is contracted on behalf of the RCAF to provide the facilities, simulation, ground based training staff (GBTS), aircraft, and servicing for NFTC. The RCAF provides the flight instructors and other personnel.

[104] The common figure quoted for Ph II training is several hundred thousand dollars. The exact figure per student is not given by NFTC as the enterprise is akin to a flying university where most of the total annual costs for the year are fixed, then distributed over the student body. The cost of Ph I and of other mandatory aircrew training such as land and sea survival, aeromedical training and so on are not included. The model where total program costs of Ph II (Snow and Ice Clearing or SNIC, simulator instruction, ground school, aircraft maintenance etc) are distributed over the total students graduated each year means that at times where fewer students graduate, the cost per student is much higher.

Personal drive to succeed, or to fulfil aspirations, is a powerful motivator to sustain a high standard of effort over the lengthy training periods of daily assessment that are measured in years. It is possible to build physical resilience and teamwork though training, but intrinsic motivation appears closer to an aptitude and should be screened for during selection. Anxiousness often indicates poor psychological resilience; if not resolved through confidence-building it tends to erode performance, often to critical or unsafe levels which creates a death spiral of regressive performance. The deliberate addition of responsibility with every training increment can foster mental resilience. Students become responsible for the brief, prepping the aircraft, preparing the data-cards, maps, flight plan and for making decisions airborne in dynamic and high consequence environments. These small steps equate to large gains over time, and each step is graded, and feedback is delivered by debrief (always with a risk of failure). Over time the individual is building mental resilience and reinforcing the culture of striving to improve yesterday's performance. In addition to conscientiousness and resilience, humour, humility, exercising healthy decompression activities regularly, and healthy relationships with friends and family for support are enablers of success.

Lessons from civilian aviation CRM programs are informative. The European Union Aviation Safety Agency (EASA) knowledge requirements encompass non-technical areas, which have emerged due to increasing complexity and automation in the cockpit including SA; problem-solving and DM; workload management and task sharing; leadership and teamwork; communication and assertiveness; automation complacency; stress management; fatigue and vigilance; monitoring and intervention; resilience development; surprise and startle effects; and operator and company culture.[105] These areas have been identified based on empirical data on human factors. CRM training programs have recognized resilience as a key component of successful outcomes in dynamic environments. Resilience at the individual, crew, and organizational levels involves the ability to sustain effective operations despite unexpected conditions. Developing individual resilience in terms of mental flexibility and performance adaptation, as well as organizational resilience to maintain operations in a dynamic environment, is a relatively new concept that has been incorporated into CRM training curriculums.[106]

[105] Flin, R., 'CRM (Nontechnical) Skills', in *Crew Resource Management*, 3rd ed. (Amsterdam: Academic Press, 2019), 185–226.
[106] Flin, R.

POLICY

A common assumption regarding the contemporary Government of Canada policy is that equal representation is the highest order priority from which others follow or are subsumed. In the Canadian Prime Minister's (PM) Mandate Letter to the Minister of National Defence (MND), it is clearly stated that "as Minister of National Defence, your immediate priority is to take concrete steps to build an inclusive and diverse Defence Team."[107] This is not in opposition to the suggested rigour of an improved pilot selection system and should not be viewed as zero sum. The implementation of a more rigorous approach to pilot selection, incorporating the additional measures discussed herein, is not in conflict with the steps taken to ensure a more diverse workplace. Simply put, the aptitudes required to safely operate in the military aviation environment are distributed on an individual basis throughout our population as the previous studies on cognitive psychology have shown. Selecting on the basis of aptitude and personality is a mechanism for ensuring candidates are appropriately screened without the influence of mercurial social and cultural mores. The aptitudes and personalities that have been identified in this paper are derived from a considerable body of knowledge and their contribution to safety and effectiveness has been thoroughly explained. It is worth noting other policy statements aligned with the suggestions made by the author.

In 2021, the Canadian PM specified a commitment to "ensure the CAF is a 21st century military with the capabilities, equipment and culture to implement Canada's Defence Policy, *Strong, Secure, Engaged*, and anticipate and respond to the full range of current and emerging threats."[108] Executing this vision in the air domain requires competent pilots. Furthermore, our new CAF Ethos, *Trusted To Serve*, says of stewardship that all personnel must ensure the effective and efficient use of public money, property and resources.[109] This means we cannot ignore the high cost of failure or the burden on the training system of those candidates who ought to have been screened out during the selection process. While our doctrine *Strong, Secure and Engaged* states "we are fully committed to implementing our new comprehensive Diversity Strategy and Action Plan, which will promote an institution-wide culture that embraces diversity and inclusion . . . diversity will enhance military operational effectiveness by drawing on all of the strengths of Canada's population."[110] Certainly, this will include the strengths of all those who meet the aptitude and personality requirements necessary for successful pilot selection. Within the same document, the RCAF perspective is clearly stated: "the effectiveness of the Royal Canadian Air Force requires continued investment in professional development and education programs focused on the theory and practical application of aerospace power, training programs and systems of the highest calibre, and

[107] Rt. Hon. Justin Trudeau, 'Minister of National Defence Mandate Letter' (Government of Canada, 16 December 2021), https://pm.gc.ca/en/mandate-letters/2021/12/16/minister-national-defence-mandate-letter.
[108] Rt. Hon. Justin Trudeau.
[109] Canadian Defence Academy - Professional Concepts and Leader Development, 'Canadian Armed Forces Ethos: Trusted To Serve', 2022.
[110] Canada, 'Strong, Secure, Engaged: Canada's Defence Policy' (Ottawa: Department of National Defence, 2017), 22, https://www.canada.ca/content/dam/dnd-mdn/documents/reports/2018/strong-secure-engaged/canada-defence-policy-report.pdf.

an institutional culture placing the highest value on the maintenance of air safety and airworthiness standards."[111]

Government of Canada policy values both stewardship of resources and the highest standards of air safety. It is also true that representation and diversity are overarching themes in political discourse, but it should be clear that all these policy requirements are satisfied by a more robust selection system independent of identity and grounded in decades of advanced research in multiple countries that can more accurately isolate clusters of attributes that are enablers of individual success. The reason this bears mention at all, is the prevalence of misunderstanding around the utilization of technical solutions at a time when institutional reformation is in vogue. There may be a misconception that a credibility crisis affecting recruiting and retention is best addressed by demonstrating a willingness to cast a wider net in the hopes of capturing candidates who once may have been excluded in order to compensate for the reduced numbers of interested applicants.[112] It would be unwise to consider this a form of fairness, or a suitable solution to any problem with a safety nexus.

The RCAF is graduating fewer and fewer competent pilots every year, and this has both safety and stewardship components. Adopting industry best practice for pilot selection to provide the training system with the most capable candidates appears to be in line with both policy on safety and on stewardship of resources. The author refrained from discussing equity policy in the pursuit of objective data on actual student performance over time to ensure the paper remained agnostic of politics. A possible critique of this approach might suggest aviation procedures and engineering implicitly advantage those individuals that have a similar cultural background to those who designed the aircraft and aviation procedures themselves. Anecdotally, the author flew with foreign pilots for whom English was a second language and who had flown Scandinavian Gripens and Russian built Mig-29s prior to training on the British Aerospace Hawk Mk 115. The author flew with Royal Saudi Air Force and Royal Singaporean Air Force pilots whose customs, cultures and languages are very different from those of the western engineers that designed the aircraft they each flew. Fundamentally, the aptitudes required in the cockpit to assess dynamic aspect, closure, and range; or to assess turn-circle entry in three dimensions are independent of language and custom. The dial and gauge clusters of Russian-built aircraft may differ from those of western aircraft, but the aptitudes required of the pilots who fly those aircraft are universal.[113]

A growing discipline within aviation studies is that of *human factors.* Display design and its effect on human perception, our various physiological limitations, human cognition and more fall within this field of study which has provided numerous improvements to data representation in the cockpit, ergonomic factors and the

---

[111] Canada, 39.
[112] Department of National Defence, 'Canadian Armed Forces Retention Strategy', 2022.
[113] Though these observations are anecdotal, there is ample evidence of highly competent pilots successful on one type of airframe transitioning to airframes designed by other manufacturers, countries, cultures and from different periods. In fact, test pilots of all backgrounds do exactly this.

identification of deficiencies in cockpit design that negatively affect pilot performance.[114] Use of colour, active warning systems, passive system monitoring, rate of change of displayed information, and countless other visual and aural perception factors have been widely studied and incorporated into the design of aircraft to improve safety; and this process is ongoing. Expert guidance on the requirements of humans in the socio-technical aviation environment has always been grounded in the pursuit of safety and effectiveness and we ignore expert testimony at our peril. Recent advances in the use of brain imaging equipment in medical and cognitive psychology fields have produced a growing body of literature on brain differences between males and females and, predictably, these differences can manifest in the cockpit environment. Unfortunately, this also creates a potential problem: accounting for sex differences in selection and training while simultaneously upholding equity policies.

Forgues (2014) identified in her thesis several studies demonstrating differences in performance between the sexes, notably that males outperformed females on all measures of psychomotor performance on both the AFOQT and BAT for the USAF[115] and males were more likely to pass the legacy CAPSS selection than females. A Chi-square analysis of the gender effects of CAPSS testing was significant: "71.5% of male candidates were in Class 1 (high CAPSS scores), whereas only 37.5% of females were. The opposite pattern was shown in Class 2 (low scores)."[116] Forgues conjectured reasons for this, citing prevailing literature on social factors and brain differences. The availability of data on brain connectome differences using biomedical image analysis is consistent on the male advantages in spatial awareness and psychomotor ability and female advantages in memory and social cognition.[117] Differences in cerebellar connections (pertinent to the previous discussion on EF), on grey and white matter, and the way these differences manifest in terms of performance have been demonstrated unequivocally using modern neuroimaging technology.[118] This is problematic when choosing aptitude testing systems that must not disadvantage females in comparison to their male peers. The replacement of CAPSS with the RAFAAT (which was far more equitable by comparison) presumably leverages advantages in female memory that may offset male advantages in spatial awareness and psychomotor ability.[119] It is worth noting that the subtle differences in brain connectome are normally distributed among the test populations, and high achievers in both sexes will demonstrate abilities superior to the population average.

There are countless examples of talented aviators of all cultures, both female and male, underscoring the fact that capable and motivated individuals will meet arduous standards. However, subtle differences between the sexes mean that in the upper percentiles of high-performance human pursuits where small differences manifest in ways

---

[114] Monica Martinussen And David R Hunter, *Aviation Psychology and Human Factors, Second Edition*, Book, Whole (CRC Press, 2017); Dan Maurino Eduardo Salas, ed., *Human Factors in Aviation*, Second (UK: Academic Press, 2010).
[115] Forgues, 'Aptitude Testing of Military Pilot Candidates', 24.
[116] Forgues, 54.
[117] Madhura Ingalhalikar et al., 'Sex Differences in the Structural Connectome of the Human Brain', *PNAS* 111, no. 2 (14 January 2012): 823–28.
[118] Madhura Ingalhalikar et al.
[119] Forgues, 'Aptitude Testing of Military Pilot Candidates'.

that aren't evident day-to-day, there is unlikely to be a perfectly representative outcome. Therefore, misguided pursuit of parity in performance through dogmatic application of equity measures is unscientific. If the science supports the implementation of a selection system that excludes those not competent to perform in the required operational environment, and policy overrides prevailing wisdom, it does so out of an ideological motivation not a scientific one. This is not to say that it will not happen, but merely to say that we can measure the consequences of ignoring required aptitudes, we understand the cost of doing so, and we can predict the likely outcome.

If any further rationale is required for placing aptitude above all else in the aviation domain, it is the fact that pilots will operate internationally, in both civil and military airspace in times of peace and war. They will be subject to the laws and procedures of host nations and of international organizations that supersede policy statements of the sitting government.[120] The indifferent nature of the flying environment includes random emergencies, system failures, bad weather and enemy action which also gets a vote. Thriving in this arena over the course of a career requires high standards, as any successful female or male career pilot in any country will attest. In fact, the consequences of misinterpreted policy will be felt long after the policy itself has changed, and nothing will hurt the credibility of the CAF and Canada writ large more than a catastrophic accident on foreign soil which was predicted from selection, through training, to the operational environment over a period of years, but for which no interventions were enacted. Fortunately, our policy documents are clear on the requirement to exercise stewardship in training and the primacy of air safety. There can be no confusion that selection systems which provide only the most capable candidates for training, should satisfy our nation's policy objectives.

CONCLUSION

Our selection and training systems should not be stove piped separately from one another. Feedback from the training system, must inform our selection methods. There is a strong but unpopular case that can be made for increasing the stringency of selection requirements to provide fewer (but more capable) candidates to our training system. The ability of our training system to absorb student pilots is more limited now than any time over the last decade and its finite resources are easily consumed by the PRB and grievance processes to the detriment of pilot output. A stated observation of the 2014 comparative study of the RAFAAT and CAPSS was that candidates who would have been successful under the criteria for the RAFAAT were screened out by CAPSS[121] and it is left to conjecture whether this may have featured in its replacement. But the data from the years that followed shows that preferential adoption of less stringent selection methods not only fails to address the pilot shortage by quantity, but there is also a deleterious effect on pilot output by quality. This is not to suggest that our current selection is completely flawed, but it raises a valid argument that it ought to be improved.

---

[120] Air Traffic Controllers in dense European airspace are ambivalent to Canadian cultural trends and should care only that our pilots are safe and effective while operating in their airspace.

[121] Forgues, 'Aptitude Testing of Military Pilot Candidates'.

The RCAF can begin by thoroughly assessing the EF, WM, and SA concerns raised by Forgues[122] and throughout this research to ensure that the additional subtests incorporated in the PB11 actually address the deficiencies noted in her research and those noted in the preceding sections. Specifically in terms of high-level WM tasks and combined measures for spatial ability and PS, and for psychomotor and EF tasks. The information-dense environment of modern cockpits warrants the inclusion of thorough EF screening during aptitude testing of updating, shifting and inhibiting attributes under varying workload. However, arguments that modern automation has decreased the importance of psychomotor ability are unfounded. The unanimity of historic and contemporary research on the predictive validity of psychomotor ability when measured alongside PS and spatial awareness tasks should not be disregarded. Moreover, the apparatus used in the assessment of psychomotor ability has a significant effect on its utility as a discriminator. Based on data from studies on the USAF BAT and the RAFAAT a strong case can be made that these tests do not adequately measure psychomotor ability as compared with CAPSS and the DOT which combine spatial ability and PS more effectively with psychomotor ability and have far higher predictive validities for this aptitude.

Educational background, aviation interest and experience, general cognitive intelligence, verbal reasoning, and mathematical ability are all positively correlated with pilot training success. Verbal reasoning and mathematical ability are tested as part of the CFAT, and the other elements emerge through the interview process and elsewhere during RCAF pilot selection, including CFASC. What is deficient, however, are comprehensive and repeatable techniques for personality assessment against an objective standard. These could include the utilization of MMPI2, NEO-PI-R (or NEO-PI3), or Hogan's test during selection with particular attention paid to the various personality facets mentioned earlier, particularly regarding conscientiousness and resilience. Throughout this research, the most glaring area for improvement has been effective personality screening. The author's own experience, having served in more than one air force and having successfully completed four selection programs (RAF, RAAF, RNZAF, RCAF) is that the RCAF is in a distant last place in terms of personality assessment.

The contribution of cognitive psychology is growing in many domains from professional sports to military special forces units. The attributes long understood to positively correlate with superior performance in specific domains can be measured increasingly accurately with access to ever evolving datasets. The RCAF ought to adopt additional screening mechanisms from this field of study and rigorously apply them to aircrew selection.

The centrality of flying performance measured by GPA to student pilot success rates during and after Ph II pilot training underscores another course of action for consideration. Using only the current model for selection, with no further addition of personality screening mechanisms or a modification of our psychomotor screening apparatus, the required performance for successful pilot selection (prior to Ph II) ought to be raised. Airline studies discussed in this research showed a six-fold

[122] Forgues.

increase in likelihood of encountering training problems for those in the bottom 20% of their classes; nearly 80% of all failures at FPC over a six year period were those who scored below the mean Ph II GPA of their stream cohort; and the resources consumed by struggling pilots are at a level now that severely restricts the training system capacity to generate pilots at the rate required. Selecting fewer candidates per year may not meet recruiting metrics, but the preference towards quantity is creating a much larger and longer-term issue by forcing ill-suited candidates on an already strained training system. Recommendations to improve our pilot selection system are as follows:

- Implement rigorous personality testing for baseline selection and include ongoing assessment for deselection over a probationary period. Areas of focus ought to include (at a minimum) emotional stability and self-regulation; conscientiousness; resilience; attitudes towards risk; extraversion; agreeableness; neuroticism; anxiousness; stress tolerance; depression; acceptance of criticism; humility; team behaviours; humour.
- Retain verbal and mathematical ability testing, along with objective assessment of *g*F, *g*C, previous aviation knowledge and experience, and educational background.
- Modify the mechanism by which psychomotor ability is tested from the current alphanumeric keyboard and joystick design offered by the RAFAAT. This assessment should involve spatial awareness and PS components. A suggestion that will satisfy the requirement for increased emphasis on key EF processes in an information-dense environment is to modify the CAPSS model to incorporate a modern glass cockpit where the presentation of data involves inhibiting, updating and shifting EFs under varying WM and PS loads.
- During the development cycle of the enhanced psychomotor and WM testing suggested above, raise the minimum requirement for success during pilot selection under the current paradigm.

AREAS FOR FUTURE STUDY

To justify a modification of our psychomotor testing apparatus, testing could be conducted to further understand the role of EF in the development and maintenance of SA—particularly in the modern context. A sample of randomly selected operational pilots could undergo aptitude testing to measure their inhibiting, updating, and shifting abilities using established methods like the flanker n-back and Stroop tests. Perceptual speed would be evaluated through a search test that increases visual perceptual loads with interleaved information. The population would then be divided into two groups based on their combined mean EF and PS scores: high and low. Both EF groups would participate in an exercise involving a dynamic system that can measure level 1, 2, and 3 situational awareness (SA) in a simulated cockpit environment. A secondary task that varies in perceptual load would also be applied, and it is hypothesized that as perceptual load increases, SA would break down (single-tailed). The same procedure would be conducted for the PS groups tested using SAGAT. The performance of each group in terms of their ability to maintain SA under increasing perceptual load demands would be compared using the Chi-square test with the categories being high/low EF scores and high/low perceptual load resilience. If the groups are significantly different at the 0.05 level, it could be concluded that individuals with a higher aptitude for EF are capable of handling higher perceptual loads while maintaining SA. The same comparison would be conducted for high/low perceptual speed versus high/low perceptual load resilience. When refined, this method could provide a low-cost proxy for predicting individual capacity to maintain SA in the modern cockpit while subject to psychomotor and other cognitive demands. A slightly more complex model could also introduce an ER component to simultaneously assess candidates' self-regulation.

Personality screening by analogue assessment is another area that warrants research. The Japanese Space Agency (JAXA) utilises an analogue for long duration missions whereby candidates are placed in a confined environment with one another over a period of days during which every social interaction is assessed.[123] It is useful for screening out individuals who have succeeded on written psychological assessments but who still may be unsuited to that environment. While this may seem obtrusive, complex, and expensive for pilot selection there are low-cost methods that could suffice. During pilot selection in the RNZAF two decades ago, the author's own experience involved barrack accommodation of all candidates in one communal room, and teamwork and leadership assessments throughout multiple days under the supervision of a psychologist. This was in addition to written batteries of psychological tests alongside mathematical and other aptitude tests, including psychomotor and divided attention assessments. Incorporating lessons learned from the CANSOFCOM model is another option, whereby candidates who are successful during initial assessment could remain under probation with subsequent psychological testing at intervals over a prescribed period deselecting those who invalidate the results of their initial test over time. Verification and comparison of models such as the NEO-PI3 and Hogan tests ought to be a high priority for inclusion in the pilot selection process.

[123] Te Koeti, T., 'Situational Awareness in Astronaut Selection: Assessment of Methods. ASCI 691 Graduate Capstone Project.' (Embry-Riddle Aeronautical University, 2020).

BIBLIOGRAPHY

Adams, M.J., Tenney, Y.J., & R.W. Pew. 'Situation awareness and the cognitive management of complex systems'. *Human Factors* 37, no. 1 (1995): 85-104.

Adriansen, A., Patriarca, R., Smoker, A., & J. Bergstrom. 'A socio-technical analysis of functional properties in a joint cognitive system: a case study in an aircraft cockpit'. *Ergonomics* 62, no. 12 (2019): 1598-1616. doi:10.1080/00140139.2019.1661527

Aiken, Edwin W., and Robert K. Merrill. 'Results of a Simulator Investigation of Control System and Display Variations for an Attack Helicopter Mission'. Moffett Field CA: Army Research And Technology Labs Aeromechanics Lab, 1980.

Arguello, J., & B. Choi. 'The effects of working memory, perceptual speed, and inhibition in aggregated search'. *ACM Trans. Inf. Syst* 37, no. 3 (2019): 1-34. doi:10.1145/3322128

Bartram, D. 'The predictive validity of the EPI and 16PF for military flying training'. *Journal of Occupational and Organizational Psychology* 68, no. 3 (1995).

Biggerstaff, Sean. Factor Analysis of the U. S. Navy's Aviation Interest Subtest. Pensacola FL: US Navy Advancement Center, 1998.

Boyd, J.E., Patterson, J.C., & B.T. Thompson. 'Psychological test profiles of USAF pilots before training vs. type aircraft flown'. *Aviation, Space, and Environmental Medicine* 76, no. 5 (2005): 463-468.

Bradley Purvis. 'An Evaluation of B1B Pilot Performance During Simulated Instrument Approaches With and Without Status Information'. Final Report AL-TR-1992-0088. Air Force Materiel Command Wright Patterson Air Force Base, Ohio: Armstrong Laboratory, 1992.

Brannigan, Michael. 'America"s First Cosmonauts: Reflections on the Human Cost of Shuttle-Mir'. *Astropolitics* 15, no. 1 (2017): 27–50.

Brooks, Mary R., Sean M. Puckett, David A. Hensher, and Adrian Sammons. 'Understanding Mode Choice Decisions: A Study of Australian Freight Shippers'. *Maritime Economics & Logistics* 14, no. 3 (2012): 274–99.

Burke, E.G. 'A short note on the persistence of retest effects on aptitude scores'. *Journal of Occupational and Organizational Psychology* 70 (1997): 295-301.

Butcher, J.N. 'Personality assessment of airline pilot applicants'. In Bor, R. et al. (Eds.), *Pilot mental health assessment and support*. Routledge, 2016.

Caldwell, John A., Qiwen Mu, Jennifer K. Smith, Alexander Mishory, J. Lynn, Gordon Peters, David L. Brown, and Mark S. George. 'Are Individual Differences in Fatigue Vulnerability Related to Baseline Differences in Cortical Activation?' *Behavioral Neuroscience* 119, no. 3 (2005): 694–707.

Canada. Strong, Secure, Engaged: Canada's Defence Policy. Ottawa: Department of National Defence., 2017.

Canada. Department of National Defence. ‘Royal Canadian Air Force Strategy’, 2023. https://www.canada.ca/content/dam/rcaf-arc/documents/reports-publications/royal-canadian-air-force-strategy.pdf.

Canada. Department of National Defence. ‘Canadian Forces Joint Publication 1-0: Military Personnel Management Doctrine’. Ottawa: Department of National Defence, 2008.

Canadian Defence Academy - Professional Concepts and Leader Development, ‘Canadian Armed Forces Ethos: Trusted To Serve’, 2022.

Carretta, T.R., & M.J. Ree. ‘Pilot candidate selection method (PCSM): What makes it work?’ TX: Armstrong Laboratory, 1993.

Carton, Andrew M. ‘“I’m Not Mopping the Floors, I’m Putting a Man on the Moon”: How NASA Leaders Enhanced the Meaningfulness of Work by Changing the Meaning of Work’. *Administrative Science Quarterly* 63, no. 2 (2001): 323–69.

Castellan, J. N., ed. *Individual and Group Decision Making: Cultural Issues*. New Jersey: Lawrence Erlbaum Associates, Inc, 1993

CDS. ‘HPMA Handbook’. 1 Canadian Air Division Air Force Standards Advanced Training Centre, 2015.

Cheung, Bob, Oshin Vartanian, Kevin Hofer, and Fethi Bouak. ‘General Recommendations on Fatigue Risk Management for the Canadian Forces’, 2010.

Coram, Robert. Boyd: The Fighter Pilot Who Changed the Art of War. Book, Whole. Boston: Little, Brown, 2002.

Cormier, D.C., Bulut, O., McGrew, K.S., & D. Singh. ‘Exploring the relations between Cattel-Horn-Carroll (CHC) cognitive abilities and mathematics achievement’. *Applied Cognitive Psychology* 31 (2017): 530-538. doi:10.1002/acp.3350

Crozier, M.S., Ting, H.Y., Boone, D.C., O’Regan, N.B., Bandrauk, N., Furey, A. … & M.P. Hogan. ‘Use of human patient simulation and validation of the team situation awareness global assessment technique (TSAGAT): A multidisciplinary team assessment tool in trauma education’. *Journal of Surgical Education* 72, no. 1 (2015): 156-163.

Curtis Chow, CANSOFCOM. Discussion on personality testing. Telephone, 27 March 2023.

Department of National Defence, ‘Canadian Armed Forces Retention Strategy’, 2022.

Doherty, Patrick J. ‘Electronic Checklists on Multi-Purpose Displays: A Better Way For Fighter Pilots to Manage Information and Situational Awareness during Periods of High Workload’. 2001.

Du, X., Zhang, Y., Tian, Y., Huang, W., Wu, B., J. Zhang. ‘The influence of spatial ability and experience on performance during spaceship rendezvous and docking’. *Frontiers in Psychology* 6 (2015): 1-12. doi:10.3389/fpsych.2015.00955

Eastman, R. F., and R. L. McMullen. ‘The Current Predictive Validity of the Flight Aptitude Selection Test’. Alexandria VA: Army Research Inst For The Behavioral And Social Sciences, 1978.

Endsley, M. ‘Design and evaluation for situational awareness enhancement’. *Proceedings of the Human Factors Society 32nd Annual Meeting*. 97–101, 1988.

Endsley, M.R. ‘Situation awareness global assessment technique (SAGAT)’. *Proceedings of the IEEE National Aerospace Electronic Conference* 3 (1988): 789–795.

Endsley, M.R. ‘Toward a theory of situation awareness in dynamic systems’. *Human Factors* 37, no. 1 (1995): 32–64.

Endsley, M.R. & C.A. Bolstad. ‘Individual differences in pilot situation awareness’. *International Journal of Aviation Psychology* 4, no. 3 (1994): 241-264.

Endsley, M., Sollenberger, R.L., Nakata, A., & E.S. Stein. ‘Situational awareness in air traffic control: enhanced displays for advanced operations’. DOT/FAA/CT-TN00/01. Atlantic City, NJ: FAA, 2000.

Endsley, M.R. ‘Situation awareness misconceptions and misunderstandings’. *Journal of Cognitive Engineering and Decision Making* 9, no. 1 (2015): 4-32. doi:10.1177/1555343415572631

Falconer, Boyd. ‘Human Factors and the “can-Do” Attitude: Strength or Weakness?’ *Australian Defence Force Journal*, no. 169 (2005): 23–32.

Falleti, Marina G., Paul Maruff, Alexander Collie, and David G. Darby. n.d. ‘Qualitative Similarities in Cognitive Impairment Associated with 24 h of Sustained Wakefulness and a Blood Alcohol Concentration of 0.05’. *Journal of Sleep Research* 12, no. 4: 265–74.

Farrow, D.R. ‘A Regulatory Perspective’. In Crew Resource Management, 3rd ed., 465–87. Amsterdam: Academic Press, 2019.

Ferrier, D.E., Bassett, H.H., & S.A. Denham. ‘Relations between executive function and emotionality in preschoolers: Exploring a transitive cognition-emotion linkage’. *Frontiers in Psychology* 5, no. 487 (2014): 1-12. doi:10.3389/fpsyg.2014.00487

Fisk, J.E., & P. Warr. ‘Age and working memory: The role of perceptual speed, the central executive, and the phonological loop’. *Psychology and Aging* 11, no. 2 (1996): 316-323.

Flight Instructor’s Handbook. A-PD-050-001/PF-001. Central Flying School, 2005.

Flin, R. ‘CRM (Nontechnical) Skills’. In Crew Resource Management, 3rd ed., 185–226. Amsterdam: Academic Press, 2019.

Flouris, Triant. ‘Integrated Airline Organizational Frameworks and Crew Resource Management Effectiveness’. Journal of Aviation/Aerospace Education and Research 9, no. 3 (2000).

Ford, Jane, Robert Henderson, and David O’Hare. ‘The Effects of Crew Resource Management (CRM) Training on Flight Attendants’ Safety Attitudes’. *Journal of Safety Research* 48 (2014): 49–56.

Forgues, S.L. *Aptitude testing of military pilot candidates* [Unpublished master's thesis]. Queen's University, 2014.

Friso-van den Bos, I., van der Ven, S.H.G., Kroesbergen, E.H., & J.E.H. van Luit. 'Working memory and mathematics in primary school children: A meta-analysis'. *Educational Research Review* 10 (2013): 29-44. doi:10.1016/j.edurev.2013.05.003

Gibb, Randy and Rob Gray. Ashgate Studies in Human Factors for Flight Operations : Aviation Visual Perception : Research, Misperception and Mishaps. Farnham, GB: Routledge, 2010.

Gillespie, Brigid M., Karleen Gwinner, Nicole Fairweather, and Wendy Chaboyer. 'Building Shared Situational Awareness in Surgery through Distributed Dialog'. *Journal of Multidisciplinary Healthcare* 6 (2013): 109–18.

Goldstein, B. Cognitive psychology connecting mind, research, and everyday experience. Fifth. Cengage, 2019.

Haas, M. W., L. M. Hirshfield, P. V. Ponangi, P. Kidambi, D. Rao, N. Edala, E. Armbrust, M. Fendley, and S. Narayanan. 'Decision-Making and Emotions in the Contested Information Environment'. *EAI Endorsed Transactions on Security and Safety* 1, no. 2 (2013): 1–6. https://doi.org/10.4108/trans.sesa.01-06.2013.e3.

Harris, Don. 'Airline Safety Management'. In Human Performance on the Flight Deck, 273–96. Ashgate Publishing, 2011.

Hecht, Heinrich. 'The World's First Turbojet Fighter – Messerschmitt Me 262'. Schiffer, 1990.

Helmreich, R.L., Merritt, A.C., & J.A. Wilhelm. 'The evolution of crew resource management training in commercial aviation'. *The International Journal of Aviation Psychology* 9, no. 1 (1999): 19-32. doi: 10.1207/s15327108ijap0901_2

Helmreich, R.L., Butler, R.A., Taggart, W.R., & J.A. Wilhelm. 'Behavioral markers in accidents and incidents'. *NASA Technical Report 95-1*, 1995.

Hettinger, Lawrence J., Bart J. Brickman, Merry M. Roe, and W. Todd Nelson. 'Effects of Virtually-Augmented Fighter Cockpit Displays on Pilot Performance, Workload, and Situation Awareness'. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 40, no. 2 (1996): 30–33.

Hoffman, C. & A. Hoffman. 'The role of assessment in pilot selection'. In Bor, R. et al. (Eds.), *Pilot mental health assessment and support*. Routledge, 2016

Holbrook, J. B., M. J. Stewart, B. E. Smith, L. J. Prinzel, B. L. Matthews, I. Avrekh, C. T. Cardoza, O. C. Ammann, V. Adduru, and C. H. Null. 'Human Performance Contributions to Safety in Commercial Aviation'. *NASA*. Farnham: CRC Press, 2019.

Houghton, Robert J., Chris Baber, Malcolm Cowton, Guy H. Walker, and Neville A. Stanton. 'WESTT (Workload, Error, Situational Awareness, Time and Teamwork): An Analytical Prototyping System for Command and Control'. *Cognition, Technology & Work* 10, no. 3 (2008): 199–207.

Hubbard, Todd, Ray King, Carina Eriksen, and Robert Bor. Pilot Selection: Psychological Principles and Practice. Book, Whole. CRC Press, 2019. https://doi.org/10.4324/9780429492105.

Hunter, Monica Martinussen And David R. Aviation Psychology and Human Factors, Second Edition. Book, Whole. CRC Press, 2017.

Inoue, N., & S. Tachibana. 'An isolation and confinement facility for the selection of astronaut candidates'. *Aviation, Space, and Experimental Medicine* 84, no. 8 (2013): 867-871.

J, Oksa, Hämäläinen O, Rissanen S, Salminen M, and Kuronen P. 'Muscle Fatigue Caused by Repeated Aerial Combat Maneuvering Exercises.' *Aviation, Space, and Environmental Medicine* 70, no. 6 (1999): 556–60.

Janic, M. 'High-Speed Rail and Air Passenger Transport: A Comparison of the Operational Environmental Performance'. *Proceedings of the Institution of Mechanical Engineers* 217, no. 4 (2003): 259–69.

Jewsbury, P.A., Bowden, S.C., & M.E. Strauss. 'Integrating the switching, inhibition, and updating model of executive function with the Cattel-Horn-Carroll model'. *Journal of Experimental Psychology: General* 145, no. 2 (2016): 220-245.

Ji, Ming, Can Yang, Ying Li, Quan Xu, and Rong He. 'The Influence of Trait Mindfulness on Incident Involvement among Chinese Airline Pilots: The Role of Risk Perception and Flight Experience'. *Journal of Safety Research* 66 (2018): 161–68.

Johnson, J.F., Barron, L.G., Carretta, T.R., & M.R. Rose. 'Predictive Validity of Spatial Ability and Perceptual Speed Tests for Aviator Training'. *International Journal of Aerospace Psychology* 27, no. 3–4 (2017): 109–20. https://doi.org/10.1080/24721840.2018.1442222.

Kanki, B.G., Anca, J., & T.R. Chidester. *Crew resource management. Third*. Amsterdam: Academic Press, 2019.

Keiser, H.N., Moclaire, C.M., King, K.M., Brown, N.L., Foroughi, C.K., Sibley, C., & J.T. Coyne. 'Updating the direction orientation task: An aviation selection tool'. *Proceedings of the Human Factors and Ergonomics Society 2019 Annual Meeting*, 2019.

Kharoufah, H., Murray, J., Baxter, G., & G. Wild. 'A review of human factors causations in commercial air transport accidents and incidents: From 2000-2016'. *Progress in Aerospace Sciences* 99 (2018): 1-13.

King, R.E., Carretta, T.R., Retzlaff, P., Barto, E., Ree, M.J., & M.S. Teachout. 'Standard cognitive psychological tests predict military pilot training outcomes'. *Aviation Psychology and Applied Human Factors* 3, no. 1 (2013): 28-38.

Kozuba, J., & J. Pila. Chosen aspects of pilots situational Awareness/Odabrani aspekti situacijske svijesti pilota. *Nase More* 62*, no.* 3 (2015): 175-180.

Landon, L.B., Rokholt, C., Slack, K.J., & Y. Pecena. 'Selecting astronauts for long-duration exploration missions: Considerations for team performance and functioning'. *REACH – Reviews in Human Space Exploration* 5 (2017): 33-56. doi:10.1016/j.reach.2017.03.002

Lockheed Martin. 'F-104 Starfighter'. Lockheed Martin, webpage accessed 8 April 2023.

Li, Wen-Chin, and Don Harris. 'The Evaluation of the Effect of a Short Aeronautical Decision-Making Training Program for Military Pilots'. *The International Journal of Aviation Psychology* 18, no. 2 (2008): 135–52.

Li, Wen-Chin, and Don Harris. 'Identifying Training Deficiencies in Military Pilots by Applying the Human Factors Analysis and Classification System'. *International Journal of Occupational Safety and Ergonomics* 19, no. 1 (2013): 3–18.

Luciano, M., Posthuma, D., Wright, M.J., Geus, E.J.C., Smith, G.A., Geffen, G.M., Boomsma, D.I., & N.G. Martin. 'Perceptual speed does not cause intelligence and intelligence does not cause perceptual speed'. *Biological Psychology* 70 (2005): 1-8. doi:10.1016/j.biopsycho.2004.11.011

Madhura Ingalhalikar, Alex Smith, Drew Parker, Theodore D. Satterthwaite, Mark A. Elliott, Kosha Ruparel, Hakon Hakonarson, Raquel E. Gur, Ruben C. Gur, and Ragini Verma. 'Sex Differences in the Structural Connectome of the Human Brain'. PNAS 111, no. 2 (14 January 2012): 823–28.

Mahler, Julianne and Maureen Hogan Casamayou. *Organizational Learning at NASA: The Challenger and Columbia Accidents*. Washington, D.C: Georgetown University Press, 2009.

Mansell, Tony. 'Flying Start: Educational and Social Factors in the Recruitment of Pilots of the Royal Air Force in the Interwar Years'. History of Education (Tavistock) 26, no. 1 (1997): 71–90. https://doi.org/10.1080/0046760970260105.

Mansikka, Heikki, Kai Virtanen, Don Harris, and Petteri Simola. 'Fighter Pilots' Heart Rate, Heart Rate Variation and Performance during an Instrument Flight Rules Proficiency Test'. *Applied Ergonomics* 56 (2016): 213–19.

Martinussen, M. 'Pilot selection: An overview of aptitude and ability assessment'. In Bor, R. et al. (Eds.), *Pilot mental health assessment and support*. Routledge, 2016.

Martinussen, M., & D.R. Hunter. 'Aviation Psychology and Human Factors'. London: CRC Press, 2017.

Maruff, Paul, Marina G. Falleti, Alexander Collie, and David G. Darby. n.d. 'Fatigue-related Impairment in the Speed, Accuracy and Variability of Psychomotor Performance: Comparison with Blood Alcohol Levels'. *Journal of Sleep Research* 14, no. 1: 21–27.

Maschke, Peter, Viktor Oubaid, and Yvonne Pecena. 'How Do Astronaut Candidate Profiles Differ from Airline Pilot Profiles? Results from the 2008/2009 ESA Astronaut Selection'. *Aviation Psychology and Applied Human Factors* 1, no. 1 (2011): 38–44.

Mathewson, F.A.L. 'Medical Aspects of Aircrew Selection'. The Canadian Medical Association Journal, 1942, 318–22.

Metscher, Donald S., Marvin Smith, and Abdullah Alghamdi. 'Multi-Cultural Factors in the Crew Resource Management Environment: Promoting Aviation Safety for Airline Operations'. Journal of Aviation/Aerospace Education and Research 18, no. 2 (2009): 9–23. https://doi.org/10.15394/jaaer.2009.1423.

Milam, David W. 'A Pilot's Perspective on Workload in Single Seat Fighters,'. 1982.

Mittelstadt, J.M., Pecena, Y., Oubaid, V., & P. Maschke. 'Psychometric personality differences between candidates in astronaut selection'. *Aerospace Medicine and Human Performance* 87, no. 11 (2016): 933-939.

Müller, Michael P., Mike Hänsel, Andreas Fichtner, Florian Hardt, Sören Weber, Clemens Kirschbaum, Sebastian Rüder, Felix Walcher, Thea Koch, and Christoph Eich. 'Excellence in Performance and Stress Reduction during Two Different Full Scale Simulator Training Courses: A Pilot Study'. *Resuscitation* 80, no. 8 (2009): 919–24.

Musson, D.M., Sandal, G.M., & R. Helmreich. 'Personality characteristics and trait clusters in final stage astronaut selection'. *Aviation, Space, and Experimental Medicine* 75, no. 4 (2004): 342-349.

National Defence. 'Announcement Regarding the F-35 Acquisition'. Government of Canada. National Defence, 9 January 2023. https://www.canada.ca/en/department-national-defence/news/2023/01/announcement-regarding-the-f-35-acquisition.html.

'NATO Flying Training in Canada Integrated Training Plan'. 2 CDN Air Div HQ AF Trg, 2014.

Noel, Jeremy B., Kenneth W. Bauer Jr, and Jeffrey W. Lanning. 'Improving Pilot Mental Workload Classification through Feature Exploitation and Combination: A Feasibility Study'. *Computers & Operations Research* 32, no. 10 (2005): 2713–30.

O'Hare, D. 'Cognitive ability determinants of elite pilot performance'. *Human Factors* 39, no. 4 (1997): 540-552.

O'Leary, Mike. 'The British Airways Human Factors Reporting Programme'. *Reliability Engineering & System Safety* 75, no. 2 (2002): 245–55.

Ottawa Citizen. 'Airbus Deemed Only Qualified Supplier for New RCAF Refueling and VIP Aircraft'. 1 April 2021, Online edition.

Ouellet, Eric. 'Rethinking Military Leadership from a Sociological Perspective'. The Canadian Army Journal 12.1 (Spring 2009): 44–60.

Passolunghi, Maria C., Sara Caviola, Ruggero De Agostini, Chiara Perin, and Irene C. Mammarella. 'Mathematics Anxiety, Working Memory, and Mathematics Performance in Secondary-School Children'. Frontiers in Psychology 7, no. Journal Article (2016): 42–42. https://doi.org/10.3389/fpsyg.2016.00042.

Pederson, L.A., Allan, K.E., Laue, F.J., & J.R. Johnson. 'Personality theory for aircrew selection and classification'. TX: Armstrong Laboratory, 1992

Portman-Tiller, Claire A., Sean Biggerstaff, and Dave Blower. 'Relationship Between the Aviation Selection Test and a Psychomotor Battery'. Pensacola FL: US Naval Aerospace Medical Research Lab, 1998.

Progressive Digital Media Defense (incl, Airforce, Army, Navy and Homeland Security) News. 'Airbus Appoints CAE to Support RCAF's FWSAR Programme'. 15 February 2017.

Purvis, Bradley D. 'An Evaluation of B-1B Pilot Performance During Simulated Instrument Approaches With and Without Status Information'. Armstrong Laboratory, 1992.

Reason, J. Human Error. New York: Cambridge University Press, 1990.

Redick, Thomas S., Nash Unsworth, Andrew J. Kelly, and Randall W. Engle. 'Faster, Smarter? Working Memory Capacity and Perceptual Speed in Relation to Fluid Intelligence'. *Journal of Cognitive Psychology* (Hove, England) 24, no. 7 (2012): 844–54. https://doi.org/10.1080/20445911.2012.704359.

Retzlaff, Paul, J. D. Callister, J.D., and King, R.E. 'Clinical Procedures for the Neuropsychological Evaluation of U.S. Air Force Pilots'. *Military Medicine* 164, no. 7 (1999): 514–19. https://doi.org/10.1093/milmed/164.7.514.

Russell, Deborah, Jonathan K. Statz, John Ramiccio, Mike Henderson, David Still, Leonard Temme, Bethany Ranes, John Crowley, Arthur Estrada, and U.S. Army Aeromedical Research Laboratory Fort Rucker United States. 'Pilot Cueing Synergies for Degraded Visual Environments', no. Generic (2016). https://cfc.summon.serialssolutions.com/2.0.0/link/0

Rosenman, E.D., Dixon, A.J., Webb, J.M., Brolliar, S., Golden, S.J., Jones, K.A., . . . R. Fernandez. 'A simulation-based approach to measuring team situational awareness in emergency medicine: A multi-center, observational study'. *Academic Emergency Medicine* 25 (2018): 196-204.

Russell, D., Keegan, J., Ramiccio, J., Henderson, M., Still, D., Temme, L., Ranes, B., Crowley, J., & A. Estrada. 'Pilot cueing synergies for degraded visual environments'. USAARL 2016-10. Fort Rucker, AL: US Army Aeromedical Research Laboratory, 2016

Salas, Dan Maurino Eduardo, ed. *Human Factors in Aviation*. Second. UK: Academic Press, 2010.

Salas, Eduardo, Katherine A. Wilson, C. Shawn Burke, and Dennis C. Wightman. 'Does Crew Resource Management Training Work? An Update, an Extension, and Some Critical Needs'. *Human Factors; Santa Monica* 48, no. 2 (2006): 392–412.

Sauseng, P., Klimesch, W., Schabus, M., and Doppelmayr, M. 'Fronto-Parietal EEG Coherence in Theta and Upper Alpha Reflect Central Executive Functions of Working Memory'. International Journal of Psychophysiology 57 (2005): 97–103.

Scharinger, Christian, Alexander Soutschek, Torsten Schubert, and Peter Gerjets. 'When Flanker Meets the N-Back: What EEG and Pupil Dilation Data Reveal about the Interplay between the Two Central-Executive Working Memory Functions Inhibition and Updating'. Psychophysiology 52, no. 10 (2015): 1293–1304. https://doi.org/10.1111/psyp.12500.

Schiflett, Samuel G., Paul M. Linton, and Ronald J. Spicuzza. ‘Evaluation of a Pilot Workload Assessment Device to Test Alternate Display Formats and Control Handling Qualities’. PATUXENT RIVER MD: NAVAL AIR TEST CENTER, 1982.

Schneider, W. J., & K.S. McGrew. ‘The Cattell-Horn-Carroll model of intelligence’. In D. Flanagan & P. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues. Third. 99–144. New York: Guilford, 2012.

Schuppe, Thomas Frederick. ‘A Methodology for Predicting Pilot Workload’. ProQuest Dissertations Publishing, 1988.

Schvaneveldt, R. W., T. J. Breen, N. M. Cooke, F. T. Durso, and T. E. Goldsmith. ‘Cognitive Organization as a Function of Flying Experience’. Air Force Human Resources Laboratory, 1984.

Schvaneveldt, R.W., Branaghan, R.J., Lamonica, J., D.B. & Berringer. ‘Weather in the cockpit: Priorities, sources, delivery, and needs in the next generation air transportation system’. DOT/FAA/AM-12/7. Washington, DC: FAA Office of Aerospace Medicine, 2012.

Schweizer, S., Hampshire, A., & T. Dalgleish. ‘Extending brain-training to the affective domain: Increasing cognitive and affective executive control through emotional working memory training’. *PLoS ONE* 6, no. 9 (2011). doi:10.1371/journal.pone.0024372

Scott McPhalen. ‘Executive Function Assessment’, Email, 27 January 2023.

Signori, E.I. ‘The Arnprior experiment. A study of world war II pilot selection procedures in the RCAF and RAF’. *Canadian Journal of Psychology* 3, no. 3 (1949): 136-150.

Soeters, Joseph L., and Peter C. Boer. ‘Culture and Flight Safety in Military Aviation’. The International Journal of Aviation Psychology 10, no. 2 (2000): 111–33. https://doi.org/10.1207/S15327108IJAP1002_1.

Stanton, N. A., P. R. G. Chambers, and J. Piggott. ‘Situational Awareness and Safety’. *Safety Science* 39, no. 3 (2001): 189–204.

Stratton, G.M., McComas, H.C., Coover, J.E., & E. Bagby. ‘Psychological tests for selecting aviators’. *Journal of Experimental Psychology* 3, no. 6 (1920): 405-423.

Sulistyawati, K., Wickens, C.D., & Y.P. Chui. ‘Prediction in situational awareness: Confidence bias and underlying cognitive abilities’. *The International Journal of Aviation Psychology* 21, no. 2 (2011): 153-174. doi:10.1080/10508414.2011.556492.

Te Koeti, T. ‘Assessing situational awareness’. ASCI 670 Research Paper, Embry-Riddle Aeronautical University, 2020.

Te Koeti, T. ‘Pilot selection methods’. ASCI 634 Research Paper, Embry-Riddle Aeronautical University, 2020.

Te Koeti, T. ‘Improved aircrew selection methods: The association between executive functions, perceptual speed, and situational awareness’. MSHF 606 Research Paper, Embry-Riddle Aeronautical University, 2020.

Te Koeti, T. ‘Situational awareness in astronaut selection: Assessment of methods’. ASCI 691 Graduate Capstone Project, Embry-Riddle Aeronautical University, 2020.

Vidulich, M.A., Wickens, C.D., Tsang, P.S., & J.M. Flach. ‘Information processing in aviation’. In Eduardo Salas, D.M. (Ed.). *Human factors in aviation.* Second. 175-215. UK: Academic Press, 2010.

Webb, C.M., Gaydos, S.J., Estrada, E., & L.S. Milam. ‘Towards an operational definition of workload: A workload assessment of aviation maneuvers’. Fort Rucker, AL: US Army Aeromedical Research Laboratory, 2010.

Wendy Darr. ‘The Trait Self Descriptive (TSD) Inventory: A Facet-Level Examination’. Technical Memorandum. Director General Military Personnel Research & Analysis, February 2009.

Wickens, Christopher D. ‘Commentary on `Making Sense Out of Team Performance Errors in Military Aviation Environments’’. *Transportation Human Factors* 1, no. 3 (1999): 243.

Wong, Daniel S., Vinit M. Desai, Peter Madsen, Karlene H. Roberts, and Anthony Ciavarelli. ‘Measuring Organizational Safety and Effectiveness at NASA’. *Engineering Management Journal* 17, no. 4 (2005): 59.

Wongupparaj, P., Kumari, V., & R.G. Morris. ‘The relation between a multicomponent working memory and intelligence: The roles of the central executive and short-term storage functions’. *Intelligence* 53 (2015): 166-180. doi:10.1016/j.intell.2015.10.007

Wu, Changxu, and Yili Liu. ‘Development and Evaluation of an Ergonomic Software Package for Predicting Multiple-Task Human Performance and Mental Workload in Human–Machine Interface Design and Evaluation’. *Computers & Industrial Engineering* 56, no. 1 (2009): 323–33.

Yeh, Michelle. ‘Attention and Trust Biases in the Design of Augmented Reality Displays’. ProQuest Dissertations Publishing, 2000.