APTITUDE TESTING OF MILITARY PILOT CANDIDATES

By

Susan L. Forgues

A thesis submitted to the Graduate Proogram in the Department of Education
in conformity with the requirements for the
Degree of Master of Education

Queen’s University
Kingston, Ontario, Canada
(October, 2014)

Abstract

Flying a military aircraft is a cognitively complex activity. Military pilots must not only be able to fly the aircraft but they also must be able to seamlessly integrate the aircraft into a wide range of operational situations, working to complete complex missions in hostile terrain and under difficult circumstances. The overall goal of this thesis is to examine the specific cognitive abilities and/or demographic characteristics of Canadian Forces pilot candidates in aircrew selection using three aptitude test batteries.

There were three purposes of this study: to investigate relationships amongst the three aptitude test batteries completed by the pilot candidates, to determine if there were specific indicators that defined successful pilot candidates, and to examine the patterns of performance in flight simulator testing. Analysis of the relationships identified three factors, which were significant in a number of analyses and confirmed that candidates who were successful at aircrew selection possessed a number of common abilities. Specific groups of candidates were also identified based on their performance in the simulator. Candidates who scored well on Psychomotor Ability and Spatial Reasoning subtests were successful at pilot selection and Gender was consistently a significant factor in aptitude testing, with female candidates experiencing greater difficulty passing selection.

The development of systemically complex aircraft may have reduced the need for strong psychomotor abilities and instead generated an increased requirement for improved problem solving abilities and situational awareness. The current study demonstrated some movement towards this new dynamic by showing the importance of a Reasoning factor based on problem solving and critical thinking abilities, and an ability to work quickly and accurately under time constraints. Successful completion of pilot selection required candidates to be competent in a number of ability domains. More diverse abilities testing may select military pilot candidates whose performance during flight training is of a higher calibre as a result of their expanded skill set and who are better equipped to meet the challenges of today's complex and ever-changing air environment.

Acknowledgements

Completion of this thesis would not have been possible without the support of my family and the professors, staff and graduate students at the Faculty of Education of Queen's University. Thank you to my husband Pierre for his unwavering support and input. As always, we worked together and I am so thankful you were willing to proofread chapters and review data analysis over the past months. Thank you also to my Mom who listened politely as I recited a litany of difficulties to which she provided thoughtful and useful solutions and encouragement.

I owe Dr. John Kirby, my supervisor, a great debt of thanks for his unwavering support and encouragement from the very beginning of my Master's degree. Although military pilot selection is well outside his area of research, Dr. Kirby never hesitated to explore new avenues of inquiry and investigate new data analysis methods. The quality of this thesis is a reflection of his professionalism and dedication. Thanks also to Dr. Richard Reeve for his assistance and membership on my committee. To Danielle Lapointe-McEwen, Sean Cousins, Sana Tibi, Mary Bouchard, Yan Wei, Jess Chan, and Natalie Simper: thank you for being my sounding boards and general escape from the vagaries of writing and studying. I wish you success in your own endeavours wherever they may take you.

The Canadian Forces was the driving force behind the acquisition of this archival dataset and I want to thank Susan Truscott for the research idea, Major-General David Miller for his assistance in getting the data released to me, Dr. Wendy Darr at DGMPRA who provided critical information and assistance throughout the thesis-writing process, Lieutenant - Colonel Klammer for her review of the finished product, and Major Dawn Herniman whose guided tour through the Aircrew Selection Center in Trenton made all the difference in my approach to writing about the subtests.

To my close friends Val Arthur and Lisa Boyd, thank you for always taking my calls and listening to the updates on my progress (or lack thereof). Finally, thank you to Chance who was with me through most of the journey but not the end; I miss you.

Table of Contents

List of Figures

List of Tables

List of Tables (continued)

List of Abbreviations

CAPSS	Canadian Automated Pilot Selection System
CFAT	Canadian Forces Aptitude Test
RAF	Royal Air Force
RAFAAT	Royal Air Force Aircrew Aptitude Test

Chapter 1

Introduction

Flying a military aircraft is a cognitively complex activity. Military pilots must not only be able to fly the aircraft but they also must be able to seamlessly integrate the aircraft into a wide range of operational situations, working to complete complex missions in hostile terrain and under difficult circumstances. In light of the wide array of cognitive demands placed on military pilots, the aptitude testing and selection of pilot candidates needs to be a rigorous, multi-faceted process designed to assess the skills and capabilities of the pilot candidate in a variety of domains (Damos, 2003; Hilton & Dolgin, 1991; Wickens, 2007). Selection systems are methods of prediction, which can be tracked over time to maximize the quality of learning and achieve a greater degree of success in a chosen field (Cook & Ward, 1996). Selection systems act like filters to increase the likelihood of success in training or they can be used to select candidates who can master a satisfactory level of performance in a core skill at a faster rate (Cook & Ward, 1996).

The overall goal of this thesis is to examine the specific cognitive abilities and/or demographic characteristics that are markers for success of Canadian Forces pilot candidates in aircrew selection. The archival dataset used in this research comprised pilot candidate scores on three groups of measures: the Canadian Forces Aptitude Test (CFAT) which measures verbal and spatial abilities as well as problem solving acumen; the Canadian Automated Pilot Selection System or CAPSS, a computerized simulator that replicates tasks performed in flight; and selected subtests of the Royal Air Force Aircrew Aptitude Test (RAFAAT), an ability test battery developed in the United Kingdom.

The aptitude testing system designed to select Canadian Forces military pilots has, as its theoretical centre, a framework that assesses the general cognitive abilities of pilot applicants through a comprehensive aptitude test battery. The Review of Literature opens with a presentation of several influential theoretical models of human intellectual assessment and an overview of the foundation of aptitude testing. The concept of executive function (EF) is discussed as a construct consisting of interrelated but distinct components involved in goal-directed behaviour in novel situations. The importance of a comprehensive and accurate job analysis is highlighted because it identifies actual task competencies and personnel requirements.

The literature review continues with the examination of the ability domains used to classify cognitive capabilities according to the types of tasks and measures used to assess them. Where possible, recent empirical evidence assessing pilot performance in these ability domains, and how that performance may be influenced by EF, is presented. Specifics of simulator testing are also profiled given its importance in the aptitude testing of Canadian Forces military pilot candidates. The chapter concludes by outlining the research questions of the current study and describing other related studies concerning pilot selection in the Royal Canadian Air Force.

Chapter 2

Literature Review

Human Cognitive Abilities – Assessing Intelligence

General cognitive abilities influence how much and how quickly individuals learn, and predict their ability to react in innovative ways (Hunter, 1986). Charles Spearman (1904) coined the term *g* to designate general mental ability. Since Spearman, a myriad of theories and taxonomies has emerged in an effort to provide an organising scheme of human cognitive abilities. Several prominent theories have also contested the inclusion of *g* in an intelligence model, including Thurstone's Primary Mental Abilities theory (Thurstone, 1958) in which he proposed intelligence was based on seven primary abilities – spatial reasoning, perceptual speed, number facility, verbal relations, word fluency, memory, and inductive reasoning - and not on a single general reasoning factor. Sternberg (1986) distinguished three classes of intelligence: analytic, creative, and practical; Gardner's (1993) Theory of Multiple Intelligences was built on the premise that there was not one general trait of overall mental competence but many types of intelligence, ranging from musical skills to kinaesthetic intelligence. Despite these dissenting views, recent taxonomic models have been constructed around the concept of *g*. Arthur Jensen wrote:

> "The best single predictor of individual differences in the rate of learning and the level that can be attained in a great many areas of knowledge and skills that people regard as being of a mental nature is *g* …any group differences in *g* are really aggregated (or accumulated) individual differences." (cited in Miele, 2002).

The C-H-C Model. Of particular interest for this thesis is the Cattell – Horn – Carroll (C-H-C) Theory of Cognitive abilities, cited as "…the most comprehensive and empirically supported psychometric theory of the structure of cognitive and academic abilities to date" (Alfonso, Flanagan, & Radwan, 2005, p. 185). Beginning in the early 1960's, Cattell and Horn proposed the existence of two types of intelligence: fluid intelligence – Gf – which encompassed the basic abilities in reasoning as they related to higher mental processes, and crystallized intelligence – Gc – representing the extent to which an individual has been able to learn from experience and education (McGrew, 2009).

In 1993, Carroll presented a comprehensive, empirically based synthesis of factor-analytic research on the structure of human cognitive abilities. Carroll described a Three-Stratum Theory in which First Stratum abilities represented greater specialisations of abilities as a result of experience and learning. Second Stratum factors represent moderate specialisations of ability that can govern or influence behaviours in a given situation and Third Stratum abilities reflected differences in the performances of individuals in broad classes of tasks (Carroll, 1993). An example of the Three-Stratum Theory is provided by McGrew (2009): A Third Stratum ability would be *g* under which fluid intelligence, Gf, would be considered an integral Stratum II component; general sequential or deductive reasoning would then be a First Stratum ability.

The two theoretical approaches were combined to create the C-H-C theory of intelligence, a model that has had significant impact on the structure of cognitive testing. The C-H-C model has a single overarching Stratum III cognitive factor – *g*, then branches into Stratum II (broad) ability domains, which comprise up to ten broad abilities. An additional six have been suggested for inclusion so as to address human sensory domains (McGrew, 2009). In addition to Gf and Gc, the following Stratum II ability domains are of particular interest for this thesis: Gv - Visual-spatial abilities; Gsm - Short-term memory; Glr - Long-term storage and retrieval; Gs - Cognitive Processing Speed; Gt - Decision/reaction time or speed; and, Gp - Psychomotor Abilities. Stratum I or identified narrow abilities are described in McGrew (2009) and cover a wide range of cognitive abilities, including many that are detailed in the ability domains that are examined in this thesis.

Executive Function. The concept of Executive Function (EF) has been studied extensively in neuropsychological theories of behaviour control, specifically as it relates to the cognitive functions associated with voluntary control of behaviour (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). EF is not specifically named in the current theories of intelligence, including the C-H-C taxonomy, but some of its constructs like working memory (WM), attention, and inhibition, are represented in the Stratum II broad ability domains.

EF facilitates goal directed behaviour and adaptation to novel and complex situations, and allows the inhibition of automatic responses in favour of controlled, measured behaviour (Causse, Dehais, & Pastor, 2011). EF has been defined in a number of different ways: as a family of top-down mental processes needed for concentration/attention and when reliance on instinct or intuition would be ill-advised (Diamond, 2013); the ability to control cognitive actions by inhibiting impulsive task responses and manipulating/organising complex information held in working memory (Richland & Burchinal, 2013); those capacities that enable a person to engage successfully in independent, purposive, self-serving behaviour (Barkley, 2012).

There is also lack of agreement on whether EF should be considered as a unitary construct or as a set of independent components (Best & Miller, 2010). Zelazo, Carter, Reznick, and Frye (1997) proposed a problem-solving framework for EF that illustrated the manner in which distinct EF processes operate by integrating information in order to solve problems and achieve goals. In this model, four temporally distinct phases in EF problem solving are employed in the following sequence: problem representation, planning, execution, and evaluation.

In the models of Miyake et al. (2000) and Diamond (2013), the EF construct consists of the following distinct but interrelated components: inhibition, working memory (WM), and shifting. The inhibition component assists the individual in not relying on learned behaviours, instinct or intuition when confronted with novel and/or complex situations (Miyake et al., 2000). WM, as described by Baddeley (1986), represents a general cognitive workspace for concurrent processing and storage demands that are involved in complex learning activities, while shifting denotes cognitive flexibility or the ability to shift between states or tasks (Diamond, 2013).

Situational awareness is the perception of elements in the environment at a certain time and space, to include the comprehension of their meaning and the projection of their status in the near future

(Endsley & Bolstad, 1994; Vidulich, 2003). The accuracy of situational awareness depends on working memory (WM) to integrate incoming information with a coherent interpretation of current events to facilitate the prediction of the future status of a specific process or system (Sohn & Doane, 2004). McCabe et al., (2010) identified a common attention construct present in both WM capacity and EF tasks called executive attention.

Süß, Oberauer, Wittman, Wilhelm, and Schulze (2002) concluded that WM capacity was, in fact, the best predictor of intelligence and reasoning ability. They also argued that determining which specific function of WM – storage capacity, processing capacity, or both – is used to successfully perform reasoning tasks is not well understood (Süß et al., 2002). Task specific skills are necessary if an individual is to perform to a high level and these skills form the core of the job analysis component of ability assessment.

Theories of intelligence have evolved since the 1960s to make them more related to the constructs of cognitive psychology; thus the Stratum II abilities refer to terms such as short-term memory and long-term memory. More complex cognitive constructs, such as attention, working memory (and its various subsystems), metacognition, self-regulation, and executive function do not yet feature prominently in theories of intelligence, except that some authors argue that they are synonymous with *g* (Causse et al., 2011). These constructs are often discussed under the general label of Executive Functions and appear to be relevant to the skills required to fly airplanes (Causse et al., 2011).

Job Analysis

Whereas the constructs of human intelligence and cognition (like the Cattell-Horn-Carroll model or EF) can be thought of as a theoretical framework, job analysis represents the pragmatic framework of selection systems. Job analysis is an important component of assessment that should be incorporated into any selection system because it is critical to identify the actual job requirements and so refine the structure of the selection batteries for each role (Bailey, 1999; Cook & Ward, 1996; Kantor & Carretta, 1988). The work of Fleishman and Quaintance (1984) on the development of ability dimensions and measurement systems provided the foundation for the classification of tasks based on ability requirements, a critical step in the development of selection measures. In the ability requirements approach, tasks are described, contrasted, and compared in terms of the abilities they are thought to require of the operator, and then clustered into ability groups alongside other tasks with similar ability requirements (Fleishman & Quaintance, 1984).

The first step in developing new selection systems or implementing new technology is a thorough understanding of the task and a validation of the cognitive models of performance (Cook & Ward, 1996). Job analyses identify specific knowledge, skills, aptitudes, and other attributes required to perform specific tasks to a high standard (Darr, 2010a). When the Royal Air Force in the United Kingdom revised its selection system in the 1980's, they engaged subject matter experts – individuals with a thorough knowledge of the operational job requirements – who broke down each role into tasks, which were then weighted by their importance to mission success (Bailey, 1999).

Damos (1996) observed that job analyses of operational pilots were difficult to find, but essential to answering the question 'what is the job of the pilot?' In 2010, the Canadian Forces completed a pilot job analysis for each of the three pilot streams – Jet, Rotary Wing (Helicopter), and Multi-Engine – to determine commonalities and variations in the underlying knowledge and skills, associated with each stream (Darr, 2010a). Appendix A contains an excerpt from the job analysis of the Rotary Wing stream and includes a list of competency groupings. Psychomotor, mathematics, and reading skills, which would be considered Stratum II abilities Gp, Gq, and Grw respectively (McGrew, 2009), topped the list of skills that, if ignored in the selection process, would result in trouble for the novice pilot. The ability to operate under stress, to attend to multiple stimuli, to analyse the current situation, and anticipate changes were identified as the top three abilities that distinguished superior rotary wing pilots from average pilots (Darr, 2010a). These latter abilities relate to the inhibition and WM components identified by Richland and Burchinal (2013) as being part of EF and are related to the CHC Stratum II broad abilities of Gs - Cognitive Processing Speed; Gt - Decision/reaction time or speed, and Gsm - Short-term memory.

The Mental Abilities of Pilots

In this section of the literature review, ability domains are introduced. An ability domain can be considered as a broad collection of similar aptitudes; domain based composite scores were found to be more robust and reliable as they were comprised of a number of scores, each of which was derived from tests covering a range of similar aptitudes (Bailey, 1999; Carroll, 1993). Some of the ability domains described here correspond to CHC Stratum II broad ability domains; for others there are some similarities at the Stratum I narrow abilities level and these correspondences are noted where applicable. Ability domains are the selection framework used by many air forces in selecting pilots and represent the outcome of the theoretical approach to assessing intelligence combined with specific job requirements. The examination of the mental abilities of pilots begins with a description of the ability domains identified by the Royal Air Force and, where available, empirical evidence showing the results of pilot aptitude testing in these domains followed by an overview of EF and its role in pilot performance,

Aircrew-ability domains. As a result of the task analyses completed by the Royal Air Force, six aircrew-ability domains were identified (see Table 1 from Royal Air Force, 2007). These domains, known as the Legacy Cognitive Model, are defined in Table 1 and the corresponding CHC Stratum II broad abilities are identified. As part of ongoing research into aptitude testing, Canadian Forces military pilot candidates who completed Aircrew Selection between 2008 and 2013 also completed these tests. These data are the focus of the research for this thesis. The following sections examine the evidence assessing the roles of these ability domains in identifying pilot aptitudes.

Verbal Reasoning. Defined as the ability to interpret and reason with verbal information, verbal reasoning includes the assimilation and integration of information, inference, deduction, and evaluation of information (Southcote, 2004). Many journal articles addressed verbal reasoning skills as part of early cognitive development but none were found that specifically addressed the role of verbal reasoning in the pilot selection process. The ability of pilots to communicate was identified in the Darr (2010a) Job Analysis as an important consideration in the selection of helicopter pilots. Communication was defined as the ability to understand instructions in English and to speak clearly and this requirement was based on the ratings of Instructor Pilots who completed the job analysis.

Table 1

The Royal Air Force Aircrew Aptitude Test Legacy Ability Domains and Corresponding Cattell-Horn-Carroll (CHC) Stratum II Broad Ability Domains (McGrew, 2009)

Ability Domain	Definition	CHC Stratum II Factor
Verbal Reasoning	The ability to use and interpret written or spoken information, comprehend meaning from, and reason with, grammar, syntax, vocabulary, and sentence forms.	Gc, Grw
Numerical Reasoning	The ability to use and interpret information presented in the form of tables, graphs, and equations.	Gf, Gsm, Glr
Spatial Reasoning	The ability to form three-dimensional representations and manipulate diagrammatic information in ‘the mind’s eye’.	Gf, Gv
Work Rate	The ability to work accurately through routine tasks under time constraints.	Gf, Gs, Gt
Attentional Capability	The ability to deal with multiple tasks involving auditory and/or visual information, to concentrate over periods of time, noting changes, and paying attention to detail.	Gf, Gs, Gt
Psychomotor	The ability to perform tasks requiring eye-hand coordination and eye-hand-foot coordination with speed and accuracy.	Gf, Gp, Gps

Note. Gc – crystallised intelligence; Gf – fluid reasoning; Grw – reading and writing; Gsm – short-term memory; Glr – long-term storage and retrieval; Gv – visual processing; Gs – cognitive processing speed; Gt – decision and reaction speed; Gp – psychomotor abilities, Gps – psychomotor speed.

The current test battery for Canadian Forces pilot candidates contains a single test of verbal reasoning, the Canadian Forces Aptitude Test (CFAT) verbal reasoning subtest administered to all applicants to the Canadian Forces regardless of occupation. That the verbal reasoning domain is not included in the domains tested by the RAFAAT may be less an indication that it is not important for pilots, and more a reflection of the restricted sample considered for pilot selection. Applicants who do not score high enough on the CFAT simply do not proceed to aircrew selection (Darr, 2009).

Numerical Reasoning. Aptitude tests assessing numerical reasoning ability test a candidate's ability to comprehend, interpret, and use numerical information in a logical way (Southcote, 2004). Darr (2010b) observed that, with respect to measures capturing mathematical reasoning, the CFAT-Problem Solving (PS) subtest appeared to be relevant as it included a timed test requiring candidates to complete several mathematical problems. Within the RAFAAT test battery, there are two subtests in the numerical reasoning ability domain administered to Canadian Forces pilot candidates, one measuring mathematical reasoning, and the other mathematical computation.

Boccio (2009) highlighted the involvement of mathematical reasoning in aviation and the arithmetic skills required by pilots in order to obtain a private pilot endorsement from the Federal Aviation Administration in the United States. Boccio listed the following as required mathematical proficiencies: the ability to mentally estimate quantities; to convert units between different systems of measurement (e.g. knots to miles-per-hour or nautical miles to statute miles); to calculate angles to intercept desired navigation tracks; to perform vector operations to calculate headwinds, tailwinds, and cross-winds; to calculate square roots (e.g. to determine hydroplaning speed); and to read and interpret graphs.

Several published studies were found that concerned numerical reasoning requirements for pilots, however they were reports of rankings of pilot abilities completed by Instructor Pilots as part of more comprehensive studies identifying abilities that are critical for pilot success (Carretta, Rodgers, & Hansen, 1993; Youngling, Levine, Mocharnuk, & Weston, 1977). In Damos (2011), USAF pilots gave mathematical computation a moderate rating of cognitive ability relevance to pilot qualification but considered mathematical reasoning to be a little relevance.

Spatial Reasoning. Measures of spatial reasoning are a mainstay in pilot aptitude testing. Cooper and Regan (1982) defined spatial ability as competence in encoding, transforming, generating, and remembering internal representations of objects in space and in assessing their relations to other objects and spatial positions, while Dror, Kosslyn, and Waag (1993) suggested spatial ability encompassed the ability to rotate objects in mental images, to extrapolate motion, to scan imaged objects, to encode spatial relations between objects, and to extract the visual features of an object in the presence of visual noise. However, there is consensus that pilots should possess strong spatial aptitudes (Boer, 1991; Carretta, 2011; Carretta & Ree, 2000a; Dror et al., 1993). Boer (1991) concluded that pilots needed good spatial abilities not only because tasks such as navigation and air-to-air combat require them, but also because good spatial ability frees mental capacity for other tasks.

Maccoby and Jacklin (1974) identified three distinct categories of spatial tests: spatial perception (the ability to determine spatial relations despite distracting information); mental rotation (the ability to rotate quickly and accurately two or three dimensional figures in imagination); and spatial visualization (the ability to manipulate complex spatial information when several stages are needed to produce the correct solution). These categories are mirrored in the subtests of the Spatial Reasoning domain developed by the Royal Air Force. A fourth category of spatial reasoning subtest has been added to the RAFAAT subtest battery; spatiotemporal ability, defined as the ability to comprehend and manipulate spatial forms that have a dynamic quality (Southcote, 2004).

Dror et al. (1993) examined the spatial abilities of 16 male United States Air Force (USAF) pilots, age range 23 – 46 years (*M* = 30) and 16 male non-pilot control subjects from Harvard University and Armstrong Laboratories in Arizona, age range 21 – 44 (*M* = 29). Handedness and education levels were matched between the pilot and non-pilot subjects. The participants completed four spatial tests: mental rotation, motion extrapolation, motion scanning, and spatial relations. The mental rotation, motion scanning, and spatial relations subtests correspond to the Maccoby and Jacklin (1974) categories while the motion extrapolation subtest is more a test of spatiotemporal ability as defined by Southcote (2004).

In the mental rotation tests, the participants were required to determine if two sequentially presented objects were identical images or mirror images. Dror et al. determined that pilots were faster overall than non-pilots, $F(1, 30) = 6.75$, $p = .01$. In the motion extrapolation task, participants had to track a ball on the computer screen and then extrapolate its future position. In the motion scanning test, participants saw a ring composed of black and white squares; an arrow appeared briefly then disappeared, prompting participants to decide whether the arrow had been pointing at a black square or not. Dror et al. (1993) found no significant differences between the pilots and non-pilots in these two motion tests.

The final task assessing spatial relations abilities comprised two subtasks. In both subtasks, the stimuli comprised a narrow, horizontal bar and small X (0.4 cm^2). In the categorical subtask, participants had to decide whether, when the X appeared on the screen, was it above or below the bar. Exposure time was not specified however participants were tested at two difficulty levels: in the difficult trials, the X just touched the bar and in the easy trials, it was placed more than 2 cm above or below the bar. In the more complicated metric subtask, participants had to determine if the X was within ½ inch (1.27 cm) of the bar. Dror et al. found that pilots were better at judging metric spatial relations and were less affected by task difficulty in the metric task than were non-pilots; the pilots also made fewer errors during difficult metric conditions than non-pilots. There was no evidence that pilots judged categorical spatial relations better than non-pilots, however this may have been a result of the task simplicity (judging whether the X was simply above or below the bar). Overall, Dror et al. (1993) concluded that pilots possessed exceptional abilities in the mental rotation of objects and did not require as much extra time as the non-pilots when orientation differences increased. The faster rotation abilities of the pilots indicated that they seemed to be better at accessing spatial information in their memory and at shifting the locations of representations.

The results of the Dror et al. (1993) study provide a comprehensive overview of pilot spatial abilities however there are several areas that could be addressed should this type of testing be redone. The pilots Dror et al. (1993) tested were a restricted sample given that all USAF pilots must complete the Air Force Officer Qualifying Test, which requires that they meet minimum standards on several spatial aptitude composites. The results may have been different if civilian pilots had been used in the study as there is no spatial testing completed as part of civilian pilot licensing. Also, no details were provided concerning the professional backgrounds of the non-pilot candidates in the Dror et al. study and, although the researchers matched education levels between the pilot and non-pilot groups, proficiency in mathematics/science was not controlled in the spatial testing results.

Several studies (Lubinski, 2010; Nagy-Kondor & Sörös, 2012; Onyancha & Kinsey, 2007) have documented a link between proficiency in mathematics/science and spatial abilities. A final observation on the Dror et al. study concerns gender. In 1993, there were few serving female USAF pilots, however today the number of female pilots is likely large enough that they could be included in a spatial testing experiment to determine if gender is significant. Historically, females have scored poorly on mental rotation tests compared to their male counterparts (Hunter & Burke, 1994; Maccoby & Jacklin, 1974) so a study including female USAF pilots may substantiate or refute this gender discrepancy.

Work Rate. The domain definition of Work Rate provided by Southcote (2004) as the ability to work accurately through simple routine tasks under time constraints is vague concerning the specific cognitive abilities tested within the domain. In 2007, Southcote expanded the definition to include the specific aptitude of Perceptual Speed, defined as the ability to scan and search a visual scene quickly (Southcote, 2007). The CHC Stratum II broad ability equivalent is Gs, cognitive processing speed, which is primarily concerned with the time it takes to complete the task successfully, e.g. locating a particular letter in an array of random letters. The aptitude tests in the Work Rate ability domain can be characterised as clerical in nature (Southcote, 2004).

The vague and varied definitions of the Work Rate aptitudes made it difficult to identify empirical studies that examined the specific pilot abilities it encompasses. The RAFAAT subtests Table Reading and Visual Search assess a candidate's ability to read tables quickly and accurately, and to search for targets (letters or shapes) amongst distracters (Southcote, 2004). These subtests are scored solely on the number of correct responses in a limited time. The Work Rate domain may also include the Executive

Function (EF) components of working memory and shifting, depending upon the complexity of the task and the presence of distracting stimuli or secondary tasks to perform.

The fourth subtest in the RAFAAT Work Rate domain, entitled Vigilance, is considered a measure of both the Work Rate and Attentional Capability domains because it requires the pilot candidates to respond to both a single-step task and a multiple-step task concurrently. Southcote (2004) identified the single-step routine task as the Work Rate component of the subtest. In this task, candidates enter the coordinates of a star that must be cancelled, scoring points for the number of tasks completed but also losing points if they make errors. The multi-step priority task requires candidates to press a coloured key and then enter the coordinates of a cell where an arrow has appeared. Scoring for this task is based on the accuracy and speed with which they complete both the routine and priority tasks. Southcote (2004) identified this composite score as a measure of Attentional Capability, insofar as it tests candidates' abilities to deal with multiple tasks simultaneously, which also tasks WM resources.

Attentional Capability. The Attentional Capability aptitude domain comprises a broad range of abilities including working on multiple tasks concurrently, paying attention to details, and noting changes in those details over time (Southcote, 2004). The subtests of the RAFAAT Attentional Capability domain assess information processing abilities, situational awareness, working memory, and decision making (Royal Air Force, 2007).

Information processing. Barkhuizen, Schepers and Coetzee (2002) defined information processing as the process whereby any system associates or transforms new information in order to align it with stored information, prior to the creation of new information. They also considered information processing to be a function of intellectual ability that is representative of an individual's cognitive capacity. Bellenkes, Wickens, and Kramer (1997) identified attentional control, and its two subcomponents, perception and response, as critical components of a pilot's information processing system. Perception uses selective attention, defined as the decision to pay attention to or ignore events within and outside the aircraft, to trigger and execute a response (Bellenkes et al., 1997). Wickens (2007) expanded this two-component information-processing model to include situational awareness, working memory, and decision-making. Wickens (2007) considered these three components as overlapping components in a pilot's information processing system. The remainder of this ability domain section examines these components - situational awareness, working memory, and decision-making - as they relate to the attentional capabilities and information processing systems of pilots.

Situational awareness and working memory. The concept of situational awareness (SA) was earlier defined as one of the activities EF regulates through WM and shifting. A 2004 study by Sohn and Doane concerning WM processes and SA reinforced the interconnectivity of these information-processing components as defined by Wickens (2007). Sohn and Doane (2004) administered a series of tasks (memory span, situation recall, and SA) to 52 novice and expert pilots in order to assess the role of WM capacity and long-term working memory (LT-WM) in SA and whether those roles varied as a function of pilot expertise. The 26 novice pilots in the study had an average of 85.7 total flight hours in contrast to the expert pilots who had an average of 1116.8. No gender information was provided.

Two span tasks were administered as tests of WM capacity. In the spatial span task, participants were shown a set of five English letters (F, J, L, P, and R) and their mirror images one at a time in different orientations. Participants had to remember the orientation of each letter in the order they were presented while also deciding whether the image was normal or reversed in orientation. In the verbal span task differed in that participants were asked to recall seven English letters (G and Q were added) in the order of presentation rather than according to orientation.

The situation recall task and the SA task were both considered as measures of LT-WM. In the situation recall tasks the pilots were given either pictures or verbal descriptions of cockpit situations, and, after completing a 30 second intervening task, were asked to recall the depicted flight situation. In the SA task, the pilots viewed consecutive screens detailing a goal description (desired flight situation – altitude, airspeed, and/or heading) followed by pictures of cockpit instrumentation, whereupon they had to decide whether the aircraft in the cockpit pictures would reach the specified goal/flight situation.

Sohn and Doane concluded that WM capacity was critical for novice pilots whereas acquired LT-WM skills were important for expert pilots. In particular, because WM capacity predicted novice pilot SA, Sohn and Doane suggested that screening tests assessing the WM capacity dimension of a student's cognitive abilities might be useful in customising flight training (Sohn & Doane, 2004).

The Sohn and Doane (2004) study provided empirical evidence concerning the role of WM and SA in pilots with different expertise levels, however there are several caveats that must be addressed in the application of these results. Although little information besides total flight hours was provided for the two groups of pilots provided, it may be that the novice pilots were students at the flight schools and the expert pilots were their instructors. The unlicensed students would have been much less familiar with cockpit instrumentation with only a rudimentary understanding of the implications of flight instrument readings used in testing. A more useful gauge of the role of WM in SA may have been to compare moderately experienced civilian pilots who had all met a single flight test standard with the more experienced pilots. Including instructors in the experiment introduces confounding variables given that they are much more experienced than the students in terms of the type of flying they had completed i.e. cross country trips and instrument flying, and the instructors would have most likely flown several different types of aircraft, giving them more familiarity with cockpit instrumentation and better recall of the depicted flight situations in the situation recall/LT-WM task.

The use of total flight hours as a measure of expertise is open to debate. O'Hare (2003) observed that there were few differences in the information acquisition or decision making prowess between novice and experienced pilots when they were grouped based on total flight hours. In contrast, a number of differences in these cognitive processes were noted when pilots were grouped on the basis of cross-country flight experience as was done by Wiggins, Stevens, Howard, Henley and O'Hare (2002). In this study, novice and experienced pilots were identified based on task-specific experience i.e. cross-country flying rather than general flying experience leading to performance differences in problem-solving, information acquisition, and decision-making (Wiggins et al., 2002).

Decision-Making. Wickens (2007) included decision-making in the action section of the information processing system. Barkhuizen and Schepers (2002) considered the rate of decision making as a function of the complexity of available information and, like Wickens, considered pilots to be information processing devices interposed between the external environment and the controls of the aircraft. Zelazo et al. (1997) also included decision-making in their problem-solving model of executive function (EF). Decision-making is often characterised as the act of choosing between alternatives under conditions of uncertainty (O'Hare, 2003). At its most basic, decision-making comprises preparation and execution where preparation entails sensing and organizing information, while execution entails analysing and responding to the needs of the situation (O'Hare, 2003).

Aeronautical decision-making (ADM) will be discussed in detail in in the Review of Literature section entitled EF and pilots. ADM concerns the decision-making processes of pilots, when, in the face of uncertainty, the pilot must seek and acquire information from available sources, then process this data to reach a wise decision from a limited number of alternatives (O'Hare, 2003). ADM is an extensive field of research (e.g., Li & Harris, 2001; O'Hare, 1992, 2003) and should be considered as an integral component of the pilot's information processing system.

The abilities assessed in the Attentional Capability domain provide a comprehensive introduction to the final ability domain addressed in the literature review: psychomotor ability. Chaiken, Kyllonen, and Tirre (2000) identified situational awareness and mental capacity as contributors to an individual's psychomotor abilities. Ree and Carretta (1996) argued that WM, information processing, and psychomotor ability measure an aspect of *g* and are therefore important predictors of success in pilot training.

Psychomotor Ability. Subtests in the psychomotor ability domain assess different kinds of physical coordination and the ability to perform physical acts with both speed and accuracy (Southcote, 2004). Current computer-based testing of psychomotor ability enables test designers and administrators to present dynamic visual displays and to compile large data sets of psychomotor scores quickly and accurately (Fatolitis, Jentsch, Hancock, Kennedy & Bowers, 2010). Aptitude testing researchers like Fleischman (1972) and Carroll (1993) did not consider psychomotor ability to be a cognitive ability, however, current research includes psychomotor ability as a cognitive construct; its components, including tracking and coordination, are highly correlated with *g* (Chaiken et al., 2000; Carretta, 2011; Ree & Carretta, 1994). Measures of psychomotor coordination have remained a mainstay in pilot testing batteries in most Air Forces, as they are strongly related to flying tasks (Carretta & Ree, 1997; Carretta, 2011; Griffin & Koonce, 1996; Olson, Walker, & Phillips, 2010).

Wheeler and Ree (1997) examined the test results of 1,099 USAF pilot trainees; 98% were male ($n = 1077$), all were college [university] graduates between 23 and 27 years of age. The candidate testing took place between 1982 and 1993. The psychomotor tests, described in Table 2, included in the study were computer-based and classified as either tracking or reaction time tests. These psychomotor ability scores were used as predictors of pilot candidate performance on two flying training scores. The first score was the pass/fail final school grade on Undergraduate Pilot Training (UPT), a yearlong course comprising ground school and basic flying training on a single engine fixed wing aircraft ($n = 1099$). The second score was the mean score of daily flying and flight test averages on the primary and advanced flight training courses that followed UPT ($n = 833$).

The factor analysis completed by Wheeler and Ree produced a measure of general psychomotor tracking ability, p, and three lower order factors of specific psychomotor tracking ability named for the specific psychomotor test: two-handed coordination, complex coordination, and time sharing. The general factor p, was found to be a predictor of both performance criteria in both flying training scores, however the correlation between p and UPT pass/fail rates was small, $r = .285$, as was the correlation between p and daily flying/check ride average scores, $r = .287$ (both $p < .01$). Adding the lower order psychomotor tracking factors to p did not result in a significant contribution to either outcome.

Table 2

The psychomotor ability tests used by Wheeler and Ree (1997)

Test name	Aptitude assessed	Brief description	Scoring
Two-hand coordination	Rotary pursuit/ Pursuit tracking	A target travels an elliptical path on a computer screen. The participant uses two joysticks – one for vertical movement, one for horizontal movement – to keep a cross on the target as it moves.	The scores are the horizontal and vertical tracking distance errors.
Complex Coordination	Control precision; multi-limb coordination	The participant uses a dual-axis right control joystick to keep a cursor horizontally and vertically centred on a cross on the screen. Simultaneously, participants use a left single-axis joystick to centre a vertical bar at the base of the screen.	The three scores for the test are horizontal distance tracking error, vertical distance tracking error, and the tracking distance error for bar at the base of the screen.
Time Sharing	Rate control and reaction time	Two-part test: In part 1, participants keep randomly moving cross-hairs on an airplane target. In part 2, candidates repeat the tracking task from part 1 and cancel digits that appear at random intervals on the screen using a keypad.	Three scores: Tracking errors without digit cancellation; tracking errors during digit cancellation; and digit cancellation reaction time.

Note. Test information is taken from Carretta and Ree (1997) and Wheeler and Ree (1997).

While Wheeler and Ree (1997) focused solely on the relationship between a general psychomotor ability factor, *p*, and specific psychomotor tracking abilities, Carretta and Ree (1997) addressed the nexus of cognitive and psychomotor tests. The 354 United States Air Force non-pilot personnel completed psychomotor ability tests that included a pursuit-tracking task, a complex coordination task and a time sharing/attention splitting task. The cognitive tasks were taken from the Armed Services Vocation Aptitude Battery and comprised Arithmetic Reasoning, Word Knowledge, Mathematics Knowledge, and

Paragraph Comprehension. Caretta and Ree (1997) observed significant correlations between psychomotor and cognitive scores, the highest being between Arithmetic Reasoning and psychomotor ability, $r = .46$, $p < .05$. Chaiken et al. (2000) also found a significant overlap between psychomotor abilities and cognitive abilities, concluding that individuals with high psychomotor abilities learned faster, and that cognitively able individuals tended to do very well on psychomotor tests (Chaiken et al., 2000).

As comprehensive as the Wheeler and Ree (1997) study was, the test data were collected over a period of 11 years, during which time there were significant changes to the Undergraduate Pilot Training (UPT) program. Manning (2002) detailed a number of changes in the length of the course, varying between 49 weeks and 55 weeks in length. There were also changes in the hours allocated to the different aircraft types the students flew during UPT (range 173.3 - 260 total hours). Both these changes may have had a bearing on the pilot candidates' pass/fail scores.

EF and pilots. Much of the time, pilots find themselves in complex and/or novel situations requiring EF support for decision-making, problem-solving, reasoning, and planning activities (Causse et al., 2011). "EFs appear critical for handling the flight, monitoring the engine parameters, planning the navigation, maintaining up-to-date SA [situational awareness], correctly adapting to traffic and environmental changes, and performing accurate decision making by inhibiting wrong behavioural responses" (Causse et al., 2011, p. 219). Causse et al. (2011) examined the link between three EF composites – shifting, inhibition/level of impulsivity, and updating/working memory (WM) – and pilots' flight navigation performance and decision-making capabilities during landing. The participants were 24 male, native French-speaking pilots who held visual flight rules (VFR) flight ratings (*M* age = 43.3 years, SD = 13.6). Mean total flight experience was 1,676 hours (range 57 – 13,000 hours); all pilots in the study had flown within the previous two-year period.

The EF composites were assessed using a five-test neuropsychological battery. Target-hitting measured psychomotor reaction time; a Two-Back test (i.e. does the current stimulus match the stimulus shown two items ago?) assessed WM; a deductive reasoning test involving syllogisms measured overall reasoning performance; a card-sorting test evaluated shifting abilities; and a Stroop test measured inhibition. The level of impulsivity of the pilots was measured by a self-report impulse-scale that assessed quick decision-making (11 items), motor skills/acting without thinking (11 items), and non-planning/impulsiveness (12 items).

Flight performance testing comprised a 45-minute navigation flight scenario on a PC-based flight simulator in which the pilots completed a takeoff, flew to a specified waypoint using navigation instruments, and received instructions to land at a designated airport. Performance in navigation was measured using the angular deviation from the ideal flight path, summed from take-off until arrival at the navigation waypoint. Before reaching the designated airport for landing, the pilots received meteorological information concerning crosswind conditions on landing. The pilots were required to calculate the crosswind limitations of the simulated aircraft and make a decision to land as planned or to fly to a diversion airfield with better wind conditions. This landing decision produced a binary variable: ‘correct’ if the pilot opted for the diversion airport as the crosswind landing limits of the aircraft had been exceeded, and ‘incorrect’ if the pilot opted to land at the original airport, thereby exceeding the aircraft’s limitations.

Causse et al. (2011) determined that deductive reasoning performance was most predictive of pilot performance as measured by flight path deviations and the go/no go decision to land. The researchers attributed this result to the role of fluid intelligence, which plays an essential role in adapting to novel problems (McGrew, 2009). Causse et al. also concluded that updating ability using WM resources predicted pilot performance during the navigation phase, a finding they had expected given that flying takes place in a dynamic, constantly changing environment. Causse et al. (2011) did not find a significant contribution from shifting or inhibition to pilot performance, however, the researchers conceded this might have been a result of the flight scenario not requiring pilots to use these EF skills.

WM updating performance was also significant in the landing decision scenario, confirming the pilots’ ability to integrate new meteorological information concerning crosswind speeds into an established flight scenario. This finding supported that of Morrow et al. (2003) who showed that poor WM performance degraded the ability of pilots to follow air traffic control instructions. A high level of impulsivity was also predictive of the pilots' poor landing decisions and has been identified as a contributor to hazardous aeronautical decision-making resulting in pilot error, the causal factor responsible for 85% of crashes in general aviation (Causse et al., 2011).

The Causse et al. (2011) study of EF and pilot performance overlooked the large range in pilot flight experience, which may have confounded some of the results attributed to EF, specifically WM updating. Furthermore, a more complex scenario may have compelled the pilots to involve the shifting and inhibition components of EF in their flight performance, providing a more reliable indication of their role in pilot performance. Notwithstanding these limitations, the study provides a comprehensive introduction to pilot testing. The identified EF composites – inhibition, WM, and shifting/cognitive flexibility – are among the pilot aptitudes tested in the specific aircrew-ability domain subtests that were detailed in the previous section.

Sex Differences in Abilities Testing

Sex differences manifest in a number of ability domains including spatial abilities and psychomotor abilities, and the cause of these differences has been the focus of a great deal of research. For example, Ingalhaliker et al. (2013) modelled the structural connectome or neural connections of the brains of 949 youths using diffusion tensor imaging, and determined that there are genetic differences in the basic structure of the human brain. They concluded that male brains are structured to facilitate connectivity between perception and coordinated action, whereas female brains are designed to facilitate communication between analytical and intuitive processing modes. The research of Ingalhaliker et al. (2013) was part of a larger study, which included testing in several behavioural and aptitude domains. The female subjects outperformed males on attention, word, and social cognition tests while males performed better on spatial processing and sensorimotor speed.

Sex differences in spatial abilities. Sex differences in spatial abilities have been found in the area of spatial visualization, particularly in mental rotation and mental folding tasks (Ganley & Vasilyeva, 2011; Harris, Hirsch-Pasek, & Newcombe, 2013; Hult & Brous, 1986; Nazareth, Herrera, & Pruden, 2013; Voyer, Voyer, & Bryden, 1995). Both tasks require the dynamic spatial transformation of objects with respect to their spatial structure (Harris et al., 2013). Again, explanations for why males outperform females on these types of spatial tasks are wide ranging. Better spatial acuity for males may be correlated with more males participating in sports requiring high spatial visualization skills including basketball, squash, and soccer (Hult & Brous, 1986). Males are also exposed to an increased number of sex-typed activities like mechanical drawing, building models, and carpentry (Nazareth et al., 2013). Finally, the propensity of males to use spatial skills more often in solving math problems, particularly geometry may be correlated with better spatial abilities (Ganley & Vasilyeva, 2011).

Strong spatial abilities, particularly in mental rotation tasks, play an important role in navigation, which is the ability to process spatial information (Cherney, Brabec, & Runco, 2008). Navigation is a critical skill for pilots because, during flight, they continually assess spatial relationships between landmarks and perform mental rotations of those landmarks according to the structural properties of the available cues in the environment (Verde et al., 2013). Verde et al. tested the mental rotation abilities of 41 pilots (20 male and 21 female) from the Italian Air Force and 38 non-pilots who were college students with no flight experience. All participants completed a timed mental rotation test and a sense-of-direction questionnaire containing self-referential statements about aspects of their environmental spatial cognition e.g. knowledge and use of cardinal points, outdoor and indoor orienting ability, preference for landmark-centred geospatial representations.

Verde et al. found that gender differences in mental rotation capability were present in the non-pilot group but not the pilot group. Additionally, both male and female pilots had faster responses on the mental rotation tests than the non-pilots. Verde et al. hypothesized that this difference may have been a result of working memory constraints, specifically cognitive load; gender differences emerge only in high visuo-spatial working memory load tasks like mental rotation tasks, and females have been found to have lower visuo-spatial working memory capacity than males (Halpern, 1992).

Sex differences in psychomotor abilities. Carretta (1997) and Carretta and Ree (2000b) investigated whether pilot selection instruments measured the same factors for all groups and concluded that, based on the Basic Attributes Test (BAT) data between 1993 and 1996, all mean score differences favored men and were statistically significant. The largest effect sizes were for psychomotor coordination (two handed coordination and complex coordination tests) at 1.68. Overall, female applicants were less likely to meet or exceed minimum scores on Air Force Officer Qualifying Test (AFOQT) and BAT (Carretta, 1997; Carretta & Ree, 2000b). These findings mirrored those of Burke (1995), who observed larger mean score differences favouring males on psychomotor tests used for military pilot selection.

Simulator Testing in Pilot Selection – Work Sample

The era of information technology spurred a cultural shift that has transformed education, aptitude testing, and personnel selection by making computerized virtual environments available to almost everyone (Bartram & Bayliss, 1984; Macedonia, 2002). Burke et al. (1995) reviewed computer-based assessment (CBA) in aptitude testing and noted that computerization improved the accuracy of assessment, reduced test administration time, and facilitated tailoring test items to subjects' ability levels. In addition to facilitating more efficient, multi-aptitude test batteries, improved CBA permitted the inclusion of work sample tests in the form of flight simulation-based assessment (Burke et al., 1995; Carretta & Ree, 2008). Simulator-based testing has an intuitive appeal as a selection measure because it bears a strong resemblance to parts of the job for which the applicant is being selected (Carretta & Ree, 2000b). The flexibility of simulators allows for the testing of pilot candidates in a variety of realistic scenarios, which in turn permits the identification of those who may possess the aptitude to succeed in flight training (Gress & Wilkomm, 1996; Hunter & Burke, 1995).

Simulators used for aircrew selection provide candidates with an immersive experience in which they can demonstrate aptitude in a variety of domains including spatial reasoning, attentional capability, and psychomotor ability. These correspond to the following CHC Stratum II broad ability domains: Gf, Gv, Gsm, Gs, Gt, Gq, Gp, and Gps (see McGrew, 2009). Until October 2013, pilot applicants in the Canadian Forces were required to complete a number of sessions in the Canadian Automated Pilot Selection System (CAPSS) single engine aircraft flight simulator. Research by the Canadian Forces has found that CAPSS is a good predictor of pilot success in the early phases of training but less so in the more advanced flying training phases (Darr, 2009; Woychesin, 2000). Specifics of CAPSS and the testing regimen completed by the candidates are detailed in the Method section of this thesis.

The Current Study

This introductory chapter concludes by highlighting other studies that address topics related to pilot selection using the data analysed for this thesis and presents the three research questions that guided the current data analysis.

Recent Reports on Canadian Forces Pilot Selection

The archival dataset used in the completion of this thesis has been the subject of other studies completed for the Canadian Forces. Darr (2009) completed a psychometric examination of the Canadian Automated Pilot Selection System (CAPSS) with a focus on test or measurement bias. Darr's analysis revealed a significant difference in the distributions of CAPSS scores for men and women, with males' scores being negatively skewed. Darr (2009) questioned the fairness of selection decision based on CAPSS for female candidates and recommended that CAPSS be combined with other predictors of psychomotor ability for reducing adverse impact and sub-group differences.

Darr (2010b) also examined the predictive validity of the RAFAAT using CAPSS scores as an interim outcome. Her research comprised candidates' scores for 11 RAFAAT subtests (*n* = 455), the Canadian Forces Aptitude Test (CFAT) (*n* = 291), and CAPSS (*n* = 421). She found large sex differences only for the RAFAAT Sensory Motor Apparatus subtest, which was also the strongest predictor of CAPSS scores. Darr cautioned against using CAPSS as a proxy measure of flying training performance, recommending that RAFAAT predictive validity should be assessed using flying training outcomes including academic grades or flying ratings.

Herniman (2013) focused on the role of executive functioning (EF) in pilot selection and its predictive validity for pilot training success as part of a pilot selection battery. Herniman examined pilot candidate scores on CFAT, RAFAAT, and ExamCorp, a battery of computer-based measures of EF. She found that the inhibitory and sustained attention components of EF were predictive of academic performance during early flying training but not flying performance. She recommended that the role of EF be examined in later stages of flying training when pilot candidates fly more complex aircraft in more complicated flight scenarios as these situations require higher levels of multi-tasking and decision making abilities as well as improved situational awareness.

Johnson and Catano (2013) investigated the role of cognitive ability, previous flying experience, and CAPSS in predicting success in the three phases of Canadian Forces pilot training academic and flight performance. The cognitive ability testing analyzed in their research comprised candidate scores on the Canadian Forces Aptitude Test and CAPSS simulator scores. They determined that cognitive ability had a direct effect on academic achievement in early flying training; its effects on later flying training were mediated by the job knowledge acquired in the earlier phases. Johnson and Catano (2013) found that CAPSS was a more effective predictor of early flying training. Also, CAPSS predicted later flying training performance better for candidates who had little previous flying experience, accounting for 14% variance in their flight training performance compared to 3% for candidates with previous flying experience.

Research Questions.

Research for this thesis comprised analysis of an archival dataset of aptitude testing and demographic data collected at the Canadian Forces Aircrew Selection Centre in Trenton, Ontario between 2008 and 2013. Analysis focused on answering three research questions:

- What are the relationships amongst the aptitude tests completed by the pilot candidates?

- Are there specific demographic variables or aptitude test indicators that defined successful pilot candidates?
- Are there patterns of performance evident in the flight simulator testing that differentiate successful candidates from unsuccessful candidates?

Chapter 3

Method

This chapter is organized in three sections. The initial section contains a description of the participants whose demographic data and aptitude test scores comprise the dataset. This is followed by a description of the individual aptitude test batteries completed by the candidates and includes a description of each test, the aptitude abilities it assesses and its reliability. The final section of this chapter details how the aptitude tests were administered and explains the differences in *n* between the measures.

Participants

The dataset received from the Canadian Forces contained data for 1172 pilot candidates. Once the duplicate data had been removed, demographic information and aptitude test scores were available for 1067 candidates. Demographic data, available for the majority of candidates, included Age at Testing, Gender, and Educational Background. Age of the pilot candidates ranged between 17 and 49 years (*n* = 919); the mean age at testing = 22.6 years (SD = 5.3 years). Gender information was available for 1040 of the 1067 candidates (97.4%); 921 males and 119 females completed testing. Highest educational level achieved was available for 953 pilot candidates (89.3%). There were over 150 separate courses and degrees contained in the original data; these were recoded into three levels: candidates who completed high school (*n* = 510) were coded as 1; CEGEP/college graduates were coded as 2 (*n* = 108), and candidates who completed university/graduate school (*n* = 335) were coded as 3.

Measures

There were three groups of measures: the Canadian Forces Aptitude Test (CFAT), the Canadian Automated Pilot Selection System (CAPSS), and the Royal Air Force Aircrew Aptitude Test or RAFAAT.

Canadian Forces Aptitude Test (CFAT). Canadian Forces Recruiting Centers across the country administer the CFAT to all persons applying to be enrolled in the Canadian Forces, regardless of chosen occupation. All pilot candidates had completed the Canadian Forces Aptitude Test (CFAT) prior to their arrival at aircrew selection. The CFAT is a cognitive ability test used to screen Officer and Non-Commissioned Member applicants to the Canadian Forces and to classify applicants into various military occupations (Donohue, 2006). The CFAT is a speeded test; items increase in difficulty as the test progresses; test items that are not completed are scored as incorrect (Black, 1999).

The following information on the CFAT is taken from the practice version of the CFAT produced by Director Military Personnel Operational Research (DMPORA). The first section of the test assesses the candidates' verbal skills, specifically the ability to understand words and their meanings. This section comprises 15 multiple-choice questions for which the candidate chooses which one of four answers is the best one. The answers are marked on an answer sheet; the score on the verbal skills section of the CFAT is the number of correct answers. The second section tests candidates' spatial awareness. There are two types of problems in this section: in the one called PATTERN, the candidate is to find the form that can be made by folding a cardboard pattern and fitting it together. In the second, called FORM, the candidates are to determine what the cardboard pattern would look like if the form were unfolded. The CFAT Aptitude test spatial ability score is the number of correct answers. The third section of the CFAT concentrates on problem solving. The 30 questions are multiple choice and the candidates must choose which one of the four answers is the **best** one. The problems are numerical, verbal and spatial in nature and the score is the number of correct answers.

The dataset contained scores on the CFAT for 1052 candidates. Gender information was not available for all candidates; *n* = 920 male pilot candidates; *n* = 118 female pilot candidates. Black (1999) found Cronbach's alpha coefficients of .87, .88, and .91 for the Verbal Skills, Spatial Ability, and Problem-Solving scales respectively.

The Canadian Automated Pilot Selection System (CAPSS). Until October 2013, applicants seeking entry into the pilot occupation in the Canadian Forces were required to complete four one-hour sessions on the Canadian Automated Pilot Selection System (CAPSS). CAPSS is a computerized simulator of a single engine light aircraft, which presents pilot candidates with the information necessary to perform flight manoeuvres using instrument flying procedures (Woycheshin, 2000). Each session reflected an increasing complexity with respect to flying manoeuvres and flight patterns. A number of basic flight manoeuvres were tested: basic flight instruments and controls; straight and level flight, straight climb, straight descent, take off, climb out and level off, level turns, standard rate turns, climbing turns, descending turns, and airport traffic patterns.

Candidates were assessed on their accuracy in maintaining assigned altitude, airspeed, and heading, their speed of response in correcting errors, and the smoothness and coordination of the operation of the flight controls (Woychesin, 2000). Based on their accuracy, applicants received a score ranging between .000 and 1.000 on each session, with a higher score reflecting better performance. The CAPSS selection score was based on a cut-off score of .70 on session 4 (Darr, 2010). In order to obtain the CAPSS Pass/Fail variable, CAPSS 4 session scores at or above .70 were coded as 1 (pass) and scores that were below the cut-off were coded as 0 (fail).

The number of candidates completing the Canadian Automated Pilot Selection System (CAPSS) changed for each of the four sessions as detailed in Table 3. Differences in *n* resulted from poor scores on the early simulator sessions. In some cases, CAPSS testing ceased for candidates with scores lower than .2 on sessions 1 and 2, which affected four male candidates and five female candidates. Also, ‘crashing’ (exceeding the flight limitations of) the CAPSS five times on the same maneuver resulted in immediate cessation of testing for that candidate.

Table 3

Number and Gender of candidates completing CAPSS Testing by Session

CAPSS Session (*n* total)	*n* Male candidates	*n* Female candidates
1 (1026)	888	111
2 (1014)	884	107
3 (1013)	884	106
4 (1011)	882	106

Note. Gender information was available for 92.6% candidates (998/1011). Changes in *n* were caused by candidates' failure to achieve required performance levels to move to the next CAPSS session.

Royal Air Force Aircrew Aptitude Tests (RAFAAT). The Royal Air Force Aircrew Aptitude Test subtests were administered by the Aircrew Selection Center in Trenton, Ontario the day before the candidates completed their CAPSS simulator sessions. Candidates were informed that their performance on the RAFAAT subtests would be used for research purposes only and would not be used in the selection process. The Royal Air Force Aircrew Aptitude Test (RAFAAT) comprised a battery of tests developed by the RAF for use in the selection of personnel for Pilot, Navigator, Air Traffic Controller and Air Engineer. All tests are self-administered on the Officer and Aircrew Selection computer-based system.

Information on the subtests comes from Royal Air Force (2007), Royal Air Force Aptitude Testing System (2013), and Southcote (2004). The subtests assess five ability domains: Numerical Reasoning, Spatial Reasoning, Work Rate, Attentional Capability, and Psychomotor Ability as detailed in Table 1 in the Literature Review of this thesis. Within each domain, the subtests are described and their associated reliabilities are reported as determined by the Royal Air Force. No reliability information was available for candidate testing by the Royal Canadian Air Force.

Numerical Reasoning domain. Numerical reasoning subtests assess the candidates' ability to interpret and reason with numerical information, to identify patterns in presented information, and solve problems using a logical approach. Subtests included in this domain are Mathematical Reasoning and Numerical Operations.

Mathematical Reasoning. Mathematical Reasoning is a test of a candidate's ability to solve mathematical problems. Candidates are required to solve aircraft-related, time/speed/distance problems, which require mathematical reasoning skills rather than the ability to perform mental arithmetic. The duration of the test is 18 minutes. Scores reflect the number of correct answers, range 0 - 24. Bradshaw (1997, cited in Southcote, 2004) found a test-retest reliability coefficient of .75, n = 832.

Numerical Operations. Numerical Operations is a test of mental arithmetic. Each item is a basic arithmetic problem based upon the following operators: addition, subtraction, multiplication, and division. Candidates use a numerical keypad to answer each test question. The average test duration is 2.5 minutes. Scores reflect the number of correct answers, range 0 - 50. Bradshaw (1997, cited in Southcote, 2004) found a test-retest reliability coefficient of .92, n = 40.

Spatial Reasoning domain. This domain assesses candidates' ability to form mental pictures and mentally manipulate spatial information. Subtests included in this domain are Critical Reasoning; Angles, Bearings, and Degrees; Directions and Distances; Instrument Comprehension 1; and Instrument Comprehension 2.

Critical reasoning – diagrammatic subtest. The three parts (verbal, numerical, and diagrammatic) of this subtest are designed to assess general reasoning aptitude. The diagrammatic test was the only portion used in testing Canadian Forces candidates. This subtest assesses spatial reasoning aptitude and the ability to manipulate diagrammatic or pictorial information. The test is 15 minutes in duration. Scores reflect the number of correct answers, range 0 -16. Bailey and Southcote (2007 cited in Royal Air Force, 2007) found the internal consistency reliability for CRBD to be .362 (Cronbach's Alpha) and .406 (KR-21).

Angles, bearings & degrees. The Angles, Bearings, and Degrees score should be interpreted as a measure of one part of the spatial aptitude and should be used only in conjunction with other spatial tests in order to give a more reliable estimate of spatial ability. The test comprises two parts: Angles, Bearings, and Degrees 1 (Angles) measures a candidate's ability to judge the size of angles. Angles, Bearings, and Degrees 2 (Bearings) measures the ability to judge the bearing of one object from another. Both parts include practice questions and actual multiple-choice test items. The tests are timed: 3.5 minutes for each of the two subtests. Two scores are produced – one for Angles, Bearings, and Degrees 1 and one for Angles, Bearings, and Degrees 2. The executive score is the combined number of correct items, range 0 – 30 for each of the two tests, total range 0 - 60. Bradshaw (1997, cited in Southcote, 2004) found a test-retest reliability coefficient of .64 for Angles, Bearings, and Degrees 1, *n* = 125, and .84 for Angles, Bearings, and Degrees 2, *n* = 123.

Directions and distances. Directions and Distances is a spatial reasoning test of the candidate's ability to use and interpret verbal descriptions of spatial relations. The candidate reads a paragraph of text giving the relative distance and directional relationship of a variety of objects and then answers questions regarding the distance and bearing of two objects in particular. Alternatively, the paragraph might describe a route taken by an 'actor' and the question asks the distance and bearing of the actor's final position from a given point. The test duration is 11.5 minutes. Scores reflect the number of correct answers, range 0 - 15. Bradshaw (1997, cited in Southcote, 2004) reported a test-retest reliability coefficient of .93, *n* = 41.

Instrument comprehension 1 and 2. These two subtests assess candidates' spatial visualization abilities using spatial, numerical, and verbal information. Part 1 presents five three-dimensional pictures of an aircraft in different orientations and two pictures of aviation instruments – artificial horizon and compass. The candidate must inspect the instrument readings and identify which of the five aircraft orientations accurately corresponds with the instrument readings. Part 2 presents six aircraft instruments (altimeter, artificial horizon, airspeed, vertical speed, compass, turn & bank) in the top half of the screen while in the bottom there are five verbal descriptions of an aircraft's orientation. The candidate must inspect the instrument readings and select the description that corresponds with the readings. The test is timed: 9 minutes per subtest. Two scores are produced, one for each part. The scores are based on the number of correct answers. Bailey and Southcote (2007, cited in Royal Air Force, 2007) found a test-retest reliability coefficient of .698, $p < .01$, $n = 580$.

Work Rate domain. Subtests in the Work Rate domain assess the ability of candidates to scan and cross-reference tables quickly and accurately or search for a target amongst a number of distractors (Southcote, 2007). Subtests included in this domain are Table Reading, Visual Search 1 - Letters, Visual Search 2 - Shapes, and Vigilance.

Table reading. Table Reading requires candidates to look up hard copy table chart data for answers to a set of multiple-choice items to test work rate and ability to work accurately through simple tasks under a time constraint; each part of the test has a time limit of three minutes. The test consists of two parts: Part 1 requires candidates to cross-reference two given row and column numbers to find a third tabulated value in a numerical reference table. Part 2 requires candidates to use a set of tables that describe the relationship between wind velocity, wind angle, drift correction, and ground speed for different airspeeds. In the questions, one of the values is missing; candidates must solve for that value using the table. The score is the total number of correctly answered items from both parts, range 0 -86. Bradshaw (1997, cited in Southcote, 2004) reported a test-retest reliability coefficient of .73, $n = 843$.

Visual search 1 and 2. These subtests are measures of the candidates' ability to look for a target amongst a set of distracters. Visual Search 1 involves searching for a particular letter in the matrix of letters. Visual Search 2 involves searching for a specified shape in a matrix of shapes. The candidate is shown a matrix of tiles; on each tile is a large object (e.g. letter E or a shape) and a small reference number in the bottom right corner. Candidates are given the tile object to search for in a matrix of tiles, and must enter the reference number once it is found. Each part of the test has a time limit of 1.25 minutes. There are two scores – one for each part, range 0 – 74 – reflecting the number of correctly identified targets. Bradshaw (1997, cited in Southcote, 2004) reported a test-retest reliability coefficient of .73, $n = 125$ for VIS 1 and .71 ($n = 125$) for VIS 2.

Vigilance. The Vigilance subtest is a measure of ability to detect infrequently occurring events under high workload. Candidates are presented with a 9 x 9 matrix on the screen. Each cell in the matrix is identified by two reference numbers (1-9) running along the top and down the left-hand side of the matrix. Candidates are required to attend to two tasks, one routine and the other priority. The routine task involves the cancellation of stars and the priority task the cancellation of arrows. An arrow is cancelled by a two-step procedure. The test duration is 8 minutes. Scoring is a derived from three scores: the first score, range 0 – 589, is based on the number of stars cancelled correctly; the second score, range 0 – 700, is based on the errors made while cancelling the stars; the third score, range -400 - +316, is based upon the speed and accuracy of the cancelling both the stars and arrows. The alpha reliability has been reported as .908 (Bailey & Southcote, 2007 cited in Royal Air Force, 2007)

Attentional Capability. Attentional Capability assesses the efficiency with which an individual can deal with visual and auditory information in real time. It is related to working memory capacity and attentional flexibility (Royal Air Force, 2007). Subtests included in this domain are Digit Recall; Colours, Letters, and Numbers; and Digit Recognition.

Digit recall. This subtest captures attentional capacity, particularly the ability to retain information in short-term memory. The task requires candidates to remember a string of digits presented on the screen for five seconds, and to accurately enter this information from memory. The total test duration is 4 minutes. The score is the total number of correctly reported digits, range 0 - 135. A reported digit is judged to be correct if it is entered in the same position as originally presented. Bailey and Southcote (2007, cited in Royal Air Force, 2007) found the alpha reliability to be .877.

Colours, letters, and numbers. Colours, Letters, and Numbers is a triple task test designed to assess how effectively candidates are able to multi-task under increasingly demanding conditions. The test is based on the following three sub-tasks: a simple continuous monitoring and tracking task (Colours), a short-term verbal memory task (Letters), and a mental arithmetic task (Numbers). In the Colours task, coloured diamonds move in straight lines across the screen and enter three coloured vertical bands. When a diamond is masked (in a band of the same colour) it may be cancelled by pressing a same-coloured key on the keyboard. In the Letters task, a target string of 5 – 8 letters is presented briefly; it is removed and after 12 seconds four different answer strings are presented. The candidate must key in a letter corresponding to the correct letter string. The numbers task is simple mental arithmetic; candidates enter their answer using the numeric keypad. The test is 22 minutes in duration.

Scoring varies by subtest. In the Colours subtest, candidates are rewarded for correct responses but penalized for incorrect responses. For the Letters task, candidates receive 1 point for each correct answer with no penalties. For the Numbers task, candidates are awarded 1 point for each correct answer but lose 1 point for each incorrect answer. The individual test scores are combined into one overall score. An internal-consistency reliability coefficient of .506 (Cronbach's Alpha) was found; the test –retest reliability coefficient for CLAN was .764 (n = 2254) (Bailey & Southcote, 2007 cited in Royal Air Force, 2007).

Digit recognition. Digit Recognition is a test of candidates' working memory. Candidates are shown a string of digits for a few seconds. The string of digits is then removed and immediately afterward the candidates are asked to indicate, using a keypad, how many times a particular digit appeared in the string. The size of the string presented increases throughout the test, beginning with seven digits and increasing to 15. The test duration is approximately 4.5 minutes. The score is the number of correct items, range 0 - 15. Bradshaw (1997, cited in Southcote, 2004) reported a test-retest reliability coefficient of .77, n = 125.

Psychomotor Ability. Psychomotor ability pertains to different kinds of physical coordination and encompasses the ability to perform physical acts with both speed and accuracy. Subtests included in this domain are Control of Velocity and Sensory Motor Apparatus.

Control of velocity. Control of Velocity is a pursuit-tracking psychomotor test measuring hand-eye coordination. Candidates must follow red circular targets as they follow an oscillating path descending from the top of the screen. Candidates use a pointer controlled by a joystick to hit as many of the descending targets as possible. There is an element of anticipatory tracking as the candidate is aware of only a portion of the track at any given time. Test duration is five minutes. The candidate scores one point for each target hit (maximum score 250). An alpha reliability or .938 was found; the test-retest reliability was determined to be .801, $p < .01$, $n = 2266$ (Bailey & Southcote, 2007, cited in Royal Air Force, 2007).

Sensory motor apparatus. Sensory Motor Apparatus is a compensatory tracking test that measures hand-eye-foot coordination. Candidates use a joystick and rudder pedals to move a pointer (small circle) both horizontally and vertically on a computer screen. In the center of the screen is a graticule (cross-hair). During the test, the pointer appears to move about the screen in a random manner and the candidate's task is to bring the pointer back to the center of the graticule using the joystick and rudder pedals. The test duration is five minutes. Performance on the SMA is indicated using an error score. The screen is separated into two areas, an inner area and an outer area. If the candidate fails to keep the pointer in the inner area then an error is recorded; thus higher scores equal more errors. For purposes of analysis, the scores were reversed (subtracted from 300) so that the lower scores represent poorer performance.

Summary

These five ability domains are the Legacy Domains as they were the initial set of ability-domain classifications developed for use by the Royal Air Force (Bailey, 1999). These domains comprise two separate groups of RAFAAT subtests that were administered to pilot candidates at the Canadian Forces Aircrew Selection Center (see Table 4). The RAFAAT Group 1 subtests were administered to a larger group of pilot applicants between 2008 and 2013; the Group 2 subtests were administered to a smaller number of candidates in 2012 and 2013. All candidates who completed the subtests of RAFAAT 2 also completed the subtests of RAFAAT 1.

Table 4

Subtests of Royal Air Force Aircrew Aptitude Tests (RAFAAT) Grouped by Legacy Domain (n = Number of Candidates Who Completed Each Subtest)

RAFAAT Group 1	RAFAAT Group 2
	Numerical Reasoning
	Mathematical Reasoning ($n = 560$)
	Numerical Operations ($n = 544$)
Spatial Reasoning	**Spatial Reasoning**
Critical Thinking - Diagrammatic ($n = 1067$)	Angles, Bearings, and Degrees ($n = 557$)
	Directions and Distances ($n = 544$)
	Instrument Comprehension 1 ($n = 544$)
	Instrument Comprehension 2 ($n = 544$)
Work Rate	**Work Rate**
Table Reading ($n = 1053$)	Vigilance ($n = 583$)
Visual Search 1 – Letters; ($n = 1053$)	
Visual Search 2 – Shapes ($n = 1053$)	
Attentional Capability	**Attentional Capability**
Recall Numbers ($n = 1053$)	Colours, Letters, and Numbers ($n = 560$)
	Digit Recognition ($n = 544$)
Psychomotor Ability	
Control of Velocity ($n = 1024$)	
Sensory Motor Apparatus ($n = 1036$)	

Chapter 4

Results

This chapter is organised according to the three purposes of this study: to investigate relationships amongst the three aptitude test batteries completed by the pilot candidates, to determine if there were specific demographic variables or aptitude test indicators that defined successful pilot candidates, and to examine the patterns of performance in the flight simulator testing completed as part of the pilot selection process.

Relationships amongst the measures

Descriptive statistics for the Canadian Forces Aptitude Test (CFAT) and the Royal Air Force Aircrew Aptitude Test (RAFAAT) Group 1 and Group 2 subtests are presented in Table 5. The RAFAAT Group 1 subtests can be identified by the larger number of candidates who completed them. A correlation table for these measures can be found in Appendix B. In general, higher correlations were observed within the same ability domain. Exceptions include Instrument Comprehension 1 and 2 (Spatial Reasoning domain) having correlations of .191 and .179 ($p < .01$) respectively with the Spatial Ability subtest of the CFAT. Correlations of Digit Recognition, identified as an Attentional Capability subtest, are low with all other subtests, including another Attentional Capability subtest: Colours, Letters, and Numbers, $r = .130$, $p < .01$.

A principal axis factor analysis with direct oblimin rotation was conducted to determine the dimensions underlying candidate abilities on the CFAT and RAFAAT Group 1 subtests ($n = 1024$). RAFAAT Group 2 subtests were not included because n dropped substantially, ($n = 513$). The one, two, three, and four factor solutions were evaluated with the following criteria: eigenvalues > 1.0, scree plot, variance accounted for, and interpretability. The scree plot for the factor analysis is in Figure 1 and shows that there are four eigenvalues > 1.0; there is a large difference between the first and second unrotated factors but then the differences diminish. Note that the eigenvalues are those provided by SPSS that come from a principal components solution.

Table 5

Descriptive Statistics for the Canadian Forces Aptitude Test (CFAT) and All Royal Air Force Aircrew Aptitude Tests in Six Ability Domains

	Domain	N	Min.	Max.	Mean	S.D.
CFAT verbal skills	VR	1052	2	15	10.65	2.53
CFAT spatial ability	SR	1052	4	15	11.64	2.19
CFAT problem solving	VR/SR	1052	3	30	24.32	3.71
Mathematics Reasoning	NR	560	1	21	9.87	3.738
Numerical Operations	NR	544	13	50	34.79	8.67
Critical Thinking Diagrammatic	SR	1067	1	15	7.23	2.25
Angles, Bearings, and Degrees	SR	557	20	56	42.37	5.73
Direction and Distance	SR	544	1	15	8.28	2.61
Instrument Comprehension 1	SR	544	1	23	11.45	4.05
Instrument Comprehension 2	SR	544	3	19	12.31	3.03
Table reading	WR	1053	15	86	60.48	10.18
Visual Search 1 letters	WR	1053	36	74	56.92	5.87
Visual Search 2 shapes	WR	1053	0	72	55.60	6.11
Vigilance	WR	583	-3	215	145.25	28.47
Recall Numbers	AC	1053	59	134	97.56	13.07
Colours, Letters, and Numbers	AC	560	-1676	674	119.10	219.34
Digit Recognition	AC	544	4	15	9.75	2.00
Control of Velocity	PA	1024	0	141	103.46	15.50
Sensory Motor Apparatus	PA	1036	25	297	181.29	43.66

Note. VR – Verbal Reasoning; SR – Spatial Reasoning; NR – Numerical Reasoning; WR - Work Rate; AC - Attentional Capability; PA - Psychomotor Ability.

The one factor solution (see Appendix C) had large factor loadings for both of the Work Rate measures and two of the four subtests in the Verbal and Spatial Reasoning domains. Both Psychomotor Ability subtests had only moderate loadings. The two-factor solution showed a strong Work Rate factor and a combined Verbal/ Spatial Reasoning and Psychomotor Ability factor while the four-factor solution had a singleton in factor four, the CFAT Spatial Ability subtest. The three-factor solution showed three clear factors in which almost all of the tests were involved, and so it was selected (see Table 6; correlations between factors are shown in Table 7). Details of the one, two, and four-factor solutions can found in Appendix C.

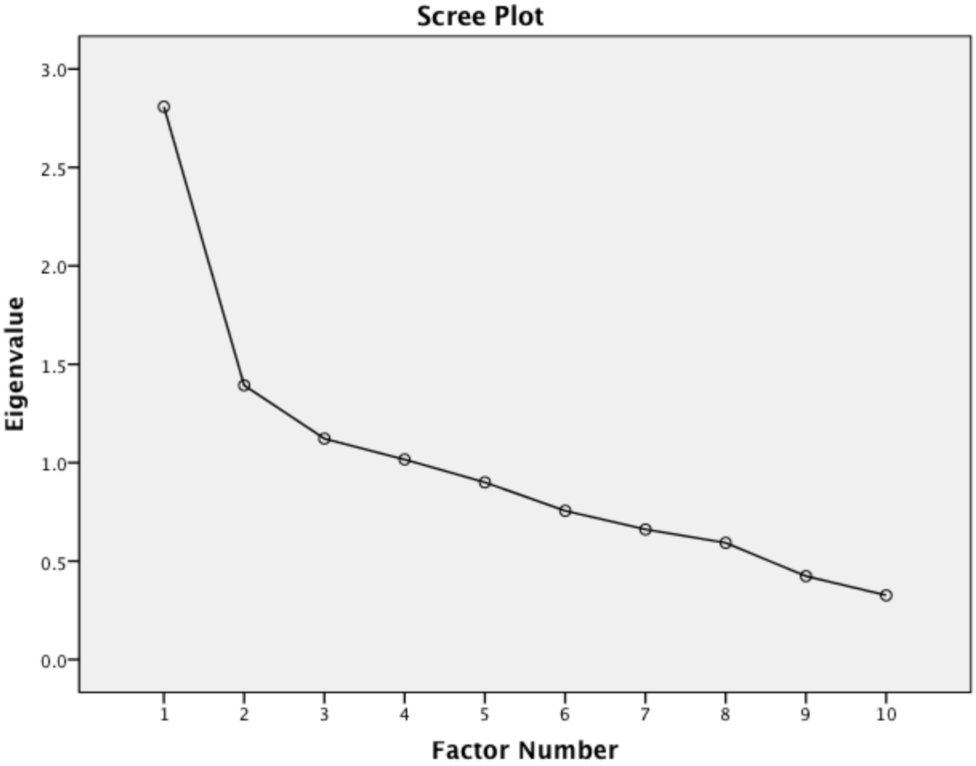

Figure 1. Scree plot for the Factor Analysis.

The first factor of the three-factor solution was defined by the three subtests from the Work Rate ability domain, and the second factor by the two psychomotor subtests from the RAFAAT battery. The Critical Reasoning subtest from the RAFAAT battery and the Problem Solving subtest from the CFAT had the largest loadings on the third factor, with smaller loadings from the CFAT Verbal Skills and Spatial Ability subtests.

Table 6

Principal Axis Factor Analysis with Direct Oblimin Rotation for RAFAAT Group 1 and CFAT Subtests (N = 1007)

		Factor		
Measure	Domain	1	2	3
Table Reading	WR	**0.603**	0.150	0.149
Visual Search 1 - Letters	WR	**0.910**	-0.074	-0.077
Visual Search 2 - Shapes	WR	**0.780**	-0.004	-0.064
Recall Numbers	AC	0.217	0.032	0.132
Sensory Motor Apparatus	PA	-0.062	**0.791**	-0.039
Control of Velocity	PA	0.037	**0.486**	0.004
Critical Reasoning - diagrammatic	SR	0.116	0.170	**0.324**
CFAT Problem Solving	PS/SR	0.017	-0.052	**0.747**
CFAT Spatial Ability	SR	0.077	-0.011	0.289
CFAT Verbal Skills	VR	-0.055	0.010	0.260

Note. WR - Work Rate; AC - Attentional Capability; PA - Psychomotor Ability; VR - Verbal Reasoning; SR - Spatial Reasoning; PS – Problem Solving. **Bold** denotes factor loadings > .300.

Recall Numbers, the sole subtest from the Attentional Capability domain, did not contribute to the factor structure. The three factors were labeled as Work Rate, Psychomotor Ability, and Problem Solving/Spatial Reasoning (shortened to Reasoning). Regression factor scores were calculated for use in subsequent analyses.

A correlation table for these three factor scores and the RAFAAT Group 2 subtests is shown in Table 7. Although the RAFAAT Group 2 subtests were significantly correlated with the three factors, they did not align well with them. The Mathematics Reasoning and Numerical Operations subtests from the Numerical Reasoning domain had strong, significant correlations with the Reasoning Factor, but Numerical Operations was also strongly correlated with the Work Rate factor. A similar split occurred with the Colours, Letters, and Numbers subtest (Attentional Capability domain). Three of the four Spatial Reasoning subtests were also split between the Reasoning factor and the Work Rate factor, but the fourth, Instrument Comprehension 1, clearly correlated with the Psychomotor Ability Factor. Overall, the three factors identified for the RAFAAT Group 1 subtests did not distinguish clearly among the Group 2 subtests.

Table 7

Correlations between Factor Scores and RAFAAT Group 2 Subtests

	Work Rate	Psycho -Motor	Reason	Math. Reason	Num. Ops	ABD	Dir. & Dist.	Inst. 1	Inst. 2	Vigil.	CLAN	Digit Recog.
Factor 1 Work Rate	1007	.344**	**.456****	.264**	**.435****	**.410****	.276**	.106*	**.456****	**.447****	**.509****	.217**
Factor 2 Psychomotor	1007	1007	**.536****	.291**	.170**	.354**	.324**	**.455****	.361**	.314**	.365**	.033
Factor 3 Reasoning	1007	1007	1007	**.592****	**.423****	**.470****	**.439****	.326**	**.489****	.387**	**.512****	.076
Math. Reason. (NR)	515	515	515	560	**.420****	.377**	.318**	.220**	**.410****	.266**	**.422****	.045
Numerical Ops. (NR)	513	513	513	544	544	.269**	.147**	.046	.308**	.220**	**.432****	.110*
ABD (SR)	513	513	513	557	544	557	.287**	.298**	.397**	.311**	**.411****	.092*
Directions Dist. (SR)	513	513	513	544	544	544	544	.276**	.359**	.224**	.304**	.086*
Inst. Comp. 1 (SR)	513	513	513	544	544	544	544	544	.301**	.124**	.220**	-.057
Inst.Comp. 2 (SR)	513	513	513	544	544	544	544	544	544	.319**	**.440****	.106*
Vigilance (WR)	551	551	551	544	544	544	544	544	544	583	**.434****	.116**
Colours, Letters (AC)	515	515	515	560	544	557	544	544	544	544	560	.130**
Digit Recog. (AC)	513	513	513	544	544	544	544	544	544	544	544	544

Note. * $p < .05$; ** $p < .01$; NR: Numerical Reasoning; SR: Spatial Reasoning; ABD: Angles, Bearings, and Degrees; WR: Work Rate; AC: Attentional Capability. Ability domains for subtests are in the left column. Shaded areas = *n* for factor/subtest; bottom of chart is *n* for individual correlations. ***Bold*** = correlations between subtests of different ability domains> .400.

Canadian Automated Pilot Selection System (CAPSS) Testing. The pilot candidates completed four sessions on the CAPSS simulator over several days as part of the selection process. Table 8 shows the descriptive statistics for CAPSS. The decrease in *n* over the sessions is a result of candidates failing to make the required score in order to proceed to the next session. Correlations between the individual CAPSS sessions, CFAT subtests, and the RAFAAT Group 1 and Group 2 subtests are in Table 9.

Table 8

Descriptive Statistics for Canadian Automated Pilot Selection System (CAPSS)

	N	Min.	Max.	Mean	S.D.
CAPSS Session 1	1026	.01	.99	.718	.181
CAPSS Session 2	1014	.00	.99	.680	.208
CAPSS Session 3	1013	.03	.99	.691	.241
CAPSS Session 4	1011	.02	.99	.640	.282

The correlations between all four CAPSS sessions and the RAFAAT subtests in both the Spatial Reasoning and Psychomotor Ability domains were significant, $p < .01$, as was the Table Reading subtest in the Work Rate domain. The highest correlations were found between the two Psychomotor Ability subtests and CAPSS, all *p*s < .01, and the highest correlation overall was between the Sensory Motor Apparatus subtest and CAPSS Session 3, $r = .506$.

The final analysis completed for research question one focused on the correlations between the three factor scores and the four CAPSS simulator scores, shown in Table 10. Work Rate had a significant correlation only with CAPSS session 4, however the correlation was very weak. Psychomotor Ability had strong, significant correlations with all CAPSS sessions. Reasoning had significant correlations with all CAPSS sessions, however they were weak to moderate in strength

Table 9

Correlations between CAPSS, CFAT, and RAFAAT Group 1 and 2 Subtests

Subtest (*N* for subtest)	Domain	CAPSS 1 (*n* = 1026)	CAPSS 2 (*n* = 1014)	CAPSS 3 (*n* = 1013)	CAPSS 4 (*n* = 1011)
CFAT Verbal Skills (*n*=998)	Verbal Reasoning	-.008	.021	.050	.015
CFAT Problem Solving (*n*=998)	Verbal/Spatial	.024	.060	.152**	.103**
Mathematics Reasoning (*n*=513)	Numerical Reasoning	.063	.091*	.126**	.141**
Numerical Operations (*n*=497)	Numerical Reasoning	-.072	-.079	-.051	-.018
CFAT Spatial Ability (*n*=998)	Spatial Reasoning	.118**	.146**	.156**	.144**
Critical Reasoning (*n*=1009)	Spatial Reasoning	.090**	.127**	.161**	.151**
Angles,Bearings,Degrees (*n*=510)	Spatial Reasoning	.130**	.139**	.222**	.192**
Directions & Distances (*n*=497)	Spatial Reasoning	.129**	.159**	.179**	.187**
Instrument Comp. 1 (*n*=497)	Spatial Reasoning	.287**	.350**	.413**	.338**
Instrument Comp. 2 (*n*=497)	Spatial Reasoning	.135**	.160**	.183**	.174**
Table Reading (*n*=995)	Work Rate	.106**	.132**	.140**	.183**
Visual Search 1– Letters (*n*=995)	Work Rate	-.004	.010	-.006	.033
Visual Search 2– Shapes (*n*=995)	Work Rate	.012	.034	.018	.034
Vigilance (*n*=536)	Work Rate	.043	.051	.097*	.115**
Recall Numbers (*n*=995)	Attentional Capability	-.003	.008	.067*	.053
Colours,Letters,Numbers (*n*=513)	Attentional Capability	.041	.027	.053	.080
Digit Recognition (*n*=497)	Attentional Capability	-.069	-.084	-.095*	-.055
Control of Velocity (*n*=991)	Psychomotor Ability	.158**	.184**	.229**	.158**
Sensory Motor Apparatus (*n*=1005)	Psychomotor Ability	.367**	.439**	.503**	.465**

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

.

Table 10

Correlations between Factor Scores and CAPSS Scores (N for Individual Measures)

Factor (*N*)	CAPSS 1 (*n* = 1026)	CAPSS 2 (*n* = 1014)	CAPSS 3 (*n* = 1013)	CAPSS 4 (*n* = 1011)
Work Rate (*n* = 1007)	.052	.060	.061	.091**
Psychomotor Ability (*n* = 1007)	.373**	**.423****	**.500****	**.447****
Reasoning (*n* = 1007)	.137**	.159**	.252**	.208**

Note. * $p < .05$; ** $p < .01$; Shaded areas = *n* for factor/subtest; bottom of chart is *n* for individual correlations. ***Bold*** = correlations between subtests of different ability domains > .400.

Summary. The relationships between the CFAT and RAFAAT subtests were explored using correlations and factor analysis. In general, higher correlations were observed between subtests in the same ability domain. There were some exceptions however, most noticeably with the Digit Recognition in the Attentional Capability domain, which not only had low correlations with other subtests in that domain, but also with all other subtests. The factor analysis of the CFAT subtests, and RAFAAT Group 1 subtests identified three factors, which in turn, did not correlate distinctively with the RAFAAT Group 2 subtests. The CAPSS simulator scores were significantly correlated with several of the CFAT and RAFAAT subtests, particularly those in the Psychomotor Ability domain. These results were consistent when the CAPSS scores were correlated with the three factors; the strongest correlations were with the Psychomotor Ability factor. Correlations were also significant but more moderate with the Spatial Reasoning subtests, a finding confirmed by the significant but moderate correlations between CAPSS scores and the Reasoning factor.

Successful and Unsuccessful Candidates

Until October 2013, pilot candidates were considered successful at aircrew selection if they passed testing on the Canadian Automated Pilot Selection System or CAPSS. CAPSS exposes candidates to a sample of the flight skills required to fly a single-engine, light aircraft and the CAPSS cut-off mark for selection was a score of .70 on session 4 (Darr, 2010). To facilitate analysis, an aircrew selection

Pass/Fail variable was created in the data set: CAPSS 4 session scores at or above .70 were coded as 1 (pass) and scores that were this score were coded as 0 (fail). Three methods of analysis were used to determine if there were specific demographic variables or aptitude test indicators that defined successful pilot candidates: MANOVA, discriminant analysis, and regression. The MANOVA, using the three factors identified in Table 6 and the three demographic variables ($n = 851$), was significant, Wilks' $\lambda = .831$, F (6, 844) = 28.633, $p < .01$. Significant univariate effects (see Table 11) were obtained for Gender, Factor 2 (Psychomotor Abilities), and Factor 3 (Reasoning).

Table 11

Between-Subjects Effects For Aircrew Pass/Fail on Demographic Variables and Factor Scores ($N = 851$)

Dependent Variable	F	p	ηp^2
Age at Testing	1.498	.221	.002
Gender (male = 1; female = 2)	19.654	≤.001	.023
Education Level	1.896	.196	.002
Factor 1 – Work Rate	2.978	.085	.003
Factor 2 – Psychomotor Abilities	153.782	≤.001	.153
Factor 3 – Reasoning	18.275	≤.001	.021

Note. Education Level: 1 = High school; 2 = College/CEGEP; 3 = University.

A chi-square test was conducted to examine the Gender effect identified in the MANOVA. The results shown in Table 12 were significant, χ^2 (1, $n = 986$) = 23.075, $p < .01$, indicating female candidates experienced greater difficulty passing CAPSS testing (30/104 or 28.8%) than male candidates (474/882 or 53.7%).

Table 12

Chi-Square Gender/CAPSS Pass/Fail – Actual Count (Expected)

	CAPSS Fail	CAPSS Pass
Gender Male (*n* = 882)	408 (431)	474 (451)
Gender Female (*n* = 104)	74 (51)	30 (53)

A discriminant analysis was completed using the same variables and ability factor scores as predictors of group membership; the significance of the function is of course identical to that of the MANOVA. The structure matrix is in Table 13 and the classification results are in Table 14.

Table 13

Structure Matrix for Discriminant Function Analysis (N = 851)

Predictor Variable	Correlation with Function
Factor 2 – Psychomotor Ability	.943
Gender (male = 1; female = 2)	-.337
Factor 3 – Reasoning	.325
Factor 1 – Work Rate	.131
Education Level	.105
Age	-.093

Note. Education Level: 1 = High school; 2 = College/CEGEP; 3 = University.

The structure matrix is defined largely by the Psychomotor Ability factor consistent with the results of the MANOVA. The discriminant analysis correctly classified 68.4% of the candidates. Generally, it was better able to classify candidates who passed CAPSS testing than those who failed.

Table 14

Classification Results for Discriminant Function Analysis: Number of Candidates (Percentage)

Pass/Fail Aircrew Selection: Fail = 0; Pass = 1	Predicted Group Membership		Total
	0	1	
0	254 (62.4%)	153 (37.6%)	407
1	116 (26.1%)	328 (73.9%)	444

The final analysis for research question two was a hierarchical regression using the demographic variables and three factor scores to predict the CAPSS session four (continuous variable) scores. These variables were chosen because the number of candidates was higher (n = 850) than the number of candidates who completed the RAFAAT Group 2 subtests (n = 513). The results are in Table 15. Demographic variables were entered in Step 1, the three factor scores were entered in Step 2, and the candidate scores from CAPSS sessions 1, 2, and 3 were entered in Step 3. The three-step model accounted for 66% of the variance.

In Step 1, all three demographic variables were significant, $p < .01$, however only 6% of variance was accounted for. Step 2 accounted for a further 16.6% of the variance, with Psychomotor Ability being the only significant ability factor; the demographic variables were reduced in magnitude. Step 3 added a further 44.1 % of the variance, with both CAPSS 2 and 3 being significant. Step 3 decreased the magnitude of the demographic variables further, with none being significant. Of the ability factors, only Psychomotor Ability was significant, $p < .05$.

Table 15

Hierarchical Regression Analysis Predicting Canadian Automated Pilot Selection System (CAPSS) Session Four Score (N = 850)

Step	Predictor	*ΔR² (Step)*	*β* Step 1	*β* Step 2	*β* Step 3
1	Age at Testing	.058**	-.108**	-.073*	-.025
	Highest Education Level		.102**	.066	.043
	Gender (M = 1; F = 2)		-.225**	-.090*	.009
2	Work Rate	.166**		-.041	.014
	Psychomotor Ability			.464**	.071*
	Reasoning Ability			-.044	-.022
3	CAPSS Session 1	.441**			-.042
	CAPSS Session 2				.371**
	CAPSS Session 3				.506**
	Total R^2	.665**			

Note. * $p < .05$; ** $p < .01$. Education Level: 1 = High school, 2 = College/CEGEP, 3 = University.

Summary. Candidates who were successful at CAPSS testing had several common abilities/characteristics that separated them from those who were not. The MANOVA identified Gender, Psychomotor Ability, and Reasoning as significant components of strong CAPSS performance. These results were confirmed by the discriminant analysis in which Psychomotor Ability dominated the structure matrix with Gender and Reasoning making moderate contributions. The regression analysis showed that in Step 3 only the Psychomotor Factor remained a significant predictor of CAPSS Pass/Fail.

Patterns of Performance on the Canadian Automated Pilot Selection System (CAPSS)

The third research question concerning patterns of performance in the CAPSS simulator was investigated using latent class analysis (LCA) in *Mplus* (*Mplus* Demo Version 7.2, 2014). Latent class analysis focuses on grouping participants with similar performance patterns across a set of variables (Geiser, 2013). *Mplus* methodology and model fit information criteria are shown in Appendix D. Once the latent classes were identified, they were compared in terms of demographic variables and aptitude test scores. Because of the exploratory nature of this investigation, a variety of solutions were attempted with different numbers of classes.

Two-class model. Results of the LCA Two-Class Model are depicted in Figure 2.

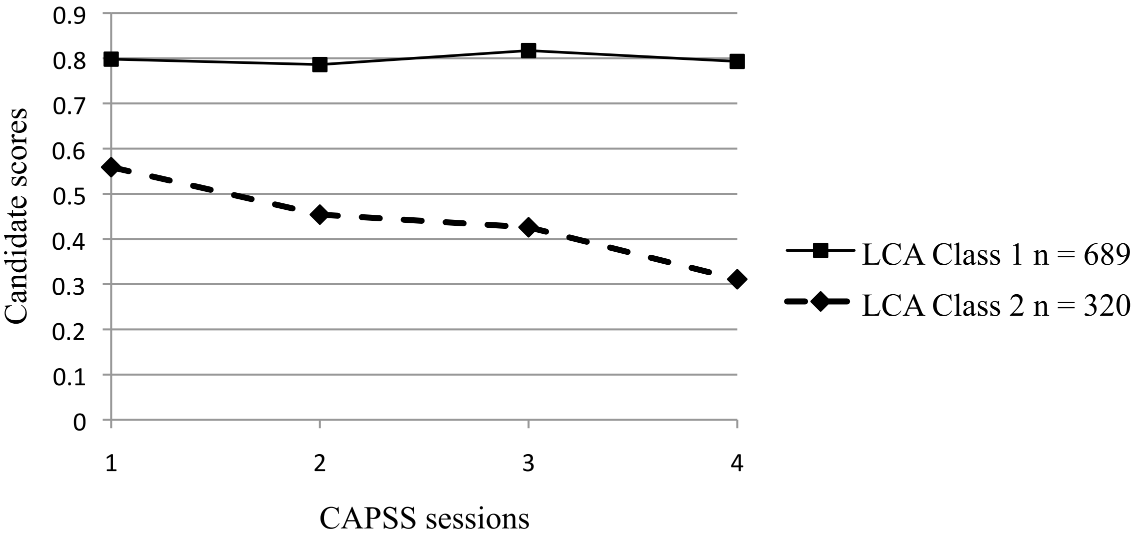

Figure 2. The two-class model for Latent Class Analysis of CAPSS scores.

Specific model fit information criteria for the two-class model are shown in Appendix E. Based on model fit information criteria, entropy scores, and the probability of latent class membership, the two-class model provided the best model fit for the CAPSS data. Members of Class 1 performed well across all sessions, whereas those in Class 2 started with lower scores and continued to decrease. One-way analyses of variance (ANOVA) were used to compare classes on the demographic and aptitude variables (see Table 16). MANOVA was not used because of the differences in *n* across measures. Table 16 and subsequent tables show the *N*'s in each class for the RAFAAT Group 1 subtests and demographics; the *N*'s for each class are approximately half as large for the RAFAAT Group 2 subtests.

Table 16

Summary of Analysis of Variance Results Comparing Latent Class Analysis Two-Class Model on CFAT Subtests, Factor Scores, RAFAAT Subtests, and Demographic Variables

Variable	Domain	Class 1 $n = 689$ Higher scores M (SD)	Class 2 $n = 320$ Lower scores M (SD)	*df*	*F*	ηp^2
CFAT Verbal	VR	10.66 (2.44)	10.54 (2.70)	1, 997	.507	.001
CFAT Spatial	SR	11.86 (2.09)	11.25 (2.33)	1, 997	17.05**	.017
CFAT Problem Solving	VR/SR	24.49 (3.73)	24.01 (3.57)	1, 997	3.620	.004
Factor 1 – Work Rate	WR	.0452 (.906)	-.0589 (.929)	1, 979	2.761	.003
Factor 2 – Psychomotor	PA	.2459 (.741)	-.4620 (.704)	1, 979	199.783**	.170
Factor 3 - Reasoning	VR/SR	.0931 (.794)	-.1784 (.755)	1, 979	25.598**	.026
Mathematics Reasoning	NR	10.16 (3.79)	9.48 (3.71)	1, 512	3.77	.007
Numerical Operations	NR	34.61 (8.41)	35.21 (9.31)	1, 496	.519	.001
Critical Reasoning	SR	7.49 (2.26)	6.86 (2.12)	1 , 1008	17.42**	.017
Angles, Bearings,Degrees	SR	42.81 (5.61)	41.76 (5.76)	1, 509	4.04*	.008
Directions and Distances	SR	8.57 (2.56)	7.74 (2.57)	1, 496	11.79**	.023
Instrument Comp. 1	SR	12.71 (3.84)	9.56 (3.50)	1, 496	81.04**	.141
Instrument Comp. 2	SR	12.72 (3.00)	11.81 (2.86)	1, 496	10.767**	.024
Table Reading	WR	61.32 (10.01)	58.54 (9.48)	1, 994	17.11**	.017
Visual Search 1	WR	56.92 (5.70)	56.85 (6.17)	1, 994	.03	.000
Visual Search 2	WR	55.66 (6.22)	55.52 (5.92)	1, 994	.11	.000
Vigilance	WR	146.86 (27.06)	143.32 (30.25)	1, 535	1.91	.004
Recall Numbers	AC	98.02 (12.80)	96.87 (13.85)	1, 994	1.65	.002
Colours, Letters,Numbers	AC	126.37 (230.82)	108.29 (204.90)	1, 512	.78	.002
Digit Recognition	AC	9.64 (1.90)	10.02 (2.23)	1, 496	4.17*	.004
Control of Velocity	PA	105.71 (13.95)	99.31 (17.08)	1, 990	38.96**	.038
Sensory Motor Apparatus	PA	194.63 (39.99)	155.52 (36.80)	1, 1004	218.83**	.179
Gender	--	1.06 (.23)	1.21 (.41)	1, 985	52.47**	.051
Age at Testing	--	22.32 (4.85)	22.76 (5.25)	1, 877	1.50	.002
Education Level	--	1.83 (.93)	1.81 (.92)	1, 900	.10	.000

Note: * $p < .05$; ** $p < .01$; VR: Verbal Reasoning; SR: Spatial Reasoning; NR: Numerical Reasoning; WR: Work Rate; AC Attentional Capability; PA: Psychomotor Ability. Gender: Male = 1, Female = 2; Education: 1 = High school, 2 = CEGEP/College, 3 = University/Graduate school.

Members of Class 1 performed significantly better than those in Class 2 on the CFAT Spatial Ability subtest and had higher factor scores on Factor 2 Psychomotor Ability and Factor 3 Reasoning. The Class 1 candidates also had better scores on all five RAFAAT Spatial Reasoning subtests, the Table Reading subtest from the Work Rate domain, and both Psychomotor Ability subtests. Class 1 members were also more likely to be male. The less successful candidates (Class 2) scored higher on Digit Recognition, but this was only at the .05 level.

A chi-square test was conducted to examine the Gender effect with class membership in the two-class LCA. The significant Pearson chi-square value = 49.919, *p* < .001 indicated that males and females were not evenly distributed across classes. As shown in Table 17, 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).

Table 17

Chi-Square Analysis of LCA Two-Class Model by Gender; Actual Count (Expected in Parentheses) and Percent of Each Gender

	LCA Class 1 (High scores)	LCA Class 2 (Low scores)
	n = 689	*n* = 320
Gender Male	632 (600)	250 (282)
	71.7%	28.3%
Gender Female	39 (78)	65 (33)
	37.5%	63.5%

Several of the variables that distinguished Class 1 from Class 2 in the LCA were the same as those that distinguished successful from unsuccessful candidates in research question two. The RAFAAT Psychomotor Ability subtests, which comprised Factor 2, showed the largest differences between Class 1 and Class 2 candidates as it did in the MANOVA (Table 11) and the Discriminant Analysis (Table 13) in classifying successful and unsuccessful candidates. The Reasoning Factor (CFAT Problem Solving and

RAFAAT Critical Thinking subtests) was also a predictor of CAPSS success in the MANOVA, although it made only a moderate contribution to the structure matrix of the discriminant analysis. Gender was also significant; female candidates were overrepresented in the low scoring Class 2 and had greater difficulty passing CAPSS.

Three-Class Model. Results of the LCA Three-Class Model are depicted in Figure 3.

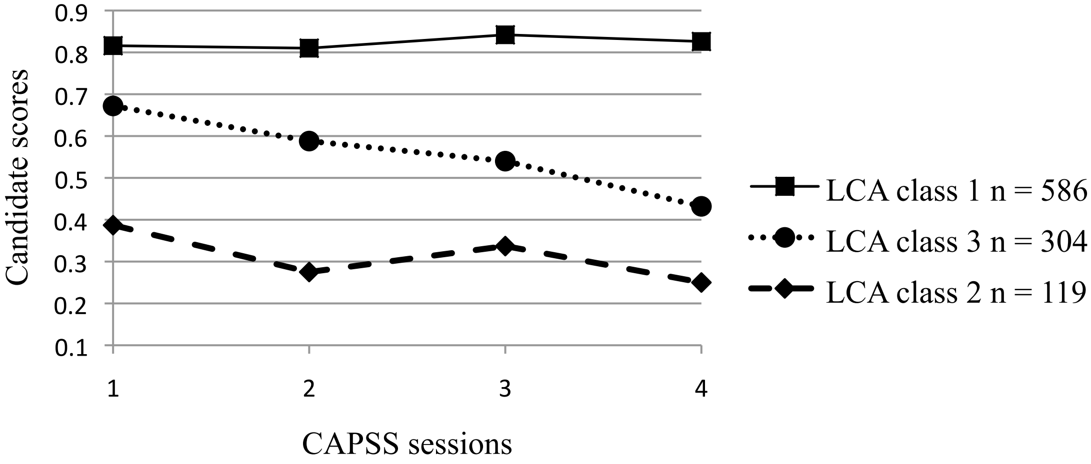

Figure 3. Latent Class Analysis three-class model of CAPSS scores.

One-way ANOVAs with follow-up Bonferroni t-tests were used to compare classes on the demographic and aptitude variables (Table 18). Members of Class 1 performed significantly better than those in Classes 2 and 3 on the CFAT Spatial Ability and Problem Solving subtests and had higher factor scores on the Psychomotor Ability and Reasoning factors. The Class 1 candidates also had significantly higher scores on all of the five Spatial Reasoning subtests, the Table Reading subtest, and both Psychomotor Ability subtests. Class 3, the candidates who started with high CAPSS scores but dropped, had higher scores on two of the five spatial reasoning subtests and both psychomotor ability subtests than the low scoring Class 2 candidates. However, they also had significantly lower scores on the CFAT Problem Solving subtest. Gender was significant, F (2, 985) = 25.48, $p < .01$ indicating that Class 1 candidates were more likely to be male.

Table 18

Summary of Analysis of Variance Results Comparing Latent Class Analysis Three-Class Model on CFAT Subtests, Factor Scores, RAFAAT Subtests, and Demographic Variables

Variable (Domain)	Class 1 $n = 586$ Higher scores M (SD)	Class 3 $n = 304$ High-low M (SD)	Class 2 $n = 119$ Lower scores M (SD)	*df*	*F*	ηp^2	Post hoc comparisons
CFAT Verbal (VR)	10.67 (2.45)	10.57 (2.56)	10.55 (2.73)	1, 997	.192	.000	
CFAT Spatial (SR)	11.93 (2.09)	11.37 (2.24)	11.16 (2.34)	1, 997	10.33**	.020	1 > 3 = 2
CFAT Problem Solving (VR/SR)	24.67 (3.65)	23.73 (3.76)	24.28 (3.44)	1, 997	6.51**	.013	1 > 3
Factor 1 – Work Rate (WR)	.0587 (.892)	-.0569 (.936)	-.0421 (.959)	1, 979	1.78	.004	
Factor 2 – Psychomotor Ability (PA)	.3136 (.731)	-.2985 (.704)	-.6050 (.680)	1, 979	121.38**	.199	1 > 3 > 2
Factor 3 – Reasoning (VR/SR)	.1412 (.778)	-.1892 (.796)	-.1590 (.719)	1, 979	20.51**	.040	1 > 3 = 2
Mathematical Reasoning (NR)	10.09 (3.64)	9.78 (4.11)	9.51 (3.39)	2, 512	.79	.003	
Numerical Reasoning (NR)	34.59 (8.08)	34.14 (9.53)	37.61 (9.01)	2, 496	3.79	.015	
Critical Reasoning (SR)	7.57 (2.20)	6.96 (2.23)	6.76 (2.24)	2, 1008	11.45**	.022	1 > 3 = 2
Angles, Bearings, Degrees (SR)	43.15 (5.56)	41.86 (5.41)	40.78 (6.44)	2, 509	5.90*	.023	1 > 3 = 2
Directions and Distances (SR)	8.61 (2.56)	7.97 (2.66)	7.58 (2.57)	2, 496	5.58**	.022	1 > 3 = 2
Instrument Comprehension 1 (SR)	12.88 (3.84)	10.16 (3.78)	9.60 (3.25)	2, 496	36.66**	.129	1 > 3 = 2
Instrument Comprehension 2 (SR)	12.76 (2.97)	12.09 (2.96)	11.61 (2.87)	2, 496	5.10**	.020	1 > 3 = 2
Table Reading (WR)	61.59 (9.95)	58.81 (9.76)	58.89 (9.48)	2, 994	9.48**	.019	1 > 3 = 2
Visual Search 1 – Letters (WR)	56.95 (5.64)	56.71 (6.02)	57.13 (6.41)	2, 994	.27	.001	
Visual Search 2 – Shapes (WR)	55.70 (6.18)	55.67 (6.08)	55.08 (6.00)	2, 994	.52	.001	
Vigilance (WR)	146.73 (21.18)	145.20 (27.82)	141.84 (33.29)	2, 535	.87	.003	
Recall Numbers (AC)	97.98 (12.94)	96.71 (13.14)	98.42 (14.10)	2, 994	1.15	.002	
Colours, Letters, and Numbers (AC)	124.91 (221.94)	122.32 (204.39)	91.44 (250.58)	2, 512	.60	.002	
Digit Recognition (AC)	9.59 (1.91)	9.92 (2.08)	10.19 (2.32)	2, 496	2.84	.011	
Control of Velocity (PA)	105.98 (14.05)	101.28 (16.13)	98.49 (16.85)	2, 990	17.53**	.034	1 > 3 > 2
Sensory Motor Apparatus (PA)	198.47 (39.49)	164.79 (36.42)	146.67 (36.25)	2, 1004	135.66**	.213	1 > 3 > 2
Gender	1.05 (.22)	1.16 (.37)	1.24 (.43)	2, 985	25.48**	.049	1 < 3 < 2
Age	22.18 (4.75)	22.83 (5.28)	22.92 (5.26)	2, 877	1.98	.005	
Education	1.80 (.93)	1.85 (.93)	1.82 (.93)	2, 900	.24	.001	

Note: * $p < .05$; ** $p < .01$; VR: Verbal Reasoning; SR: Spatial Reasoning; NR: Numerical Reasoning; WR: Work Rate; AC Attentional Capability; PA: Psychomotor Ability. Gender: Male= 1, Female= 2; Education: 1=High school, 2= CEGEP/College, 3= University/Grad school.

A chi-square test was conducted to examine the Gender effect with class membership in the three-class LCA. The significant Pearson chi-square value = 48.946, $p < .001$ indicated that males and females were again not evenly distributed across classes. As shown in Table 19, 61.5% of male candidates were in Class 1 (high CAPSS scores) compared to 27.9% of females. The opposite pattern was shown in Class 2 (low CAPSS scores). More than forty-five percent of female candidates were found in Class 3 (whose CAPSS scores started high then decreased) compared to 28.3% of male candidates.

Table 19

Chi-Square LCA three classes: Gender by Class membership – Actual count (expected) and percent of each Gender

	Class 1 (*n* = 586) Higher scores	Class 3 (*n* = 304) High to low scores	Class 2 (*n* = 119) Lower scores
Male *n* = 882	542 (511)	250 (266)	90 (105)
	61.5%	28.3%	10.2%
Female *n* = 104	29 (60)	47 (32)	28 (12)
	27.9%	45.2%	26.9%

Four-class model. The final Latent Class Analysis completed in *Mplus* was a four-class model shown in Figure 4.

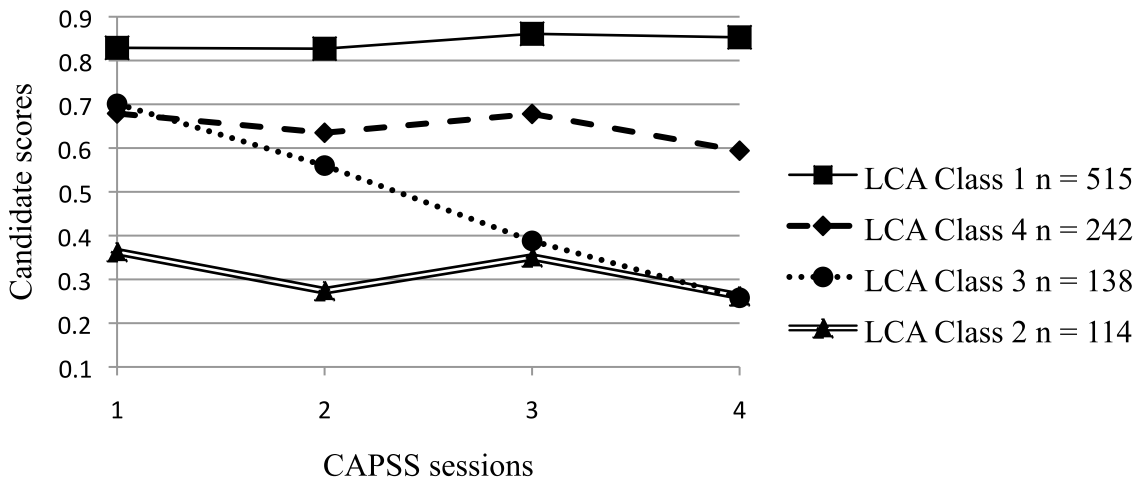

Figure 4. Latent Class Analysis four-class model of CAPSS scores.

This model contained a class of candidates (Class 4) who started with average CAPSS scores and maintained those scores throughout testing. One-way ANOVAs revealed significant main effects for the CFAT Spatial Ability and Problem Solving subtests, the Psychomotor Ability and Reasoning factors, all five Spatial Reasoning subtests, Table Reading from the Work Rate domain, both Psychomotor Ability subtests, and Gender as shown in Table 20. The Recall Numbers subtest from the Attentional Capability domain was also significant, $p < .05$, but, surprisingly, it was the low scoring Class 2 candidates who had the highest mean scores. Class 2 candidates also had the second highest scores on the CFAT Problem Solving subtests, outscoring the Class 3 candidates (high to low scores) and those in Class 4 (medium scores).

Bonferroni post hoc testing revealed that Class 1, the high scoring class, was significantly different from the other three classes on seven of the eleven statistically significant subtests while the Class 4 candidates (who maintained moderate CAPSS scores) had significantly higher scores than Class 3 (high changing to low CAPSS scores) on two of the eleven statistically significant subtests and the two factor scores. Gender was also significant for Class 4, with more male candidates than Class 3 (high to low scores) and Class 2 (low scores).

A chi-square test of independence revealed a statistically significant Pearson chi-square value = 48.946, $p < .001$, indicating that, once again, males and females were not evenly distributed across classes. As shown in Table 21, 53.9% of male candidates were in Class 1 (high CAPSS scores), whereas only 24% of females were. The opposite pattern presented in Class 2 (low CAPSS scores). Class 3, whose members started with high scores but decreased rapidly, contained approximately 28% percent of female candidates compared to 12% of male candidates. In Class 4, the numbers of male and female candidates were as expected, with only a slightly higher percentage of male candidates than female candidates.

Table 20

Summary of Analysis of Variance Results Comparing Latent Class Analysis Four-Class Model on CFAT Subtests, Factor Scores, RAFAAT Subtests, and Demographic Variables

Variable (Domain)	Class 1 *n* = 515 High scores *M* (SD)	Class 4 *n* = 242 Medium *M* (SD)	Class 3 *n* = 138 High-Low *M* (SD)	Class 2 *n* = 114 Low scores *M* (SD)	*df*	*F*	ηp^2	Post hoc testing
CFAT Verbal (VR)	10.68 (2.43)	10.71 (2.41)	10.29 (2.84)	10.61 (2.72)	1, 997	.97	.003	
CFAT Spatial (SR)	12.00 (2.09)	11.39 (2.15)	11.30 (2.33)	11.21 (2.32)	1, 997	8.26**	.024	1>4=3=2
CFAT Problem Solving (VR/SR)	24.69 (3.67)	24.15 (3.67)	23.27 (3.75)	24.42 (3.43)	1, 997	5.73**	.017	1 > 3
Factor 1 – Work Rate (WR)	.0656 (.889)	.0155 (.959)	-.1513 (.870)	-.0386 (.962)	1, 979	2.11	.006	
Factor 2 – Psychomotor (PA)	.3529 (.729)	-.0871 (.702)	-.4995 (.655)	-.6094 (.682)	1, 979	92.43**	.221	1>4>3=2
Factor 3 – Reasoning (VR/SR)	.1631 (.778)	-.0699 (.785)	-.3250 (.784)	-.1339 (.712)	1, 979	17.00**	.050	1>4>3=2
Mathematical Reasoning (NR)	10.31 (3.71)	10.02 (3.97)	8.89 (3.77)	9.52 (3.36)	3, 512	3.13*	.018	1=4>3=2
Numerical Reasoning (NR)	34.80 (8.17)	34.40 (9.35)	33.55 (9.02)	37.36 (8.85)	3, 496	2.37	.014	
Critical Reasoning (SR)	7.62 (2.21)	7.12 (2.35)	6.79 (1.93)	6.75 (2.23)	3, 1008	42.82**	.025	1>4=3=2
Angles, Bearings, Degrees (SR)	43.53 (5.46)	41.85 (5.55)	41.24 (5.44)	40.98 (6.37)	3, 509	6.13**	.035	1>4=3=2
Directions and Distances (SR)	8.76 (2.56)	8.06 (2.66)	7.74 (2.57)	7.54 (2.47)	3, 496	5.88**	.035	1>4=3=2
Instrument Comp. 1 (SR)	13.09 (3.87)	10.79 (3.78)	10.03 (3.75)	9.51 (3.17)	3, 496	25.05**	.132	1>4=3=2
Instrument Comp. 2 (SR)	12.94 (2.89)	12.06 (3.18)	11.96 (2.74)	11.57 (2.85)	3, 496	5.35**	.032	1>4=3=2
Table Reading (WR)	62.04 (9.95)	59.51 (9.77)	57.41 (9.32)	58.81 (9.59)	3, 994	10.61**	.031	1>4=3=2
Visual Search 1 – Letters (WR)	56.92 (5.63)	56.98 (5.93)	56.47 (6.02)	57.13 (6.45)	3, 994	.323	.001	
Visual Search 2 – Shapes (WR)	55.69 (6.23)	55.92 (6.32)	55.19 (5.49)	55.14 (5.99)	3, 994	.67	.002	
Vigilance (WR)	147.58 (27.01)	146.92 (25.66)	140.90 (31.06)	141.16 (33.26)	3, 535	1.80	.010	
Recall Numbers (AC)	97.93 (12.61)	98.22 (13.25)	94.70 (13.86)	98.78 (14.07)	3, 994	2.79*	.008	1=4=3=2
Colours, Letters, Numbers (AC)	129.81 (230.89)	133.59 (216.14)	84.15 (169.46)	98.37 (254.14)	3, 512	1.21	.007	
Digit Recognition (AC)	9.62 (1.88)	9.90 (2.12)	9.66 (2.03)	10.25 (2.32)	3, 496	1.80	.011	
Control of Velocity (PA)	105.96 (14.22)	104.05 (14.69)	99.46 (16.71)	97.76 (16.92)	3, 990	13.36**	.039	1=4>3=2
Sensory Motor Apparatus (PA)	200.51 (39.49)	175.65 (36.12)	154.62 (35.48)	146.91 (36.17)	3, 1004	99.95**	.230	1>4>3=2
Gender	1.05 (.22)	1.09 (.29)	1.21 (.41)	1.25 (.43)	3, 985	20.35**	.059	1=4<3=2
Age	22.11 (4.56)	22.53 (5.43)	23.42 (5.37)	22.74 (2.43)	3, 877	2.31	.008	
Education	1.82 (.93)	1.79 (.91)	1.90 (.94)	1.81 (.93)	3, 900	.38	.001	

Note: * $p < .05$; ** $p < .01$; VR: Verbal Reasoning; SR: Spatial Reasoning; NR: Numerical Reasoning; WR: Work Rate; AC Attentional Capability; PA: Psychomotor Ability. Gender: Male = 1, Female = 2; Education: 1 = High school, 2 = CEGEP/College, 3 = University/Graduate school.

Table 21

Chi-Square Analysis of LCA Four Class Model by Gender; Actual Count (Expected in parentheses) and Percent of Gender

	Class 1 (*n* = 515) High scores	Class 4 (*n* = 242) Medium scores	Class 3 (*n* = 138) High to low	Class 2 (*n* = 114) Low scores
Gender Male	475 (447)	215 (212)	107 (121)	85 (101)
	53.9%	24.4%	12.1%	9.6%
Gender Female	25 (53)	22 (25)	29 (14)	28 (12)
	24%	21.2%	27.9%	26.9%

Summary. Overall, the two-class model provided the best model fit for the CAPSS testing data but there were patterns evident in the performance of the candidates in the high and low scoring classes in all three latent class analyses. Those with high CAPSS scores were predominantly male, and the candidates with higher scores on the Spatial Reasoning and Psychomotor Ability subtests, and the Psychomotor and Reasoning factors. The opposite was true for the low scoring candidates.

In the three-class LCA, a class of candidates emerged who started well but whose scores dropped steadily. These Class 3 candidates had the lowest mean scores of all four classes on half of the RAFAAT subtests, the lowest factor score on the Reasoning factor, and contained a larger percentage of female than male candidates. Class 4 in the four-class LCA (candidates who maintained medium scores throughout testing) were roughly even for the percentage of male and female candidates but had higher mean scores on all 16 RAFAAT subtests and the Psychomotor and Reasoning factors than their Class 3 counterparts who started with high scores then dropped.

Summary of Results

The relationships amongst the ability measures (CFAT and RAFAAT) were analysed using correlations. Contrary to expectations, there were low but significant intra-domain correlations between RAFAAT subtests (Digit Recognition and Colours, Letters, and Numbers in the Attentional Capability domain) as well as high inter-domain correlations (CFAT Problem Solving and RAFAAT Mathematics

Reasoning). Factor Analysis was used to assess the relationships between the demographic variables, CFAT subtests, and RAFAAT Group 1 subtests. The analysis identified three factors, which were significant in a number of analyses. These results are summarised in Table 22.

Table 22

Summary of Results: Levels of Significance for Factor Scores and Gender

	Ability Domain	CAPSS Scores[1] (Significant correlations)	MANOVA[2]	Discrim. Analysis[3]	Hierarchical Regression[4]		
					Step 1	Step 2	Step 3
Work Rate (RAFAAT Table Reading, Visual Search)	WR			.131	N/A		
Psychomotor Ability (Sensory Motor /Control of Velocity	PA	**	**	.943	N/A	**	*
Reasoning (RAFAAT Critical Reasoning and CFAT Problem Solving)	VR/SR	**	**	.325	N/A		
Gender	N/A	N/A	**	-.337	**	*	

Note. [1] Table 10; [2] Table 11; [3] Table 13; Discriminant analysis column contains structure matrix outcomes; [4] Table 15. N/A indicated not applicable or not done.

Correlations between the CAPSS scores and factor scores showed that both Psychomotor Ability and Factor 3 Reasoning were significant, $p < .01$, however the correlations between the CAPSS scores and Psychomotor Ability were much stronger than those for Reasoning. The Table Reading subtest in the Work Rate ability domain was significantly correlated with all four CAPSS session scores and, although it was one of the subtests in Work Rate, neither of the other subtests was significant nor was the Work Rate factor.

MANOVA, discriminant analysis, and hierarchical regression were used in the second research question to determine if there were specific demographic variables or aptitude test indicators that defined successful pilot candidates. The results, summarised in Table 22, identified Psychomotor Ability in the MANOVA and the regression analysis as the main predictor of successful completion of CAPSS. Psychomotor abilities also defined the discriminant analysis structure matrix, with only small contributions from Gender and Reasoning.

Research question three identified different subgroups within the data set. In general these groups corresponded to more or less successful candidates, with intermediate groups being added in the three and four class analyses. The only group that differed from this pattern was Class 3 in the four-class solution, which started with high scores and then declined. The groups differed on many of the predictor measures. Table 23 is a summary of the *p* values for these significant subtests, factors, and the demographic variable Gender that were associated with overall higher performance on CAPSS.

Candidates with higher scores on the CFAT Spatial Ability and Problem Solving subtests and on the Spatial Reasoning and Psychomotor Ability subtests were more likely to pass CAPSS testing as were those who did well on the Table Reading subtest in the Work Rate domain. Gender was also a significant factor in the candidates' CAPSS performance. Female candidates obtained lower scores on the CFAT and RAFAAT subtests and were overrepresented in the lower scoring classes of CAPSS performance in each LCA, and in the classes that started with high CAPSS scores but dropped over the course of testing.

Table 23

Summary of Research Question Three Results: Levels of Significance for Statistically Significant Subtests and Gender for Mplus Latent Class Analyses

	Ability Domain	LCA 2	LCA 3	LCA 4
CFAT Spatial Ability	SR	**	**	**
CFAT Problem Solving	VR/SR		**	**
Factor 2 – Psychomotor	PA	**	**	**
Factor 3 – Reasoning	VR/SR	**	**	**
Critical Reasoning	SR	**	**	**
Angles, Bearings, and Degrees	SR	*	*	**
Directions and Distances	SR	**	**	**
Instrument Comprehension 1	SR	**	**	**
Instrument Comprehension 2	SR	**	**	**
Table Reading	WR	**	**	
Sensory Motor Apparatus	PA	**	**	**
Control of Velocity	PA	**	**	**
Gender	---	**	**	**

Note. * $p < .05$; ** $p < .01$. VR: Verbal Reasoning; SR: Spatial Reasoning; PA: Psychomotor Ability

Chapter 5

Discussion

The goal of this thesis was to examine the specific cognitive abilities and demographic characteristics that are markers for success of Canadian Forces pilot candidates in aircrew selection. The first research question examined the relationships amongst the test batteries used in pilot selection: the Canadian Forces Aptitude Test (CFAT) which is administered to all Canadian Forces members regardless of occupation; the Royal Air Force Aircrew Aptitude Test (RAFAAT) administered solely to pilot candidates; and the Canadian Automated Pilot Selection System (CAPSS), a single-engine aircraft flight simulator. The second research question focused on the specific demographic variables and aptitude test indicators that differentiated successful candidates from unsuccessful candidates. The third, and final, research question addressed the patterns of performance evident in CAPSS flight simulator testing.

In the remainder of this chapter, each of these research questions is addressed in turn. The implications of these findings are reviewed, followed by an overview of the limitations encountered during this research, and recommendations for future research directions in abilities testing for military pilot candidates.

Relationships Amongst of the Measures

Examining the relationships amongst the test batteries was an important first step in this research as it showed how the subtests were statistically associated with each other and facilitated the identification of common underlying factors that were used in subsequent analyses. The relationships amongst the subtests of the CFAT and RAFAAT test batteries yielded both expected and unexpected results. The subtests are grouped into ability domains developed by the Royal Air Force (RAF) and are a broad collection of similar aptitudes (Bailey, 1999). It was therefore expected that subtests found within the same ability domain would correlate well with each other and form factors that were consistent with C-H-C theory (e.g. McGrew, 2009). The results confirmed this expectation, however, all correlations were weak to moderate. One of the highest correlations found amongst the subtests was an inter-domain correlation between the RAFAAT Mathematics Reasoning subtest and the CFAT Problem Solving subtest. While both subtests assess numerical reasoning abilities, the Mathematics Reasoning subtest is focused on solving aircraft related problems, whereas the CFAT Problem Solving subtest contains more generic number-based problems that are verbal and spatial in nature. This diverse content may also explain why it had statistically significant correlations with every other subtest except Digit Recognition. Although the Problem Solving subtest was grouped in the Spatial Reasoning domain for this study, the diverse nature of its questions suggest that it could also be placed in either the Numerical Reasoning or Verbal Reasoning ability domains.

One of the weakest correlations was found between two Attentional Capability domain subtests: Digit Recognition and Colours, Letters, and Numbers. The Digit Recognition subtest is described by the RAF as a test of working memory (WM) whereas Colours, Letters, and Numbers is considered to be a divided attention task. This disparity in subtest content may explain the weak correlation and suggests that Digit Recognition may be testing abilities similar to those assessed by the subtests in the Work Rate domain, as it had significant correlations with all four subtests in that ability domain.

Correlations between the CAPSS scores and the CFAT Spatial Ability and Problem Solving subtests as well as those between CAPSS and the Spatial Reasoning subtests of the RAFAAT batteries were statistically significant, albeit weak to moderate. Stronger correlations were found between CAPSS and the two Psychomotor Abilities subtests. Unexpectedly, the Table Reading subtest (the Work Rate domain), which is considered to be a clerical-type task, had significant correlations with all four CAPSS sessions. Subtests in the Work Rate domain are described as assessing the ability to work accurately through simple routine tasks under time constraints. This ability aligns well with Gs, the cognitive processing speed Stratum II broad ability in the C-H-C model (McGrew, 2009) and is an ability that the Work Rate domain shares with simulator testing. This overlap may account for some of the similarity in the abilities being tested, however, the Table Reading subtest does not assess any of the other C-H-C abilities identified as components of simulator testing, i.e., Gt (decision and reaction time), Gv (visual spatial abilities), and Gp (psychomotor abilities), leaving the reason for the relationship between these two very different aptitude tests largely unexplained. The other subtests in the Work rate domain, the Attentional Capability subtests, and the Numerical Reasoning subtests had very weak correlations with the CAPSS scores.

Factor analysis. The factor analysis of the CFAT and RAFAAT Group 1 subtests identified three clear factors with a simple structure. These three factors, identified as Work Rate, Psychomotor Ability, and Reasoning, correspond roughly to the Stratum II broad abilities Gs (cognitive processing speed), Gp (psychomotor abilities), and a combined Gv (visual-spatial abilities)/Gt (decision and reaction time) respectively, and accounted for slightly better than half the variance in the aptitude test scores. The Reasoning factor was defined by the CFAT Problem Solving and RAFAAT Critical Thinking subtests which cover a wide range of reasoning abilities. This may explain why it had some of the largest, statistically significant correlations with a number of ability domains.

Only Recall Numbers, the sole subtest from the RAFAAT Attentional Capability domain, did not group in any of the factors. The Attentional Capability domain assesses candidates on information retention, how they deal with multiple tasks simultaneously, and their attention switching capability (Southcote, 2004). The Recall Numbers subtest measures short-term memory or information retention only so its limited scope may account for its poor fit within the factor analysis. The Recall Numbers subtest had moderate, statistically significant correlations with all four Work Rate domain subtests suggesting that it may be a test of candidates’ ability to work accurately through a routine task rather than an assessment of information retention.

The RAFAAT Group 2 subtests had significant correlations with the three identified factors, however they did not load cleanly like the Group 1 subtests and several subtests had strong correlations with more than one factor. For example, Numerical Operations from the Numerical Reasoning domain was strongly correlated with the Reasoning factor but also strongly with the Work Rate factor. Three of the four Spatial Reasoning subtests were also split between the Reasoning and Work Rate factors, and the fourth, Instrument Comprehension 1, was strongly correlated with the Psychomotor Ability factor. This suggests that a reassessment of the abilities that are tested by the Group 2 subtests, particularly those of the Work Rate and Attentional Capability domains, may provide a more accurate assessment of candidates' abilities.

Early simulators, like those that predated CAPSS, were designed to test candidates' psychomotor abilities (Macedonia, 2002) so it was not surprising that the current study found the Psychomotor Ability factor had strong significant correlations with all four CAPSS simulator sessions. The CAPSS scores were also moderately correlated with the Spatial Reasoning subtests, including the CFAT Spatial Ability and RAFAAT Critical Reasoning subtests that were part of the Reasoning factor. These correlations also hint at the involvement of problem solving, cognitive processing speed (Gs in the C-H-C model), decision making (Gt), and visual spatial abilities (Gv) which Grimm and Wilkomm (1996) found were measured in more complex simulations.

In summary, the correlations between the CFAT and RAFAAT indicated that, generally, higher correlations were observed between subtests in the same ability domain. The factor analysis of the CFAT subtests and RAFAAT Group 1 subtests identified three distinct factors, however the Group 2 subtests did not align well with the three factors, indicating a revision of the domain and subtest content may be necessary in order provide a better assessment of candidate abilities. Not surprisingly, the CAPSS simulator scores were significantly correlated with the Psychomotor Ability and Spatial Reasoning subtests however, there were unexplained correlations with the Table Reading from the Work Rate domain subtest as well. The significant correlations of the CAPSS scores with the Reasoning factor underscored the role of problem solving and critical thinking in simulator testing. These results reinforce the requirement to select pilot candidates who demonstrate aptitude in a wide range of abilities and not only those traditionally associated with pilot selection: psychomotor ability and spatial reasoning.

Successful and Unsuccessful Candidates

The second research question examined the ability of the CFAT and RAFAAT subtests to distinguish successful pilot candidates from unsuccessful candidates. When the data used in this research were collected, success at aircrew selection was based on candidate scores on CAPSS. Analysis of demographic information and the three factors identified several commonalities in successful candidates.

The MANOVA identified the Psychomotor Ability factor, the Reasoning factor, and Gender as having significant effects on whether the candidates passed or failed aircrew selection testing. Male candidates and those who had high scores on the CFAT Spatial Abilities subtests and the RAFAAT Spatial Reasoning and Psychomotor Abilities subtests did well on CAPSS. The Work Rate factor, candidate age, and Education Level were not significant. The discriminant analysis indicated that the dimension distinguishing between successful and unsuccessful candidates was largely defined by Psychomotor Ability with only small contributions from Gender and the Reasoning Ability factor. Psychomotor ability has consistently been identified as a key component in pilot performance (Darr, 2010a; Carretta & Ree, 1997a; Olson et al., 2010) and its significance in the current research supports the findings of Chaiken et al. (2000) who concluded that individuals with high psychomotor abilities learned faster, and that cognitively able individuals tended to do very well on psychomotor tests.

The first two steps of the hierarchical regression analysis, using the same variables as the MANOVA and discriminant analysis, accounted for almost a quarter of the variance in CAPSS session 4 scores. The three demographic variables initially were significant but, when the Psychomotor Ability factor was added, the variance accounted for quadrupled and the significance of the demographic variables was reduced dramatically. In the final step, when the CAPSS scores were added to the regression, a further 45% of variance was accounted for, however only the Psychomotor Ability factor from the earlier two steps remained significant.

Gender was a significant factor in identifying successful candidates, with female candidates experiencing more difficulty passing CAPSS testing, however its significance varied amongst the three analyses. The MANOVA identified Gender as having a significant effect on passing CAPSS testing and the discriminant analysis confirmed its role in classifying successful candidates. However, in the hierarchical regression, Gender was a strong predictor only before the ability variables were entered. Its significance decreased when the three factor scores were entered, and when the CAPSS scores from the first three sessions were entered in the final step, Gender was eliminated as a predictor. These results indicate that much of the Gender variance is shared with the ability tests and relatively little of the variance is due to gender alone. These findings are consistent with those of Burke (1995) who observed large differences ($d \geq 0.5$) favouring males on both spatial and psychomotor ability tests.

Overall, high scores on the subtests in the Spatial Reasoning and Psychomotor Abilities domains were the best predictors of success on CAPSS testing. Discriminant analysis confirmed that psychomotor ability was the major characteristic of successful candidates and that the Reasoning factor and Gender were found to be only moderate predictors of success in CAPSS testing.

Patterns of Performance on the Canadian Automated Pilot Selection System (CAPSS).

The third research question focused on whether patterns of performance in the CAPSS simulator would identify homogeneous sub-groups within the larger sample that constituted meaningful groups or classes of individuals. This analysis was instrumental in identifying groups of candidates whose performance differed from those who scored consistently high or consistently low on the four CAPSS sessions. Assessment of CAPSS performance also confirmed the findings of previous analyses in differentiating between successful and unsuccessful candidates.

In the two, three and four-class models of Latent Class Analysis (LCA), members of the class with the highest CAPSS scores in each model were predominately male candidates and those who had scored well on the CFAT Spatial Ability and Problem Solving subtests and had high factor scores on the Psychomotor Ability and Reasoning factors. The high scoring CAPSS candidates also did well on the RAFAAT Spatial Reasoning subtests. Conversely, the lowest scoring class in each of the models had a higher than expected number of female candidates and contained the candidates who had low scores on the aforementioned ability tests and factors. The two-class model containing a high scoring group and a low scoring group provided the best model fit for the CAPSS data however, the three and four-class models showed distinct groups that did not follow the performance patterns of either the top group or the bottom group.

Of particular interest were the Class 4 candidates in the four-class LCA model. These 242 candidates had CAPSS session 1 scores of just below .70 (the pass mark needed on CAPSS 4) and fluctuated only slightly over the following CAPSS sessions to finish testing with scores near .60. More importantly however, the Class 4 candidate scores on the CFAT and RAFAAT subtests were consistently higher than the either the candidates in Class 3 (who started with high CAPSS scores then dropped precipitously) or the Class 2 candidates (who had low CAPSS scores throughout). It is possible that Class 4 candidates may have passed CAPSS testing if given one more session in the simulator. They certainly would have passed pilot selection on the basis of their RAFAAT scores. Unfortunately, because they failed CAPSS testing, these candidates were not selected for pilot training and the Air Force missed the opportunity to train a number of pilot candidates whose high subtest scores in a number of ability domains may have enabled them to successfully complete pilot training.

Implications for Pilot Selection

The present results indicate that candidates who were successful at aircrew selection possessed a number of common abilities. The Psychomotor Ability factor was a significant predictor of the pilot candidates' ability to pass CAPSS testing and dominated the discriminant analysis structure matrix. Additionally, the high scoring candidates in all three Latent Class Analysis models of CAPSS performance had high psychomotor subtest scores. Simulators like CAPSS are excellent tests of psychomotor abilities and are representative of the types of basic flying manoeuvres that are tested in the early stages of pilot training. However, more complex flight scenarios, like those found in the later stages of training, as well as the development of systemically complex aircraft, have reduced the need for strong psychomotor abilities and instead generated an increased requirement for improved problem solving abilities and situational awareness (Ebbatson, 2009; Wiener, Chute, & Moses, 1999). The current study demonstrated some movement towards this new dynamic by showing the importance of a Reasoning factor, based largely on the CFAT Problem Solving subtest, in identifying candidates who were successful at CAPSS testing. The Work Rate subtest Table Reading that assesses cognitive processing speed (Gs from the C-H-C model) was also statistically significant for the candidates with high CAPSS scores in all three LCA models.

Spatial ability was found to be a consistent contributor to success in pilot selection. The CFAT Spatial Ability subtest and all five spatial reasoning subtests of the RAFAAT test battery were contributors to the pass /fail performance of the pilot candidates and all three latent class analyses identified high scores on the spatial ability subtests as one of the characteristics of the candidates who had the highest CAPSS scores. These results support the findings of the pilot job analysis completed by Darr (2010) in which spatial awareness was identified as one of the characteristics that distinguished superior helicopter pilots from average ones. Spatial ability plays an essential role in map reading and navigation activities (Cherney et al., 2008), both of which are crucial skills for pilots operating in complex flight environments. Spatial testing, particularly mental rotation and spatial visualization abilities, should therefore remain as one of the essential ability domains in which pilot candidates are tested.

Although abilities like spatial reasoning and psychomotor abilities were clearly identified in the CFAT and RAFAAT batteries, tests of other aptitudes considered important for pilots, including WM, situational awareness, and decision making (Wickens, 2007), are missing in the RAFAAT battery. Sohn and Doane (2004) confirmed that WM was critical for novice pilots particularly because it predicted situational awareness, defined by Endsley and Bolstad (1994) as the perception of elements at a certain time to include their meaning and the projection of their status in the near future. 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. There are no RAFAAT tests that specifically assess situational awareness. The

Instrument Comprehension 2 subtest, part of the Spatial Reasoning domain, is similar to the test 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.

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. The cRontribution of EF to flight performance is not well defined. Herniman (2013) found that components of EF were predictive of academic performance but were not predictive of student flight performance during basic flying training. This may indicate that EF may only make a difference once basic flying skills have been acquired and the pilot candidates move on to more complex flight scenarios which were not included in Herniman's study. Additional research into the role of EF in flight performance will assess the need for its inclusion in pilot selection test batteries.

Amongst the demographic variables, Gender was consistently a significant factor in aptitude testing, particularly in the Psychomotor Ability domain, and female candidates experienced greater difficulty passing CAPSS testing. Each LCA found Gender to be significant, with female candidates consistently overrepresented in the lower scoring class. These findings are consistent with those of Darr (2009) who determined that using CAPSS testing as the selection criterion resulted in a lower selection rate for female candidates. The female candidate scores were also generally lower on the CFAT and RAFAAT subtests, confirming earlier research by Carretta and Ree (1997) who found large mean differences favouring male pilot applicants, particularly for measures of psychomotor ability, spatial ability, and technical knowledge. Their research determined that female pilot applicants were also less likely to meet or exceed the minimum scores on the aptitude tests used in pilot selection. In the current study, the consistent overrepresentation of female candidates in the low scoring CAPSS class in the LCA models, and the lower scores on the aptitude tests across all ability domains in this study indicates that Gender remains a predictor of success or failure during pilot selection.

New technologies. The role of new technologies was discussed briefly in reference to psychomotor ability testing. Technologically complex subsystems in the form of computerized displays, weapons arrays, countermeasures systems, and digital communication generate enormous amounts of information that are presented to the pilot for immediate analysis. It follows therefore that pilot selection systems must assess the pilot candidates' abilities to keep pace with these new processing demands. In the current results, candidates with the highest CAPSS scores in all three latent class analyses had high Reasoning factor scores and high scores on the Table Reading subtest from the Work Rate ability domain. The Work Rate domain assesses cognitive processing speed and, to a lesser degree, WM, both of which have been identified as mission-critical abilities for pilots completing complex tasks (Causse et al., 2011). As such, ability testing for pilots should include subtests that assess the ability to process large amounts of information and to make timely decisions in the presence of distractions and secondary tasks.

Limitations and Future Directions

The current results are based on a restricted sample of pre-screened military pilot candidates and therefore, the results may not be generalisable to more diverse samples of pilots, e.g. civilian pilots or university students studying aviation. All candidates in the archival dataset had been previously selected based on their performance on the CFAT and personality testing using the Trait Self Descriptive Inventory (Darr, 2011). Range restricted samples can produce estimates of correlations that are artificially lower than they would be in an uncensored sample (Shah & Miyake, 1996) however, Shah and Miyake

(1996) also found that the use of a range restricted sample, in this study a group of pre-screened pilot candidates, may reveal domain-specific effects more clearly, as it did in the current analysis.

The Royal Air Force developed the RAFAAT selection battery, which was then purchased by the Canadian Forces and, after a lengthy trial period, was implemented as the selection system for pilots. Candidates who completed the RAFAAT testing during the trial period did so as part of a research initiative and therefore not all of them completed every subtest. The candidate data used in this study were compiled during the aforementioned trial period. Before completing the RAFAAT battery, pilot candidates were advised that their results would be used for research only and would not be the basis of their selection for pilot training. Whether this disclosure had effects on the candidates' outcomes is unknown, however researchers may want to assess the correlations between the outcomes of this study and the outcomes when the RAFAAT battery was used as the selection criterion for success at pilot selection to determine its impact.

The RAF initially based their ability domains on the skills that experienced pilots determined were needed to be successful at flight training; specific subtests were allocated to each of the identified domains. The current study showed that not all the RAFAAT subtests were well connected with the ability domains originally created by the RAF, which lends support to the RAF decision to change the ability domains. In 2013, the RAF introduced a new RAFAAT cognitive model that was developed in recognition of the critical role cognitive processing speed and multi-tasking abilities play in operating technologically complex aircraft (Royal Air Force Aptitude Testing System, 2013). The Royal Canadian Air Force has also adopted this new model. There are seven ability domains that include Strategic Task Management, Perceptual Processing, Short Term Memory and Capacity, Symbolic Reasoning, and Central Information Processing; the Spatial Reasoning and Psychomotor Ability domains that were used in this thesis are still part of the new model. While many of the subtests used in the current analysis remain, albeit grouped into the new ability domains, many new subtests have been added that assess switching capabilities, cognitive updating skills, and system analysis capacity (Royal Air Force Aptitude

Testing System, 2013). This modification brings the ability domains used in pilot selection more in line with the C-H-C model of human intelligence and aligns them with current cognitive psychological theory that is focused on EF development and its ability to facilitate goal directed behaviour and adaptation to novel and complex situations (Best & Miller, 2010; McCabe et al., 2010; Richland & Birchinal, 2013).

The subtests of the RAFAAT examined for this thesis, along with those in the new ability domains adopted by the Canadian Forces, are now the sole measures used by the Royal Canadian Air Force to select pilot candidates for flight training. Although no specific rationale behind the transition away from CAPSS to RAFAAT testing has been offered, CAPSS had low predictive validity with success on the advanced phases of pilot training (Johnson & Catano, 2013). A single engine simulator was a reasonable job sample of the basic flying manoeuvres tested in the early phases of military flight training, however in the later phases, student pilots fly more complex manoeuvres including multi-aircraft formations, aerobatic sequences, and low-level navigation. The subtests of the RAFAAT may better reflect the abilities pilot candidates require in order to succeed in the more advanced phases of flight training.

Previous flying experience data were not included for the pilot candidates in this study so it is unknown whether the subtests that focused on aircrew-specific knowledge like the flight instruments and aircraft orientations presented in the RAFAAT Instrument Comprehension 1 and 2 helped candidates with previous flight experience achieve higher scores on these ability tests. Analysis of the effect of previous flying experience on the outcomes of the RAFAAT subtests may have identified specific ability domains in which these candidates excelled and may also have provided an improved degree of prediction of pilot selection outcomes. Darr (2009) examined the effect of previous flight experience on CAPSS testing and found that twice as many applicants with previous flight experience passed (591/702 or 84.2%) compared to those with no flying experience (344/805 or 42.9%).

There were also no data available on flying training outcomes for the candidates who completed the ability testing in the current study. The predictive validity of the Spatial Reasoning and Psychomotor

Ability domains that differentiated successful from unsuccessful pilot candidates may have been greatly improved if these data had been available and may have also confirmed the role of Reasoning and Work Rate for the candidates who went on to be successful in pilot training. Abilities in these specific domains may also have been significantly correlated with higher levels of performance on certain phases of advanced flight training e.g. formation flying, low level navigation, and instrument flying, which would provide valuable information for those who research and develop pilot selection batteries.

Summary

"Critical assessment of the pilots' requisite level of information processing and reaction time will ensure an objective method of pilot selection" (Barkhuizen et al., 2002, p. 70). In the new aircraft being brought into service with the Canadian Forces, digital instrument presentations and moving-map displays have supplanted traditional cockpit instrumentation and these innovations may necessitate additional refinements to the pilot selection system as the operational requirements for Air Force pilots continue to evolve. The results of the analyses completed for this thesis show that successful completion of pilot selection required candidates to be competent in a number of ability domains, including Work Rate, Spatial Reasoning, and Psychomotor Ability. Monitoring and evaluation of the flight training performance of the pilot candidates who had higher scores on the subtests in these domains will assess the continued importance of these abilities and may also suggest new directions for pilot candidate assessment that will focus on the specific abilities pilots need to take full advantage of widespread technological innovation.

The cessation of CAPSS testing and the development of a more comprehensive RAFAAT cognitive model may help select pilot candidates who possess the abilities needed to successfully complete more complex flying activities which involve cognitive processing speed, working memory, and situational awareness, all components of EF. The results of this research show that subtests assessing cognitive processing skills, like the CFAT Problem Solving subtest and the RAFAAT Critical Thinking subtest, contributed to success in CAPSS testing and may therefore be predictors of success in flight training. Future research may wish to focus on whether the predictive validity of the new RAFAAT ability domains for success in advanced flying training is an improvement over that obtained for CAPSS testing.

Finally, Gender differences in ability testing were a consistent outcome in the current results, particularly in the area of Psychomotor Ability and CAPSS testing. CAPSS testing is no longer used as a selection measure but candidate performance on the RAFAAT battery should be monitored. These data may verify whether testing pilot candidates in multiple ability domains as recommended by Darr (2009) affects the lower selection rate for women that was present when CAPSS testing was the sole measure of success at pilot selection.

Robust and comprehensive aptitude testing may result in a cadre of military pilot candidates who possess abilities across a wide variety of domains. More diverse abilities testing may also result in student pilots who complete military pilot training in a shorter period of time, and whose performance during flight training is of a higher calibre as a result of their expanded skill set. In either case, once the student pilots receive their wings and proceed on operational flight training, they will be better equipped to meet the challenges of today’s complex and ever-changing air environment.

References

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities. Retrieved from http://faculty.mwsu.edu/psychology/dave.carlston/IQ/alfonso.pdf

Asparouhov, T. & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes: No. 14.* Retrieved from http://www.statmodel.com/examples/webnotes/webnote14.pdf

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.

Bailey, M. (1999). *Evolution of aptitude testing in the RAF* (Report No. MP-055-25). Retrieved from the website of Directorate of Recruiting and Selection: http://ftp.rta.nato.int/public/PubFulltext/RTO/MP/RTO-MP-055/MP-055-25.pdf

Barkhuizen, W., Schepers, J., & Coetzee, J. (2002). Rate of information processing and reaction time of aircraft pilots and non-pilots. *South African Journal of Industrial Psychology, 28(2),* 67-76.

Barkley, R. A. (2012). *Executive functions: What they are, how they work, and why they evolved.* New York, NY: The Guilford Press.

Bartram, D., & Bayliss, R. (1984). Automated testing: Past, present and future. *Journal of Occupational Psychology, 57,* 221-237.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68,* 569-579.

Best, J. R. & Miller, P. H. (2010). A developmental perspective on Executive Function. *Child Development, 81,* 1641-1660.

Black, M. S. (1999). The Efficacy of Personality and Interest Measures as a Supplement to Cognitive Measures in the Prediction of Military Training Performance. (Unpublished master's thesis). Saint Mary's University, Halifax, Canada. Retrieved from http://library2.smu.ca/bitstream/handle/01/22676/black_melissa_s_masters_1999.PDF?sequence=1

Boccio, D. (November, 2009). *Aviation mathematics*. Paper presented at the American Mathematical Association of Two-Year Colleges (AMATYC) 35th Annual Conference, Las Vegas, Nevada.

Boer, L. C. (1991). Spatial ability and orientation of pilots. In R. Gal and A. D. Mangerlsdorff (Eds.), *Handbook of military psychology* (pp. 103-114). Chichester, UK: John Wiley & Sons.

Bolker, B. (2007). Likelihood and all that. Retrieved from www.ms.mcmaster.ca (Chapter 6A.pdf).

Bozdogan, H. (2000). Akaike's Information Criterion and recent developments in information complexity. *Journal of Mathematical Psychology, 44,* 66-91.

Burke, E. (1995). Male – female differences on aviation selection tests: Their implications for research and practice. Proceedings of the 21st Conference of the European Association for Aviation Psychology (EAAP) (pp. 188-193).

Burke, E., Kokorian, A., Lescreve, F., Martin, C. J., Van Raay, P., & Weber, W. (1995). Computer-based assessment: A NATO survey. *International Journal of Selection and Assessment, 3,* 75-83.

Carretta, T. R. (1997). Sex differences on U.S. Air Force pilot selection tests. Proceedings of the Ninth International Symposium on Aviation Psychology, Columbus OH (pp. 1292-1297).

Carretta, T. R. (2011). Pilot candidate selection method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1,* 3-8. doi:10.1027/2192-0923/a00002

Carretta, T. R., & Ree, M. J. (1997). Expanding the nexus of cognitive and psychomotor abilities. *International Journal and Selection and Assessment, 8,* 227-236.

Carretta, T. R., & Ree, M. J. (2000a). General and specific cognitive and psychomotor abilities in personnel selection: The prediction of training and job performance. *Ability and Personnel Selection, 8,* 227-236.

Carretta, T. R., & Ree, M. J. (2000b). *Pilot selection methods* (Report No. AFRL-HE-WP-TR-2000-0116). Wright-Patterson AFB, OH: United States Research Laboratory.

Carretta, T. R., & Ree, M. J. (2008). Pilot selection methods. In P. S. Tsang & M. A. Vidulich, (Eds.), *Principles and practices of aviation psychology* (pp. 357-396). Mahwah, NJ: Lawrence Erlbaum.

Carretta, T. R., Rodgers, M. N., & Hansen, I. (1993). *The Identification of Ability Requirements and Selection Instruments for Fighter Pilot Training* (Report No. AL/HR-TP-1993-0016). Retrieved from dtic website: http://www.dtic.mil/dtic/tr/fulltext/u2/a266340.pdf

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Causse, M., Dehais, F., & Pastor, J. (2011). Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *The International Journal of Aviation Psychology, 21,* 217-234.

Chaiken, S. R., Kyllonen, P. C., & Tirre, W. C. (2000). Organization and components of psychomotor ability. *Cognitive Psychology, 40,* 198-226.

Cherney, I. D., Brabec, C. M., & Runco, D. V. (2008). Mapping out spatial ability: Sex differences in way-finding navigation. *Perceptual and Motor Skills, 107,* 747-760. doi:10.2466/PMS.107.3.747-760

Cook, M., & Ward, G. (May, 1996). *Understanding the requirement: A review of common problems in training, selection and design.* Paper presented at the AMP Symposium on ‘Selection and Training Advances in Aviation’, Prague, Czech Republic.

Cooper, L. A., & Regan, D. T. (1982). Attention, perception, and intelligence. In R. J. Sternberg (Ed.) *Handbook of human intelligence* (pp. 123-169). Cambridge, UK: Cambridge University Press.

Damos, D. L. (1996) Pilot selection batteries: Shortcomings and perspectives. *The International Journal of Aviation Psychology, 6(2),* 199-209.

Damos, D. L. (2003). Pilot selection systems help predict performance. *Flight Safety Digest, February 2003,* 1-10.

Damos, D. L. (2011). *KSAO's for military pilot selection: A review of literature* (Report No. AFCAPS-FR-2011-0003). Retrieved from dtic website: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA546965

Darr, W. (2009). *A psychometric examination of the Canadian Automated Pilot Selection System (CAPSS)* (Report No. DGMPRA TM 2009-024). Ottawa, Canada: Defence R&D Canada.

Darr, W. (2010a). *Job Analysis: Air Force Pilots. Jet, Rotary Wing, and Multi-Engine Streams.* DGMPRA TN 2010-013, Ottawa, Canada: Defence R&D Canada.

Darr, W. (2010b). *The Royal Air Force aircrew aptitude test (RAFAAT): Preliminary evidence for validity* (Report No. DGMPRA TN 2010-015). Ottawa, Canada: Defence R&D Canada.

Diamond, A. (2013). Executive functions. *The Annual Review of Psychology, 64,* 135-168. doi:10.1146/annurev-psych-113011-143750

Director Military Personnel Operational Research (DMPORA), (2007). Canadian Forces Aptitude Test practice version. Retrieved from http://cdn.forces.ca/_PDF2010/preparing_for_aptitude_test_en.pdf

Donohue, J. J. (September, 2006). Validating the parallel Canadian Forces Aptitude Test: Two plans. Paper presented at the 48th Annual Meeting of the International Testing Association, Kingston ON.

Dror, I. F., Kosslyn, S. M., & Waag, W. L. (1993). Visual-spatial abilities of pilots. *Journal of Applied Psychology, 78,* 763-773.

Ebbatson, M. (2009). *The loss of manual flying skills in pilots of highly automated airliners.* (Unpublished master's thesis). Cranfield University School of Engineering, Bedford, UK.

Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. *The International Journal of Aviation Psychology, 4,* 241-264.

Fatolitis, P. G., Jentsch, F. G., Hancock, P. A., Kennedy, R. S., & Bowers, C. (2010). *Initial validation of novel performance-based measures: Mental rotation and psychomotor ability* (Report No. NAMRL Monograph 10-52). Retrieved from dtic website: http://www.dtic.mil/dtic/tr/fulltext/u2/a529481.pdf

Fleishman, E. A. (1972). Structure and measurement of psychomotor abilities. In R. N. Singer (Ed.), *The psychomotor domain: Movement behavior* (pp. 78-106). Philadelphia, PA: Lea & Febiger.

Fleishman, E. A., & Quaintance, M. K. (1984). *Taxonomy of human performance: The description of human tasks*. Orlando, FL: Academic Press.

Ganley, C. M., & Vasilyeva, M. (2011). Sex differences in the relation between math performance, spatial skills, and attitudes. *Journal of Applied Developmental Psychology, 32,* 235-242.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York, NY: Basic Books.

Geiser, C. (2013). *Data analysis with Mplus*. New York, NY: The Guilford Press.

Gress, W., & Willkomm, B. (May, 1996). *Simulator based test systems as a measure to improve the prognostic value of aircrew selection.* Paper presented at the AMP Symposium on 'Selection and Training Advances in Aviation', Prague, Czech Republic.

Griffin, G. R. & Koonce, J. M. (1996). Review of psychomotor skills in pilot selection research of the United States military services. *The International Journal of Aviation Psychology, 6,* 125-147.

Grimm, K. J., Ram, N., & Estabrook, R. (2010). Nonlinear structured growth mixture models in M*plus* and OpenMx. *Multivariate Behavioral Research, 45,* 887-909.

Halpern, D. F. (1992). *Sex differences in cognitive abilities.* Hillsdale, NJ: Lawrence Erlbaum.

Harris, J., Hirsh-Pasek, K., & Newcombe, N. (2013). Understanding spatial transformations: Similarities and differences between mental rotation and mental folding. *Cognitive Processing, 14, 105-115*. doi:10.1007/s10339-013-0544-6

Herniman, D. (2013). *Investigating the predictors of primary flight training in the Canadian Forces.* (Unpublished master's thesis). Carleton University, Ottawa, Canada.

Hilton, T. F. & Dolgin, D. L. (1991). Pilot selection in the military of the free world. In R. Gal, & A. D. Mangelsdorff (Eds.), *Handbook of Military Psychology* (pp. 81-101). New York, NY: John Wiley & Sons.

Hult, R. E., & Brous, C. W. (1986). Spatial visualization: Athletic skills and sex differences. *Perceptual and Motor Skills, 63,* 163-168.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340-362.

Hunter, D. R., & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *The International Journal of Aviation Psychology, 4,* 297-313.

Hunter, D. R., & Burke, E. F. (1995) *Handbook of pilot selection*. Aldershot, UK: Avebury Aviation.

Ingalhaliker, M., Smith, A., Parker, D., Satterwaite, T. D., Elliott, M. A., Ruparel, K., …Verma, R. (2013). Sex differences in the structural connectome of the human brain. Retrieved from www.pnas.org/cgi/doi/10.1073/pnas.1316909110

Johnston, P. J. & Catano, V. M. (2013). Investigating the validity of pervious flying experience, both actual and simulated, in predicting initial and advanced military pilot training performance, *International Journal of Aviation Psychology, 23,* 227-244. doi: 10.1080/10508414.2013.799352

Jung, T. & Wickrama, K. A. S. (2008). An introduction to Latent Class Growth Analysis and Growth Mixture Modeling. *Social and Personality Compass, 2,* 302 – 317.

Kantor, J.E., & Carretta, T. R. (1988). Aircrew selection systems [Supplement]. *Aviation Space and Environmental Medicine, 59*, A32-A38.

Li, W-C. & Harris, D. (2001). The evaluation of the effect of a short aeronautical decision-making training program for military pilots. *The International Journal of Aviation Psychology, 18,* 135-152.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49,* 344-351.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Macedonia, M. (2002). Games, simulations, and the military education dilemma. In M. Devlin, R. Larson, & J. Meyerson (Eds.). *Internet and the University: 2001 Forum* (pp. 157-167). Retrieved from https://net.educause.edu/ir/library/pdf/ffpiu018.pdf

Manning, T. A. (2002) Major Changes in Undergraduate Pilot Training 1939 – 2002. Retrieved from http://www.aetc.af.mil/shared/media/document/AFD-070130-081.pdf

McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. A. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24,* 222-243. doi:10.1037/a0017619

McGrew, K. S. (2009). The Cattell – Horn – Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.) *Contemporary Intellectual Assessment (*pp. 136-181). New York, NY: The Guilford Press.

Miele, F. (2002). *Intelligence, race, and genetics: Conversations with Arthur R. Jensen*. Boulder, CO: Westview Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Morrow, D. G., Menard, W. E., Ridolfo, H. E., Stine-Morrow, E. A. L., Teller, T., & Bryant, D. (2003). Expertise, cognitive ability, and age effects on pilot communication. *The International Journal of Aviation Psychology, 13,* 345-371.

Mplus Demo Version 7.2 (2014). Retrieved from www.statmodel.com/demo.shtml

Muthén, B. O. (2004). Latent variable analysis: Growth mixture modeling and related techniques for longitudinal data. In D. Kaplan (Ed.). The Sage Handbook of Quantitative Methodology for the Social Sciences, pp. 345–268. Thousand Oaks, CA.

Nagy-Kondor, R., & Sörös, C. (2012). Engineering students' spatial abilities in Budapest and Debrecen. *Annales Mathematicae et Informaticae, 40,* 187-201.

Nazareth, A., Herrera, A., & S. M. Pruden, (2013). Explaining sex differences in mental rotation: The role of spatial activity experience. *Cognitive Processing, 14,* 201-204. doi:10.1007/s10339-013-0542-8

O'Hare, D. (1992). The "artful" decision maker: A framework model for aeronautical decision making. *The International Journal of Aviation Psychology, 2,* 175-191.

O'Hare, D. (2003). Aeronautical decision making: Metaphors, models, and methods. In P. S. Tsang, and M. A. Vidulich (Eds.), *Principles and practice of aviation psychology* (pp. 201 – 237). Mahwah, NJ: Lawrence Erlbaum.

Olson, T., Walker, P. B., & Phillips, H. L. (2010). Assessment and selection of aviators in the U.S. military. In P. E. O'Connor, and J. V. Cohn (Eds.) *Human performance enhancement in high-risk environments: Insights, developments, and future directions from military research* (pp. 37-57). Santa Barbara, CA: Praeger Security International.

Onyancha, R., & Kinsey, B. (October, 2007). The effect of engineering major on spatial ability improvements over the course of undergraduate studies. *Proceedings for the 37th ASEE/IEEE Frontiers in Education Conference, October 2007, Session T1H, T1H20 – T1H24.*

Ree, M. J., & Carretta, T. J. (1994). The correlation of general cognitive ability and psychomotor tracking tests. *International Journal of Selection and Assessment, 2,* 209-216.

Ree, M. J., & Carretta, T. J. (1996). Central role of g in military pilot selection. *International Journal of Aviation Psychology, 6,* 111-123.

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development.

Psychological Science, 24, 87-92. doi:10.1177/0956797612450883

Royal Air Force (2007). *Tried and tested: The RAF aptitude testing system*. Received by e-mail; Director General Military Policy, Research & Analysis, November 2013.

Royal Air Force Aptitude Testing System (2013). UK MOD – Commercial in Confidence. Obtained from the Canadian Forces Aircrew Selection Centre, April 2014.

Schermelleh-Engel, K. & Moosbrugger, H. (2003). Evaluating the fir of Structural Equation Models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8 (2),* 23 – 74. Retrieved from http://www.dgps.de/fachgruppen/methoden/mpr-online/

Sohn, Y. W., & Doane, S. M. (2004). Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors, 46,* 461-475.

Southcote, A. (2004). *Officer and Aircrew Selection Center Aptitude Test Manuals* (Psychologist Report No. 04/04 May 04). Cranwell, UK: OASC. Received from the Canadian Forces Aircrew Selection Center, April 2014.

Southcote, A. (2007). Assessing the feasibility of a tri-service selection aptitude test battery. *Proceedings of the 49th Annual Conference of the International Military Testing Association, Gold Coast, Australia*. Retrieved from http://www.imta.info/PastConferences/Presentations.aspx

Spearman, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Sternberg, R. J. (1986). *The nature and scope of practical intelligence.* In R. J. Sternberg, and R. K. Wagner (Eds.), *Practical intelligence* (pp. 1-10). Cambridge, UK: Cambridge University Press.

Süß, H-M., Oberauer, K., Wittman, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability – and a little bit more. *Intelligence, 30,* 261-288.

Templin, J. (2008). Latent Class Analysis. Retrieved from http://jonathantemplin.com/files/dcm/ersh9800f08/ersh9800f08_lecture05.pdf

Thurstone, L. L. (1958). *Primary mental abilities.* Chicago, IL: The University of Chicago Press.

Verde, P., Piccardi, L., Bianchini, F., Trivelloni, P., Guariglia, C., & Tomao, E. (2013). Gender effects on mental rotation in pilots vs. non-pilots. *Aviation, Space, and Environmental Medicine, 84,* 726-729.

Vidulich, M. A. (2003). Mental workload and situation awareness: Essential concepts for aviation psychology. In P. S. Tsang, & Vidulich, M. A. (Eds.), *Principles and practice of aviation psychology* (pp. 115-146). Mahwah, NJ: Lawrence Erlbaum.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250-270.

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). *Psychological Methods, 17,* 228-243.

Wheeler, J. L., & Ree, M. J. (1997). The role of general and specific psychomotor tracking ability in validity. *International Journal of Selection and Assessment, 5,* 128-136.

Wickens, C. (2007). Aviation. In F.T. Durso (Ed.), *Handbook of Applied Cognition* (2nd ed., pp. 361-389). New York, NY: John Wiley and Sons.

Wiener, E. L., Chute, R. D., & Moses, J. H. (1999). *Transition to glass: Pilot training for high-technology transport aircraft.* Ames Research Center (NASA/CR-1999-208784).

Wiggins, M., Stevens, C., Howard, A., Henley, I., & O'Hare, D. (2002). Expert, intermediate, and novice performance during simulated pre-flight decision-making. *Australian Journal of Psychology, 54,* 162-167.

Woycheshin, D. E. (2000). *CAPSS: The Canadian Automated Pilot Selection System* (Report No. DTIC ADP010363). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/p010363.pdf

Youngling, E. W., Levine, S. H., Mocharnuk, J. B., & Weston, L. M. (1977). *Feasibility study to predict combat effectiveness for selected military roles: fighter pilot effectiveness*. (MDC E1634). East St Louis, MO: McDonnell Douglas Astronautics Co.

Yu, C. (2002). *Evaluating cutoff criteria of model fit indices for Latent Variable models with binary and continuous outcomes*. (Unpublished doctoral dissertation). University of California: Los Angeles.

Zelazo, P. D., Carter, A., Reznick, J., & Frye, D. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1,* 198-226.

Appendix A

Job Analysis Rotary Wing (Helicopter) Stream

This appendix contains an excerpt from the job analysis of the Rotary Wing stream completed by the Canadian Forces in 2010. The knowledge, skills, aptitudes and other characteristics (KSAOs) identified by Darr (2010a) are shown below and organised into competency groupings (in bold). Where a competency refers to a combination of related KSAOS, it is labeled to best represent the underlying construct that reflects that combination (Darr, 2010a). Interestingly, the ability to attend to multiple stimuli was considered a psychomotor ability competency and not a cognitive capacity, unlike the Royal Air Force Aircrew Aptitude Test (RAFAAT) where it is considered part of the Attentional Capability domain.

i. **Cognitive Capacity**
   a. Math skills (basic calculations);
   b. Reading skills;
   c. Ability to perform basic mental calculations.
ii. **Psychomotor Ability**
   a. Psychomotor skills (Hand/Feet coordination);
   b. Ability to attend to multiple stimuli (auditory, visual);
   c. Attention to detail.
iii. **Communication**
   a. Ability to communicate (verbal).
iv. **Thinking Skills**
   a. Analytical thinking;
   b. Decision making.

(Darr, 2010a, p. 18)

Appendix B

Correlation Matrix

Table B1

Correlations for CFAT and RAFAAT Subtests by Ability Domain – Page 1

Domain	Ability Test	VR	Mathematics Reasoning		Spatial Reasoning						
		CFAT Verbal	Math Reasoning	Numerical Operations	CFAT spatial	CFAT prob. solve	Critical Thinking	Angles, Bearings, Degrees	Direction & Dist.	Inst. Comp. 1	Inst. Comp. 2
VR	CFAT Verbal	1052	.151**	.000	.037	.184**	.111**	.032	.144**	.013	-.008
Math Reasoning	Mathematics Reasoning	560	560	.420**	.198**	**.548****	.302**	.377**	.318**	.220**	**.410****
	Numerical Operations	544	544	544	.020	**.441****	.092*	.269**	.147**	.046	.308**
Spatial Reasoning	CFAT spatial	1052	560	544	1052	.220**	.227**	.357**	.210**	.191**	.179**
	CFAT Prob. Solve	1052	560	544	1052	1052	.273**	.369**	.341**	.240**	.390**
	Critical Thinking	1052	560	544	1052	1052	1067	.274**	.309**	.263**	.308**
	ABD	557	557	544	557	557	557	557	.287**	.298**	.397**
	Direction & Distance	544	544	544	544	544	544	544	544	.276**	.359**
	Instrument Comp. 1	544	544	544	544	544	544	544	544	544	.301**
	Instrument Comp. 2	544	544	544	544	544	544	544	544	544	544
Work Rate	Table Reading	1052	560	544	1052	1052	1052	557	544	544	544
	Vis Search 1 Letters	1052	560	544	1052	1052	1052	557	544	544	544
	Vis Search 2 Shapes	1052	560	544	1052	1052	1052	557	544	544	544
	Vigilance	583	560	544	583	583	583	557	544	544	544
Attentional Capability	Recall Numbers	1052	560	544	1052	1052	1052	557	544	544	544
	CLAN	560	560	544	560	560	560	557	544	544	544
	Digit Recog.	544	544	544	544	544	544	544	544	544	544
Psychomotor Ability	Control of Velocity	1024	560	544	1024	1024	1024	557	544	544	544
	SMA	1036	560	544	1036	1036	1036	557	544	544	544

Note. * $p < .05$; ** $p < .01$; VR: Verbal Reasoning; CFAT – Canadian Forces Aptitude Test; ABD – Angles, Bearings, and Degrees; CLAN: Colours, Letters, and Numbers; SMA: Sensory Motor Apparatus. Shaded areas = *n* for subtest; bottom of chart is *n* for individual correlations. Dotted lines denote boundaries between different ability domains. Solid lines denote same ability domain boundaries. ***Bold*** = correlations between subtests in different ability domains> .400.

Table B2

Correlations for CFAT and RAFAAT Subtests by Ability Domain – Page 2

Domain	Ability Test	Work Rate				Attentional Capability			Psychomotor Ability	
		Table Reading	Vis Search 1 Letters	Vis Search 2 Shapes	Vigilance	Recall Numbers	Colours, Letters, Numbers	Digit Recognition	Control of Velocity	SMA
VR	CFAT Verbal	.053	.012	.034	.158**	.042	.070	-.034	.082**	.041
Math Reasoning	Mathematics Reasoning	.346**	.141**	.119**	.266**	.214**	**.422****	.045	.127**	.151**
	Numerical Operations	**.411****	.376**	.254**	.220**	.169**	**.462****	.110*	.038	.073
Spatial Reasoning	CFAT spatial	.121**	.131**	.157**	.092*	-.024	.141*	-.010	.076*	.080*
	CFAT Prob. Solve	.319**	.162**	.137**	.270**	.220**	**.427**.**	.071	.133**	.164**
	Critical Thinking	.246**	.180**	.206**	.219**	.096**	.286**	.011	.156**	.214**
	ABD	**.420****	.295**	.262**	.311**	.116**	**.411****	.092*	.178**	.231**
	Direction & Distance	.290**	.183**	.163**	.224**	.144**	.304**	.086*	.218**	.200**
	Instrument Comp. 1	.145**	.014	.080	.124*	.094*	.220**	-.057	.237**	**.419****
	Instrument Comp. 2	**.509****	.324**	.263**	.319**	.158**	**.440****	.106*	.156**	.225**
Work Rate	Table Reading	1053	.558**	.493**	.409**	.254**	**.503****	.136**	.192**	.255**
	Vis Search 1 Letters	1053	1053	.661**	.350**	.225**	.398**.	.205**	.098**	.066*
	Vis Search 2 Shapes	1053	1053	1053	.353**	.176**	.338**	.142**	.144**	.078*
	Vigilance	583	583	583	583	.166**	**.434****	.116**	.206**	.185**
Attentional Capability	Recall Numbers	1053	1053	1053	583	1053	.300**	.207**	.093**	.084**
	CLAN	560	560	560	560	560	560	.130**	.190**	.215**
	Digit Recognition	544	544	544	544	544	544	544	.000	.001
Psychomotor Ability	Control of Velocity	1024	1024	1024	583	1024	560	544	1024	.378**
	SMA	1036	1036	1036	583	1036	560	544	1024	1036

Note. * $p < .05$; ** $p < .01$; VR: Verbal Reasoning; CFAT – Canadian Forces Aptitude Test; ABD – Angles, Bearings, and Degrees; CLAN – Colours, Letters, and Numbers; SMA – Sensory Motor Apparatus; Shaded areas = *n* for subtest; bottom of chart is *n* for individual correlations. Dotted lines denote boundaries between different ability domains. Solid lines denote same ability domain boundaries. **Bold** = correlations between subtests in different ability domains> .400.

Appendix C

Factor Analyses of the CFAT and RAFAAT Group 1 Subtests:
One, Two, and Four-Factor Solutions

The factor loadings for the one, two, and four-factor solutions of the factor analysis can be found in Table C1. The scree plot for the factor analysis is in Figure C1 and shows that there are four eigenvalues > 1.0. There is a large difference between the first and second unrotated factors but then the differences diminish.

Table C1

Factor Loadings for Exploratory Factor Analysis (Principal Axis Factoring with Oblimin Rotation) for the CFAT and RAFAAT Group 1 Subtests (N = 1024)

	Domain	1 factor	2 factors		4 factors			
Measure			1	2	1	2	3	4
Table Reading	WR	**0.777**	**0.605**	0.259	**0.603**	0.166	0.161	-0.057
Visual Search 1	WR	**0.695**	**0.917**	-0.143	**0.881**	-0.058	-0.060	-0.004
Visual Search 2	WR	**0.672**	**0.760**	-0.038	**0.800**	0.015	-0.103	0.104
Recall Numbers	AC	0.308	0.228	0.130	0.187	0.013	0.264	-0.268
Sensory Motor Apparatus	PA	0.266	-0.060	**0.512**	-0.053	**0.760**	-0.040	-0.023
Control of Velocity	PA	0.266	-0.001	**0.418**	0.040	**0.474**	-0.001	-0.011
Critical Thinking	VR/SR	**0.393**	0.106	**0.459**	0.141	0.188	0.250	0.199
CFAT Problem Solving	VR	**0.385**	0.104	**0.445**	0.017	-0.050	**0.748**	0.030
CFAT Spatial Ability	SR	0.225	0.081	0.223	0.117	-0.020	0.162	**0.463**
CFAT Verbal Skills	VR	0.108	-0.054	0.241	-0.052	0.041	**0.217**	0.044

Note. WR - Work Rate; AC - Attentional Capability; PA - Psychomotor Ability; VR - Verbal Reasoning; SR - Spatial Reasoning. **Bold** denotes factor loadings > .300.

The one-factor solution accounted for 28% of the variance and had large factor loadings on both the Work Rate domain and two of the four subtests in the Verbal and Spatial Reasoning domains. Both Psychomotor Ability subtests had only moderate loadings. The two-factor solution accounted for 42% of variance. Factor 1 in this solution was clearly a Work Rate factor with a very high loading on the Visual Search 1 subtest. Factor 2 contained four subtests and showed a split between the Psychomotor Ability domain and Verbal/Spatial Reasoning domain. All four subtests in the factor had moderate loading. Even though Factor 2 in this solution contained Verbal/ Spatial Reasoning subtests, both the CFAT spatial ability and verbal skills had low loadings.

The four factor solution, while accounting for 63% of variance, had similar loadings to the two-factor solution but Factor 4 of this solution contained a singleton, the CFAT spatial ability subtest, and was rejected.

Figure C1. Scree plot for the Factor Analysis.

Appendix D

Data Analysis with Mplus

Mplus is a general latent variable modeling program that can be used to conduct a variety of statistical analyses including structural equation modeling (SEM) and mixture modeling (Grimm, Ram & Estabrook, 2010). M*plus* produces individual class probabilities from which latent classes can be predicted and used as predictors of outcome variables (Grimm et al., 2010). The version used for analysis in this thesis was the *Mplus* Demo version 7.2 (2014) which is limited to six observed variables that can be used in an analysis; the CAPSS testing data used for this thesis comprised four.

Mplus is a syntax-based statistical software program. Generally the input file contains these subheadings: Data, Variable, Analysis, Model, Output, and Savedata. The following is the script created for the three-class Latent Class Analysis used in this thesis:

```
title: CAPSS Latent Class Analysis
data: file is CAPSS1.dat;
variable: names = id s1 s2 s3 s4;
usevariables = s1 s2 s3 s4;
classes = c(3);
analysis: type = mixture;
plot: type is plot3;
series is s1 (1) s2 (2) s3 (3) s4 (4)
savedata: file is lca3CAPSS_save.txt;
format is free;
output: tech 11 tech 14;
```

In the syntax, the letter ‘s’ is followed by the CAPSS session number and the letter ‘c’, the requested number of classes. All input lines ending in a semi-colon are commands to *Mplus*; other lines are information only for the researcher doing the analysis. Appendix E contains information on the results of the Latent Class Analyses completed using the data from CAPSS. Tech 11 and Tech 14 in the output line are commands directing *Mplus* to test the number of classes in a mixture analysis using the Lo-Mendell-Rubin (LMR; TECH11) test and the bootstrapped likelihood ratio test (BLRT; TECH14).

Asparouhov and Muthén (2012) provide an excellent overview of both tests.

Latent Class Analysis. Latent Class Analysis (LCA) is a statistical procedure used to classify individuals into homogeneous subgroups (Geiser, 2010). Geiser defined the starting point for classification as the observed response patterns of individuals across a set of categorical items. “In an LCA, the relationships between items are explained by the presence of a priori unknown subpopulations (the latent classes)” (Geiser, 2010, p. 232). In other words, individual differences in response patterns are explained by differences in latent class membership (Muthén, 2004).

There were three goals for each Latent Class Analysis (LCA) completed on the CAPSS scores. These goals are based on those of Geiser (2010): 1) determine the number of classes necessary to sufficiently explain differences in the observed response patterns; 2) determine the most likely latent class membership for the pilot candidates who completed CAPSS testing; and 3) interpret how the identified classes differ from each other.

The Latent Class analyses completed for this thesis was exploratory not confirmatory. Similar to confirmatory factor analysis, exploratory LCA explains the relationships between categorical variables, in this application the scores on the four CAPSS sessions, through their membership in one of several latent classes (Geiser, 2010). LCA can also be confirmatory, where theories about typological differences between individuals can be tested, but model testing was outside the scope of the research completed for this thesis. The issue of selecting the number of classes is addressed in detail by Bozdogan (2000); Geiser (2010); Grimm et al. (2010); and Vrieze (2012) but generally consists of assessing model fit information criteria (Jung & Wickrama, 2008).Once the requested number of classes was specified, model fit was determined using the model fit information criteria.

Model fit information criteria. Model fit assesses the degree to which the Latent Class Analysis fits the sample data to provide information about the degree to which a model is correctly or incorrectly specified for the given data (Yu, 2002). *Mplus* assesses model fit for the LCA using multiple criteria: loglikelihood; information criteria: Akaike Information Criteria (AIC), Bayesian Information Criteria

(BIC), and Classification Quality as defined by Entropy; and Average Latent Class Probabilities for Most Likely Latent Class Membership. “As there does not exist a consensus about what constitutes a “good fit”, the fit indices should be considered simultaneously” (Schermelleh-Engel & Moosbrugger, 2003, p. 24).

Loglikelihood. Geiser (2008) wrote that “…the *log likelihood value* is a measure of the probability of the observed data given the model and is used as the basis for calculating various fit statistics” (Geiser, 2010, p. 238). *Mplus* presents loglikelihood as an H0 or null hypothesis value as a way to compare the fit of nested models, and generally, the lower the loglikelihood value, the better the model fit however, it is hard to interpret by itself and should be used with other information fit criteria (Bolker, 2007).

Akaike Information Criterion (AIC). Vrieze (2012) noted that the Akaike Information Criterion (AIC) is derived from a model’s maximum likelihood estimate by taking into consideration the number of model parameters. Templin (2008) stated that when considering which model fits the data best, the smaller absolute values represent better overall model fit (Bolker, 2007; Templin, 2008).

Bayesian Information Criterion (BIC). As explained when describing AIC, BIC is also derived from a model’s likelihood function, however there is a penalty associated with BIC that increases with *N*; statistical significance becomes more and more difficult to achieve as the sample size increases (Vrieze, 2012). For *Mplus* analyses, the smallest absolute BIC is recommended when selecting the best model fit as well as the overall quality of class membership selection (Muthén, 2004).

Entropy. Entropy is reported by *Mplus* as part of the Classification Quality; it a number between 0 and 1, and is defined as a measure of classification uncertainty (Geiser, 2010). Values near 1 indicate high certainty in the classification while values near zero indicate low certainty (Geiser, 2010).

Average latent class probabilities for most likely latent class membership. The final component of the model fit information criteria is the average latent class probability assigned to each latent class. Each candidate who completed CAPSS testing had the possibility of being in each class in each LCA model, however one probability would normally be much higher than the other(s). The Latent Class Probabilities reported in Appendix E are those that are the highest for each class, however there are lower probabilities reported for each class that represent the possibility that the candidate could belong to another class. For example, the full *Mplus* analysis for LCA three-class model for Latent Class Probability is shown at Table D1. Reading across the Assigned Class 1 information, there is a 95.1% probability that the Class 1 candidates are in the correct class, a 5% chance they could be in Class 2 but were retained in Class 1, and 0% chance that they should be in Class 3.

Table D1

Average Latent Class Probabilities for Most Likely Latent Class Membership: Three-Class Model

	Membership Probability by Class		
Assigned Class	1	2	3
1	.951	.049	.000
2	.024	.918	.058
3	.000	.029	.971

Appendix E

Model Fit Information Criteria and Standard Error Ranges for Latent Class Analyses

The model fit information criteria for the two, three, and four class models are shown in Table E1.

Table E1

Model Fit information for Mplus Latent Class Analysis

Model Fit	2 classes	3 classes	4 classes
Loglikelihood	1376.119	1701.225	1838.029
AIC	-2726.238	-3366.451	-3630.059
BIC	-2662.321	-3277.950	-3516.974
aBIC	-2703.610	-3335.119	-3590.024
Entropy	.911	.895	.862
Latent Class Probabilities			
Class 1	.962	.951	.879
Class 2	.979	.918	.906
Class 3		.971	.939
Class 4			.942
Candidate Numbers by Class			
Class 1 *n* =	320	119	242
Class 2 *n* =	689	304	138
Class 3 *n* =		586	114
Class 4 *n* =			515

The following rules were used to determine which model was the most likely fit for the CAPSS testing data:

- Loglikelihood: the lower the number the better (Bolker, 2007);
- Akaike Information Criteria (AIC): The smallest absolute value (Templin, 2008);

- Bayesian Information Criteria (BIC): The smallest absolute value (Muthén, 2004);
- Entropy: A value closer to 0 indicates high certainty in the classification (Geiser, 2010);
- Average Latent Class Probabilities: Closest to 100% is best (Geiser, 2010).

Using these model fit information criteria, the two class LCA appears to have the best fit.

Standard errors. The standard error ranges for the three LCA models were very small and are therefore provided in Table E2 and not depicted in the LCA figures in the text.

Table E2

Standard Error Ranges for Two, Three, and Four Class Models

	Two-class model	Three-class model	Four-class model
Latent Class 1	.013 - .015	.020 - .029	.001 - .051
Latent Class 2	.005 - .009	.011 - .023	.001 - .059
Latent Class 3		.004 - .009	.001 - .016
Latent Class 4			.001 - .007