Rethinking Spatial Abilities: A Multimethod Examination of its Context-Dependent Nature and Whether Tests Require Increased Conceptual, Contextual, and Perceptual Similarity to Usage Context

by

Sharmili Jong
M.A., University of West Florida, 2008

A thesis submitted to
the Faculty of Graduate and Post Doctoral Affairs
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Psychology

Carleton University
Ottawa, Canada

Abstract

Four studies investigated the role of personnel selection test and job-task similarity for identifying qualified job candidates. According to the Point-to-Point Theory (Asher, 1972), greater number of common points between a test and job task should result in better predictive tests. Test and job-task similarity was manipulated in three distinct ways: conceptually, contextually, and perceptually. A focus group (Study 1) was conducted with 15 Air Traffic Controllers to identify the abilities required on the job and the context in which the job tasks are performed. Based on these results, three laboratory studies were conducted with undergraduate students at Carleton University. Study 2 examined the role of conceptual similarity by administering spatial ability tests that had high and low conceptual similarity to an Air Traffic Control-related game. Study 3 utilized tests that were best predictive of the game performance in Study 2 and administered them under a control and two different job-relevant contexts (i.e., with increased mental workload and with interruptions). Study 4 examined whether selection tests need to be visually similar to job tasks. The results revealed that a test with both increased conceptual and contextual similarity to job tasks was the only predictor of game performance and that perceptual similarity was not necessary. This finding challenges the conventional method of personnel selection test development and administration and has implications not only for ATC but also other occupations. The results also offer an improved understanding of the spatial ability construct, one that is context dependent. A revised definition of spatial ability is proposed.

Keywords: personnel selection tests, test and job-task similarity, spatial ability, Air Traffic Controllers

Acknowledgments

I have been fortunate to have many people in my life who have helped me to accomplish my goals and dreams. Avi Parush, thank you for being patient and supportive throughout this entire process. I was fortunate to have a supervisor who allowed me to take a road less travelled and encouraged me to pursue research in a field that was of interest to me. I have learned a lot and have become a better researcher because of you. I would also like to thank my committee members for their support, discussions, and feedback in making this research better.

I would like to thank the Canadian Department of National Defence and their Aerospace Controller Officers for providing me with the opportunity to work with them on this project. I would also like to thank Aren Hunter for allowing me to use her programming codes to generate two of my tests.

I am grateful for my friends Chunyun, Elisa, Nadya, and Robin. I appreciate all you have done for me over the years, from making sure I had something to eat during the 12-hour days in the lab to helping with the data analysis and everything in between.

I would like to thank my previous supervisors, Bruce Milliken (McMaster University) and Steven J. Kass (University of West Florida) for helping me discover my interest in research and for their guidance in improving my technical writing skills.

I'm forever thankful to have parents who instilled the value of good education from a young age and have unconditionally supported me and have sacrificed much for me. Lastly, I'm fortunate to have a supporting husband who listened empathetically as I talked about my research for years; he helped me to remain calm and collected as one can be while working on a dissertation.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1: Introduction

The success of any business relies on the individuals who make up the organization. Consequently, the efficiency of that organization's personnel selection process acts as a gateway to its accomplishments. Given this, many organizations put forth effort in developing protocols and personnel selection tools to select individuals who will make the best job candidates. While the screening process can involve multiple procedures, including interviews and work samples, administering tests of essential abilities are often preferred early in the process because they are cost effective and easy to administer. As a result, these tests often act as gatekeepers to the organization. Given this vital role, it is imperative to develop and utilize the most appropriate tests for personnel selection purposes. This is especially true for high-risk occupations where selecting unqualified individuals could have consequences at the personal, organizational and, perhaps even at a global level. This research revisits the core issues surrounding the difficulty with developing a good selection test: what aspect of an ability should selection tests assess and how should this test be administered?

The purpose of this research was to explore factors that influence the relation between selection tests and job performance. The studies in this thesis utilized spatial ability as the domain and Air Traffic Control (ATC) as the occupation in which the research questions were explored. Current spatial ability tests have generally been found to have moderate validity coefficients with training success and job performance and have the lowest predictive validity compared verbal, numerical, and perceptual ability tests (Bertua, Anderson, & Salgado, 2005). Therefore spatial tests are used in personnel selection while acknowledging this limitation. While using such tests may be sufficient for some occupations, high-risk occupations, such ATC, can benefit from having better tests. Given this need, the present research utilized a novel approach to improve spatial ability tests and their administration so that the test scores are better predictive of individuals who will pass the ATC training program. A focus was given on identifying successful trainees, rather than predicting successful job performance, since factors such as training and experience can affect performance. Early identification of individuals who will be successful trainees will allow organizations to focus their resources on training individuals with high likelihood of being selected for the job.

The method proposed in this thesis for developing and administering spatial ability tests is an alternative to the traditional method. Current spatial ability tests are developed and administered under the Domain Centered Framework's (Carroll, 1993), which states that it is possible to predict how individuals will perform on various tasks using general, job-non-specific tests under standard testing conditions. This research investigated the utility of developing and administering tests using the Point-to-Point Theory's (Asher, 1972) view that more commonality between a test and a task would yield in a test score that better predicts task performance. The principle behind Point-to-Point Theory has been extracted and applied in personnel selection research to successfully predict individuals' turnover (e.g., Becton, Carr, & Judge, 2011; Becton, Matthews, Hartley, & Whitaker, 2009; Bernardin, 1987). Applying this theory to test development and administration may also benefit personnel selection.

The approach used in this thesis considers job-relevant factors that influence how well spatial ability tests predict individuals' ability to perform job-relevant spatial tasks. Research evidence indicates that individuals' performance on spatial tasks can be influenced by a number of factors, including their mental processing speed (e.g., Mumwa, Pellegrino, Kali, & Carter, 1984; Pellegrino & Kali, 1982), working memory (e.g., Carpenter & Just, 1986; Just & Carpenter, 1985; Lohman, 1988), mental image representation (e.g., Kosslyn, Thompson, Kim, & Alpert, 1995; Kozhevnikov, Hegarty, & Mayer, 2002), and strategy use (e.g., Just & Carpenter, 1985; Nori, Gradicelli, & Guisberti, 2006; Parush & Berman, 2004). Therefore, it is essential that the abilities required on the job and those assessed by tests match in order to ensure that selection tests do not require abilities beyond what is required on the job.

In addition to individuals' ability affecting spatial task performance, findings of Conteras, Colom, Hernández, and Santacreu (2003), Parush and Berman (2004), as well as Waller (1999) indicate that the context in which spatial activities are performed also influences spatial task performance. Even the tools used to assess performance (Waller, 1999) or the type of information presented to individuals for performing a spatial task (Parush & Berman, 2004) have affected how individuals performed on the tasks. In one of their studies, Colom, Contreras, Shih and Santacreu (2003) found that dynamic and static spatial tests (i.e., computerized and paper-and-pencil tests, respectively) loaded on two different factors; this indicates that although these tests measured the same construct, there could be something systematically different about the tests due to the presentation format. Taken together, it is important to consider not only test content, but also the test format and the context in which the tests are administered.

While much attention has been paid in investigating the role of job-relevant factors such as stress (e.g., El Shikieri & Musa, 2012; Motowidlo, Packard, & Manning, 1986) on task performance, there appears to be a gap in knowledge within the personnel selection literature regarding the role of conceptual-, contextual-, and perceptual-similarity between tests and job-related tasks on how well they predict task performance. This research used three distinct methods to increase selection test and task similarity:

- **Conceptual similarity**: The guidelines for personnel selection test development state that only job-relevant abilities should be assessed (Goldstein & Ford, 2002). However, the extent to which the tests should be similar is left to the test developers' interpretation and often similarity is achieved only at a theoretical level. Given this, tests with low and high conceptual similarity (i.e., traditional spatial ability tests and job-specific spatial ability tests, respectively) were used to examine how well they predicted spatial task performance.
- **Contextual similarity**: Evidence from the training transfer literature indicates that the environment in which individuals learn new skills affects how well the new skills transfer to the job environment. More specifically, an increase in similarity between the training and job environment increases the likelihood of successful training transfer to similar situations (Goldstein & Ford, 2002). Studies in this thesis examined whether there is a similar advantage with increased similarity between testing and job-task environment.
- **Perceptual similarity**: Some researchers (e.g., Lievens & Sackett, 2006; Wyatt, Pathak, & Zibarras, 2010) indicate that tests with increased visual similarity to job tasks capture more than the cognitive requirements of a job and are predictive of interpersonal aspects of job performance, which contribute to overall job performance. This thesis investigated the benefit of tests with increased perceptual similarity to job tasks for predicting task performance.

Collectively, the studies in this thesis examined whether personnel selection tests with increased similarity to job tasks are better able to identify most adept job candidates. More specifically, the four studies in this thesis addressed the following three questions:

1. What should be the conceptual-, contextual-, and perceptual-similarity between personnel selection tests and the job task for which they are selected?
2. Taking spatial abilities as a test case, is it a construct that can be understood and assessed independently of the context in which it is used?
3. Taking ATC as an occupation where spatial abilities are critical, how should spatial abilities be tested to be effective predictors of ATC-related performance?

Study 1 was conducted with Air Traffic Controllers (ATCOs) to better understand the conceptual and contextual spatial requirements of ATC tasks. Based on Study 1's findings, Studies 2, 3, and 4 were conducted with undergraduate students to examine whether increased conceptual (Study 2), contextual (Study 3), and perceptual (Study 4) similarity between spatial ability tests and spatial tasks influence how well test scores predict task performance. Study 2 utilized tests with high and low conceptual similarity to an ATC-related task and examined how well they predicted task performance. The best predictive tests in Study 2 were then administered under high and low job-relevant context in Study 3. Study 3 results indicated that spatial planning tests administered under high workload (i.e., increased contextual similarity) were best predictive of task performance. Study 4 assessed whether a spatial planning test with high perceptual similarity to task was a better predictor of task performance than a test with low perceptual similarity.

Clarification of Terminology

Spatial Ability and Spatial Abilities

Although the terms *spatial ability* and *spatial abilities* are interchangeable, *spatial abilities* is used throughout this thesis in instances where an emphasis is required to communicate that this construct is comprised of multiple, spatially-relevant abilities.

Job Tasks and Tests

The job candidate screening process often involves administering test(s) to individuals in order to predict their potential to perform job tasks. For the purpose of this thesis, a *job task* is defined as "a distinct work activity carried out for a distinct purpose" (Cascio & Aguinis, 2005, p. 212). According to this definition, a job is comprised of many duties, which in turn are made up of many tasks (Cascio & Aguinis, 2005).

Successful task completion often requires individuals to have specific abilities. Given this, ability testing is normally administered during the selection process. A *test* of ability is operationally defined as "a standardized procedure for sampling behaviour and describing it with categories or scores" (Gregory, 2011, p. 2). Furthermore, a test, by definition, has correct and incorrect answers that can be used to evaluate individuals' responses and the test scores can be interpreted to estimate ability levels of individuals (Cascio & Aguinis, 2005). The results of a test could then be used to predict other behaviours of individuals (Gregory, 2011). The tests used in behavioural research sometimes appear to be analogous, but simplified, measures of task performance. This is because the concept of a task and test can sometimes overlap; for example, a driving test has both task (i.e., driving) and evaluation (i.e., observation of performance) components.

Conceptual-, Contextual- and Perceptual-Similarity

Conceptual-, *contextual-, and perceptual-similarity* are three vital concepts for this research and are used throughout this thesis. Conceptual- and perceptual-similarity refer to two ways in which individuals process information. While attaining conceptual similarity requires individuals to use nonphysical features of things to associate them, perceptual similarity relies on the shared physical or visual properties between things (Hartfield & Conture, 2006). All valid measurements of spatial ability are conceptually similar because they assess the same underlying construct. Two Boeing 787 planes assigned to different carriers will have high level of perceptual similarity despite the different aesthetic choices made by each carrier. In this thesis, the term *conceptual similarity* is used to refer to the similarity between the specific abilities required to preform job tasks and the abilities assessed using a test. *Perceptual similarity* is used to refer to how physically alike tests are to the tasks performed on the job. *Contextual similarity* is operationally defined as the level of shared situation-induced factors amongst two or more situations. Taking an exam and giving a speech may have high contextual similarity for an individual if these activities exert comparable levels of stress, for example. In this thesis, *contextual similarity* is used to refer to the similarity between testing and job-task environments.

Job-Specific and Job-Non-Specific Tests

Job-specific and *job-non-specific* are two terms associated with the concept of conceptual similarity. Job-specific tests refer to tests that have increased conceptual, and sometimes perceptual, similarity to job tasks. In this thesis, projection, multiple objects tracking (MOT), and spatial planning are all referred to as job-specific spatial abilities because they are aspects of spatial ability that are required to perform ATC-related tasks. Consequently, tests measuring these three abilities are referred to as *job-specific tests*. *Job-non-specific tests* still refer to tests that are conceptually similar to one or more job tasks, but the relationship between the abilities assessed by these tests and job tasks are not as closely related as with job-specific tests. The Vandenberg-Kuse Mental Rotation Test (MRT; Vandenberg & Kuse, 1978) is an example of job-non-specific test for ATC because, while it assesses a relevant ability (i.e., spatial ability), the type of spatial ability assessed does not appear to be what is required to perform ATC-related tasks.

Standard Testing Condition

The *standard testing condition* refers to how personnel selection tests are currently administered; this is the traditional testing situation where test takers are placed in a relatively quiet, classroom-like setting with limited interruptions in order for them to perform their best. The use of this testing condition is challenged in this thesis and, instead, testing with increased contextual similarity to job environment is proposed.

A brief overview of the spatial ability literature is presented in Chapter 2 in order to provide the reader with a better understanding of this ability and the challenges faced in measuring this ability. Theoretical frameworks for how spatial ability is currently assessed and how it can be improved are discussed Chapter 3. A conceptual framework of the thesis is presented in Chapter 4, while the studies and their results are discussed in Chapters 5 to 8. This thesis concludes with a general discussion in Chapter 9.

Chapter 2: Spatial Ability

We rely on our spatial ability on a daily basis to perform various tasks. For example, we depend on this ability to judge distances between cars while driving and to navigate within an unfamiliar environment. Most individuals are able to perform their daily spatial tasks without much difficulty; therefore while individual differences exist, their ability levels may not affect their daily task performance in a critical fashion. Yet, some individuals are faced with highly spatial challenges on a regular basis and face great consequences for poor spatial performance.

Successful spatial task performance is important for individuals such ATCOs (Contreras et al., 2003; O*NET, 2010a), pilots (Dror, Kosslyn, & Wang, 1993; Han, Chien, Chen, Chao, & Wu, 2001), ship and boat captains, commercial drivers, and forest firefighters (O*NET, 2010b). ATCOs perform spatial tasks such as maintaining safe separation between aircraft, providing vectoring instructions to pilots, and giving clearance to pilots to use the runway. For the task of maintaining safe separation between aircraft, ATCOs must be able to spatially plan and organize the aircraft and then provide direction to pilots to ensure safe guidance of each aircraft from one point in space to another. Performing this task requires ATCOs to obtain and assimilate information from multiple sources (e.g., pilots, weather reports, navigation aids). As this example highlights, each spatial task in an occupation may require simultaneous execution and coordination of multiple subtasks, making it difficult to identify the individual spatial components of a task. Consequently, this problem has proven difficult for Industrial-Organizational (I-O) Psychologists and researchers alike in translating the spatial needs of an occupation to appropriate selection tests.

While spatial ability is often used in personnel selection (Thomas & Hersen, 2003) and its importance to multiple occupations is expressed by having a spatial component in most multi-aptitude test batteries (Gregory, 2011), many aspects of the construct are still unclear. This chapter reviews the spatial ability construct with discussions surrounding its definition, composition, and assessment.

Defining Spatial Ability

Even after a century dedicated to research in this field, research on spatial ability is still at its infancy in some regards and this construct is one that is still poorly defined (Waller, 1999). Table 1 lists a sample of the definitions that have been used in the literature. The only commonality among the various definitions appears to be a reference to the abilities needed to deal with information in space. Aside from that, most definitions are quite vague and offer very little in identifying, and consequently assessing, the exact abilities needed to perform spatial tasks. Also, while some definitions suggest that this ability is multifaceted and imply that it must be inferred from tests assessing multiple spatially-relevant dimensions, often only a single assessment is taken. Another weakness is that, based on how spatial ability is defined and how it is tested, there is a lack of clear distinction between spatial tasks and the abilities required to perform these tasks. For example, many definitions include an element of transforming or manipulating images, but is this a spatial task in and of itself or is it the ability required to perform the task. In this case, definition of ability becomes cyclical; we infer the level of an ability based on performance and performance, in part, is due to individuals' ability level.

Table 1

Spatial Ability Definitions

Source	Definition
Linn & Peterson (1985, p. 1482)	"…skill in representing, transforming, generating, and recalling symbolic, nonlinguistic information."
Lohman (1993, p. 3)	"…the ability to generate, retain, retrieve, and transform well-structured visual images."
Sjölinder (1998, p. 47)	"… cognitive functions that enable people to deal effectively with spatial relations, visual spatial tasks and orientation of objects in space."
Bailey (1999, p. 25-3)	"…the ability to form mental pictures and manipulate spatial information in the mind."
Waller (1999, p. 2)	Spatial ability is "acquired and mediated through the visual-spatial abilities identified by psychometric studies. These encompass the mental operations used to encode and process simple visual stimuli such as figures, pictures, maps, or models."
Halpern (2001, p. 129)	Ability to "…imagine what an irregular figure would look like if it were rotated in space or the ability to discern the relationship among shapes and objects."
Colom, et al. (2003, p. 92)	Spatial cognitive abilities refer to "…how individuals deal with materials presented in space."
Hegarty, Montello, Richardson, Ishikawa, & Lovelace (2006; as summarized in Melsom, 2009, p. 13)	"The ability to encode spatial information from visual stimuli, the ability to hold and manipulate representations in working memory, and the ability to draw inferences from spatial representations."

For the purpose of the present research, *spatial ability* is operationally defined as specific, cognitively-relevant, aptitudes utilized to "deal effectively with spatial relations, visual spatial tasks and orientation of objects in space" (Sjölinder, 1998; p.47). An example of spatially-relevant ability is an ATCO's capability to predict where in space an aircraft will be at a given time provided that it is travelling at a given speed and direction. A *spatial task*, on the other hand, is operationally defined as obtaining spatial information about elements in a controlled space and performing actions that can impact the spatial configuration in a controlled space (Hunt & Waller, 1999; Parush & Berman, 2004). In another words, a spatial task requires the use of spatial ability to perform some form of action. An example of a spatial task is an ATCO developing an alternate flight path, which requires using the ability to know where other aircraft are and where they are headed and constructing a new flight path (i.e., the task). An alternate way to distinguish spatial ability and spatial task is while a task generally refers to *what* is done in space, ability refers to *how* it is done. However, even with this distinction, sometimes the task and the abilities required to perform them can overlap.

Components of Spatial Ability

The results of factor analytic studies have led researchers to conclude that multiple factors make up the spatial ability construct. However, there is disagreement among the researchers on the number of factors that make up the construct and how to best describe them. Furthermore, the various factors are often highly correlated (Borich & Bauman, 1972; Goldberg & Meredith, 1975), meaning that they are not distinct from one another. An additional difficulty encountered by factor analytic studies is that they do not always reveal the same factors (Waller, 1999), leading to questions regarding the reliability of those factors. To complicate things further, while the number of factors described in studies range from two to 10, they are not always labelled using the same terminology, even when used by the same researchers (D'Oliveria, 2004). Despite these limitations, *spatial ability* appears to refer to an overall construct with factors that have similar, spatially-related, loadings within a study.

While researchers have identified many spatial factors, *spatial orientation* (also referred to as *spatial relations*) and *spatial visualization* are the two that are most often cited in factor analytic studies (e.g., Carroll, 1993; Colom et al., 2003). *Spatial orientation* refers to the "[a]bility to comprehend the nature of the arrangement of elements within a visual stimulus pattern primarily with respect to the examinee's body as a frame of reference" (Micheal, Guilford, Fruchter, & Zimmerman, 1957. p. 189). *Spatial visualization* refers to one's "[a]bility to mentally manipulate, rotate, twist, or invert a pictorially presented stimulus object" (McGee, 1979, p. 893).

Given that a number of spatial factors have been identified to represent the spatial ability construct, a comprehensive assessment of individuals' spatial ability level should require the assessment of multiple factors. For example, if researchers follow Lohman's (1988) identification of three factors (i.e., spatial visualization, spatial relations/speeded rotation, and spatial orientation), then they need to measure individuals' ability level in all three factors to better understand their capability to perform various spatial tasks. Collectively, these tests may attempt to measure the ability from different perspectives and therefore account for a large portion of the variance in the construct. However, given that these tests are often used in isolation, they likely assess only part of the ability (Shanmugaratnam & Parush, 2012a). If using multiple tests to assess one's overall spatial ability level was the original intent, then the message has been lost over time. Most often only a single test is administered, with the Vandenber and Kuse's (1978) MRT being favoured more than others.

Taken together, spatial ability appears to be a multi-faceted construct that is comprised of various factors or sub-abilities. Viewing spatial ability in this light suggests that the term *spatial ability* is merely a label given for a collection of spatially-relevant sub-abilities. Consequently, a single test assessing only one of the sub-abilities may not be sufficient to accurately infer overall spatial ability levels of individuals. It is theoretically possible to obtain more comprehensive measures of individuals' spatial ability levels by assessing multiple spatially-relevant sub-abilities. Subsequently, these measures could help predict how individuals would perform on various spatial tasks in an occupation. The next section provides a brief discussion on various spatial ability tests that have been used in the literature, including ones that will be used in this thesis.

Existing Spatial Ability Tests

It is possible that the current difficulty with using spatial ability tests for personnel selection may be due to deducing individuals' ability levels based on a single test. It is also possible that the chosen test may not measure the type of spatial ability required to perform tasks on the job. Given these limitations, a variety of spatial ability tests were used in this thesis.

While many different spatial ability tests have been used in the literature, most are used on a one-time basis or only by the developers (D'Oliveria, 2004). Spatial ability tests included in this thesis are the MRT (Vandenberg & Kuse, 1978), which is the most widely used test of spatial ability, eight tests from the Kit of Factor Referenced Cognitive

Tests (Ekstrom, French, Harman, & Derman, 1976a), which have been used by the Educational Testing Services (ETS) for over 30 years, and Hegarty, Kozhevnikov, and Waller's (2008) Object Perspective Taking Test, which has more face validity for measuring spatial ability than most tests. These 10 tests, along with two other tests that have been used in the literature are briefly discussed below.

Vandenberg-Kuse Mental Rotation Test (Vandenberg & Kuse, 1978). As the name states, this test was designed to measure individuals' mental rotation ability. Each question on the test presents individuals with a target 3D image constructed using a number of cube blocks and four similar response images, two of which are rotated versions of the target image. Individuals are required to identify the two rotated images.

Kit of Factor Reference Cognitive Tests (Ekstrom et al., 1976a). This kit contains eight spatial ability tests categorized into three spatially-relevant factors: Spatial Orientation, Spatial Scanning, and Visualization. The Spatial Orientation tests (i.e., Card Rotations and Cube Comparisons) require individuals to mentally rotate 2D or 3D images and determine which of the presented images are alike. The Spatial Scanning tests (i.e., Maze Tracing Speed, Choosing a Path, and Map Planning) require individuals to quickly trace a path through a map to reach a predetermined destination. The Visualization tests (i.e., Form Board, Paper Folding, and Surface Development) require individuals to imagine how shapes will appear when they are manipulated.

Object Perspective Taking Test (Hegarty, et al., 2008). The purpose of this test is to measure individuals' ability to mentally envision different perspectives in space. Individuals are presented with an image of seven objects and each question on the test provides a scenario where the individual is standing by one of the seven objects while facing another object. Individuals are required to point to a third object by drawing the linear relationship between the three objects on the space provided below each question.

Perspective-Taking Ability Test (Hegarty et al., 2006). This test is conceptually similar to the Object Perspective Taking Test (Hegarty et al., 2008), but with the exception that it takes place in 3D space and on a computer. For this test, individuals are placed in a room with four objects along each wall and are required to study the spatial relationship between each object. They are then removed from the room and are asked to imagine standing by one of the four objects while facing the center of the room. They are asked to indicate the location of a requested object using a keyboard.

Spatial Relations Test (Bennet, Seashore, & Wesman, 1972). This test assesses individuals' ability to visualize how a patterned image would look once it is folded. This test is conceptually similar to the Paper Folding Test (Ekstrom et al., 1976a).

Critique of Using Existing Spatial Ability Tests for Personnel Selection

It is first necessary to understand the protocol for developing or selecting personnel selection tests before we can evaluate the adequacy of using current spatial ability tests for personnel selection purposes. A *job analysis* is an investigation of the knowledge, skills, abilities, and other requirements (KSAOs) that are necessary to perform the tasks in a given job (Cascio & Aguinis, 2005, Goldstein & Ford, 2002). Organizations perform this analysis to help identify the job-specific KSAOs that should be assessed during selection. Tests assessing those KSAOs are then selected or developed and validated prior to using them to screening candidates.

Taking the job analysis process into account, a noticeable problem with how spatial ability tests are currently used for personnel selection is that current tests do not assess job-specific spatial abilities. Instead, spatial tests are most often chosen simply because they assess some aspect of spatial ability. In this regard, a given spatial test is expected to measure all aspects of the ability and is treated as a global measure of the ability. However, according to the guidelines for developing selection tests, only job-specific aspects of abilities should be assessed during selection (Goldstein & Ford, 2002). This means that ATCO candidates, for example, may need to be assessed on their ability to detect when two aircraft are on a converging path using a test that assesses their ability to project the trajectory of aircraft in space and calculate the spatial geometry between them, rather than using a test such as the MRT. The requirement of assessing only job-specific ability is because some tests may not assess certain aspects of the ability required for the job, while others may assess abilities that are irrelevant to the job. The use of job-non-specific and job-specific tests to assess ability of individuals follows two distinct theoretical frameworks: Domain Centered Framework and Point-to-Point Theory. These two frameworks are discussed in Chapter 3.

Chapter 3: Two Theoretical Frameworks for Test Development and Administration

General intelligence, *g*, is a concept that was proposed by Spearman (1904) to signify that individuals will display some consistency in their performance on different tasks. He stated that this consistency was due to a general capability that is inherent in each individual. It is believed that *g* is comprised of two types of intelligence: crystalized and fluid (Cattell, 1963). Crystalized intelligence is the result of knowledge individuals accrue over the course of their lives (Cattell & Horn, 1978) and is influenced by external factors such as training and instruction (Cattell, 1963, Stankov, 1978). An example of this type of intelligence is the knowledge of geographic locations. Fluid intelligence is not affected by training (Stankov, 1978) and relies on individuals' analytic ability (Cattell & Horn, 1978). This type of intelligence is measured in tests that require individuals to adapt to new situations (Cattell, 1963), such as those that measure math and spatial ability (Rolfhus & Ackerman, 1996). Solving puzzles require fluid intelligence.

In contrast to Spearman (1904), Thomson (1916) proposed that various information-processing factors, rather than one general factor such as intelligence, contributed to task performance. Thomson emphasized the corresponding (i.e., point-to-point) similarity between cognitive requirements of tasks for predicting performance. This view of personnel selection is in contrast to Schmidt and Hunter's (1977) claim that tests validities are not situation specific and that we can generalize validity to multiple situations. This chapter discusses the Domain Centered Framework, which parallels Spearman's view of using general, job-non-specific tests to measure constructs, and the Point-to-Point Theory (Asher, 1972), which reflects Thomson's view that increased similarity between cognitive elements (e.g., test and task) best predicts performance.

Domain Centered Framework

The Domain Centered Framework views human abilities as something that can be classified into various domains or global abilities (Carroll, 1993). Intelligence testing is an example of domain-centered testing. Other examples include verbal and numerical reasoning, attentional capacity, and psychomotor ability (Bailey, 1999). Colom, Contreras, Botella and Santacreu (2002) found that measures of spatial visualization and spatial orientation/relations load on a higher order factor, indicating that spatial ability may also be a global construct comprised of multiple, spatially-relevant abilities.

The Domain Centered Framework approach allows personnel selection tests to be job-non-specific. In fact, for a given job, a test battery assessing different domains could be administered and a composite score for each job could be obtained using weighted scores for each domain. This allows for the use of one test battery for a variety of jobs (Bailey, 1999). For the spatial ability domain, this means that a spatial ability index could be established for each individual by administering one or more job-non-specific spatial ability tests and it can be used to predict his or her spatial task performance on the job. Individuals with high scores on the test should be able to perform highly spatial tasks, regardless of the context in which the tasks are performed, provided that a valid test was used. Similarly, individuals with low scores should not be able to perform most of the spatial tasks. Research findings such as that of Lohman (1986) support this view. Lohman observed that individuals' spatial task performance did not improve regardless of the amount of time provided to perform the task. He found that individuals were either able to perform the task or they were not.

If the Domain Centered Framework is appropriate for spatial ability test development, then the current use of job-non-specific tests should be sufficient to predict individuals' spatial task performance, provided that it is a valid measure of the construct. The way in which spatial ability is currently assessed indicates that this ability is regarded as a domain-centered construct. This approach to testing places an emphasis on individuals' ability and disregards the contribution of other factors, such the testing environment, in predicting task performance. In contrast to this view, the Point-to-Point Theory emphasizes the role of contextual factors for predicting performance.

Point-to-Point Theory

According to the Point-to-Point Theory for selection, the greater the number of points the predictor and criterion variable have in common, the better (i.e., higher) the validity coefficient (Asher, 1972). The principle behind this theory is conceptually similar to that of the *Identical Elements Theory* in the training transfer domain (Gordon & Kleiman, 1976). Briefly, Thorndike's (1903) *The Identical Elements Theory of the Transfer of Training* states that the amount of training transfer to an unfamiliar situation depends on the number of elements that both the training and new situation have in common (as stated in Goldstein & Ford, 2002). Extracting from these theories, it is possible that the amount of similarity between a spatial test and a spatial task may determine how well the test predicts task performance.

Evidence from personnel selection literature supports the use of the Point-to-Point Theory for selection. In a review of 11 articles on using application blanks to predict work performance, Asher (1972) found that biographical data (i.e., questions about past behaviour) had the best predictive power, more so than measures such as intelligence, mechanical aptitude, finger dexterity, personality, and spatial relations. As an example of the theory's application in real life, Asher states that research has reliably shown that an individual's high school grade point average (GPA) is the best predictor of his or her college GPA. He also states that other researchers have shown that achievements in science, music, art, and other classes in high school were the best predictors of same achievements in college. Other researchers have also found biographical data to be good predictors of various job performance-related measures (e.g., Becton et al., 2009; Dean, 2004; Vinchur, Schippmann, Switzer, & Roth, 1998). These studies lend further support to the theory that increased point-to-point correspondence between a predictor (e.g., college music achievement) and criterion variable (e.g., high school music achievement) results in better prediction.

In addition to predicting task performance, this theory has been applied to predict individuals' behaviour. This theory has been linked to the behavioural consistency model, which states that past behaviour is a good predictor of future behaviour (Wernimont & Campbell, 1968). Becton et al.'s (2011) finding that individuals' history of changing jobs predicted their future turnover supports this theory. In this context, the point-to-point correspondence occurs between past and future behaviours.

Another illustration of the application of this theory is in the use of work sample tests to predict individuals' job performance. Work sample tests are often isolated versions of tasks performed on the job and therefore have a high point-to-point correspondence in terms of what the test and job tasks require from individuals (Asher & Sciarrino, 1974). An example of a work sample test for ATC is asking individuals to communicate with a pilot to troubleshoot a navigation-related problem. While there is support for the use of work sample tests in personnel selection, there is also evidence to caution against assuming that they are the best predictors of job performance. Despite the high point-to-point correspondence between tasks and tests evident in work sample tests, Asher and Sciarrino found that when both motor and verbal types of work sample tests were used to predict job proficiency, biographical data was still the best predictor. This could be because factors such as training and experience influence individuals' job-task performance. With that said, motor work sample tests were the next best predictor of task performance. Given that biographical data and motor work sample tests have high point-to-point correspondence with task performance, the first due to behavioural consistency and the latter due to similarity in task, these are still good predictors of task performance that support the use of Point-to-Point Theory. Verbal work sample tests, on the other hand, were not as good at predicting job performance as the motor tests (Asher & Sciarrino, 1974); this may be due to lower similarity between actually performing a task and verbally describing the process of a task.

Taken together, research evidence indicates that increasing the similarity between two things (e.g., a test and a task) would result in better prediction of individuals' behaviour (e.g., task performance). Although the Point-to-Point Theory makes conceptual sense for personnel selection test development, one difficulty with developing tests that have a high point-to-point correspondence is not knowing the extent to which similarity between a test and a task should be established. The best method to understand how an individual will perform on a job task will always be to assess his or her performance on the task itself after sufficient training. However, in order to make personnel selection tests more practical, it is necessary to identify the level of similarity required between a test and a task in order to more accurately predict task performance.

Drawing from the Point-to-Point Theory, the present research attempts to increase test-task similarity in three distinct ways: 1. by making the spatial ability tests more conceptually similar to ATC-related tasks by making the test content job-specific; 2. by making the tests contextually similar by incorporating contextual factors that are present in the job environment (i.e., increased workload and interruptions) into testing environment; and 3. by making the tests perceptually similar by having tests that looks visually similar to tasks that ATCOs perform.

Concurrent Validity of Conceptually Similar Tests

There is some evidence to indicate that job-specific spatial ability tests will be better predictors of job-task performance than job-non-specific tests. In a study that is most relevant to this thesis, Conteras and colleagues (2003) examined the influence of *static* and *dynamic* tests for measuring individuals' ability to perform spatial tasks. The static measures used in this study were the traditional, job-non-specific, paper-and-pencil spatial ability tests. The job-specific, dynamic, computerized tests required individuals to direct two moving dots to a designated location. The researchers found that while both types of tests measured spatial ability of individuals, the two types of tests tapped into different types of spatial ability. Perhaps dynamic tests captured abilities that are more relevant to the ATC occupation (Larson, 1996). Therefore, the researchers recommended that dynamic (i.e., job-specific) tests should be used in conjunction with static tests for personnel selection purposes, especially in highly spatial occupations where spatial ability is a good predictor of job performance.

Colom and colleagues (2003) further illustrated the benefit of using dynamic, job-specific spatial tests. These researchers found that their job-specific tests (same ones used by Conteras et al., 2003) were better for measuring spatial ability than the traditional static, job-non-specific tests (i.e., measures of spatial relation, spatial visualization, and reasoning). The job-specific tests loaded on one factor, accounting for 10.6% of the variance while the job-non-specific tests loaded on two different factors accounting for only 3.3% and 2.5% of the variance. This indicates that their job-specific tests were not only better for measuring spatial ability, but they were also measuring the same construct. Both studies discussed above provide evidence to indicate that dynamic tests that assess job-specific ability may be better predictors of individuals' task performance than static, job-non-specific, spatial tests.

Based on these findings, the problem currently faced when using spatial ability tests for personnel selection may be due to using tests with low conceptual similarity to any of the tasks in the target occupation. Given this, the studies in this thesis utilized three spatial ability tests (two dynamic and one static) that were job-specific, along with nine static, job-non-specific spatial ability tests to predict ATC-related task performance. The dynamic tests (MOT and Projection) were designed so that they were conceptually and perceptually job-specific. These tests assess two ATC-related abilities: multiple objects tracking and projection (Barbarino, Clark, & Phillip, 1998). A multiple objects tracking test was included since a large portion of ATCOs' job involves tracking planes, either on the radar or visually. A projection test was included because ATCOs need to continually project what will happen in the near and distant future (Barbarino, et al., 1998; Shanmugaratnam & Parush, 2012b).

One of the difficulties with developing ATC-relevant tests is that they are often required to be dynamic given the nature of the job. Consequently, the ATC-relevant tests in previous research (e.g., Colom et al., 2003; Conteras et al., 2003) were also dynamic; this leads to questions regarding which aspect of increased similarity, conceptual, perceptual, or both, affect how well the tests predict job-related task performance. In order to address this issue, spatial planning ability, which is an important ability required to perform ATC-related tasks, was assessed using both non-dynamic (i.e., Map Planning) and dynamic (i.e., Projection) tests.

Concurrent Validity of Contextually Similar Tests

While research on the role of contextual similarity in predicting individuals' behaviour is generally lacking (Furr & Funder, 2004; Sherman, Nave, & Funder, 2010), there is evidence to indicate that increasing the similarity between two situations results in more predictable patterns of behaviour. Furr and Funder (2004) observed individuals' behaviour in dyadic interactions and found that if two situations were reported to be similar (either subjectively assessed by participants or objectively assessed by how many common elements the two situations had), individuals behaved more consistently than when situations were judged as dissimilar. The researchers concluded that, "[r]egardless of how situational similarity or behavioral consistency were operationalized, greater similarity was related to greater consistency" (p. 421). Such findings have been demonstrated in recent (e.g., Sherman et al., 2010) as well as in previous (e.g., Klirs & Revelle, 1986; Lord, 1982) studies. Collectively these studies indicate that behaviours of individuals are context dependent and individuals behave similarly in similar context.

Extracting from the behavioural literature, it may be beneficial to examine how the contexts in which individuals perform job tasks and personnel selection tests affect how well the tests predict task performance. If individuals behave similarly in similar situations, then an increase in contextual similarity should result in test scores that better predict task performance. Consequently, contextualizing spatial ability tests so that they are administered in conditions similar to what is experienced by ATCOs may result in better identification of successful training candidates. In support of this view, Austin and Villanova (1992) explicitly state that individuals' work performance may be influenced by the situation in which it occurs and therefore performance measures should be taken within the situational context. They emphasize this point by stating that it may not be possible to measure general abilities of individuals in order to predict their general performance. This is in contrast to how current spatial ability tests using the Domain Centered Framework are developed and administered and advocates the use of tests based on Point-to-Point Theory.

Examining the working context of ATCOs indicates ways in which test and job-task context similarity could be increased. ATCOs are responsible for controlling air traffic flow within a designated airspace and most rely on navigational and communication aids to relay instructions and information to pilots and ATCOs in other control towers (Human Resources and Skills Development Canada, 2013). Collectively, these individuals are responsible for the safe navigation of aircraft from one airport to another and they deal with any issues that may impede the achievement of that goal. Reviewing the working condition of ATCOs revealed that these individuals perform their job tasks under high workload (Barbarino et al., 1998; Bureau of Labor Statistics, 2014;

Seamster, Redding, & Kaempf, 1997) and with interruptions due to having to communicate with multiple pilots and taking phone calls, for example. These two contextual factors were also identified during a focus group with Aerospace Control Officers (Canadian Forces' title for ATCOs and subsequently referred to as such; Shanmugaratnam & Parush, 2012b). The ATCOs believed that individuals' ability to perform job tasks under increased workload and with interruptions can affect their job performance. Consequently, the impact of these two contextual factors was examined in this thesis. There is also evidence from other domains to indicate that workload (e.g., Cox-Fuenzalida, 2007; Gonzalez, 2005) and interruptions (e.g., González & Mark, 2004; Monk, Boehm-Davis, & Trafton, 2004) can affect task performance considerably. Administering tests under these contexts should increase test and ATCO job-task similarity. Consequently, the contextualized tests' scores should be better predictive of ATC-related task performance than non-contextualized tests' scores.

Concurrent Validity of Perceptually Similar Tests

Research on test format similarity provides evidence to indicate that tests with increased perceptual similarity to job tasks are better predictors of job performance. Lievens and Sackett (2006) used situation judgment tests, which describe problem situations or scenarios that require individuals to use their KSAOs to solve (Christian, Edwards, & Bradley, 2010), and manipulated whether the test was administered via a video or a written format, while having the same content. They found that the video version of the test predicted interpersonal performance better than the written version. Similarly, Wyatt et al. (2010) found that a role-playing test was better at distinguishing between high and low performers (as measured by sales output) than a situation judgment test, which is often presented on paper. Robertson and Kandola (1982) have also found that tests requiring psychomotor ability, such as role-playing, are better predictors of job performance than decision-making tasks, such as situation judgment tests.

A recent investigation by Lievens and Patterson (2011) found that in real-life personnel selection context, both written situation judgement tests and assessment center methods (e.g., role playing) predicted job performance of general practitioners (measured using supervisory ratings of individuals' level of empathy, communication, problem solving ability, professional integrity, and ability to cope with pressure) better than a test that merely assessed individuals' knowledge. Upon further analysis, the researchers noted that while both methods were equally good at predicting cognitive elements of the job, the more interactive assessment center method was able to predict non-cognitive aspects that were required on the job (e.g., communication). This may be due to the increased similarity between the assessment center method and how tasks are performed on the job. Lievens and Patterson's findings indicate not only that different tests may reveal different results regarding individuals' capacity to perform certain tasks, but also that more perceptually similar tests may be better for measuring job performance since they may account for more than just the cognitive requirements of a job. This thesis utilized two perceptually similar and 10 perceptually dissimilar tests of spatial ability. If increased perceptual similarity between tasks and tests is an important consideration for predicting task performance, then the two tests with increased perceptual similarity should be better predictors of individuals' spatial task performance than the other tests.

Chapter 4: The Conceptual Framework of the Thesis

Many researchers have speculated and debated about the source of individual differences in spatial ability. Although a consensus has yet to be reached, the two possibilities are that it is either an innate ability (or affected only by early experiences) or one that continues to be influenced on a daily basis. The use of current, non-contextualized, spatial ability tests suggest that spatial ability is generally regarded as an ability that is rather stable within an individual, unless that individual actively seeks to improve it. If this view is correct, then adults' performance on spatial tasks should be rather stable across time and situations. Accordingly, selections tests developed using the Domain Centered Framework should be sufficient to predict individual's spatial task performance. These tests can be general and job-non-specific, advocating the continued use of current spatial ability tests (See solid line in Figure 1).

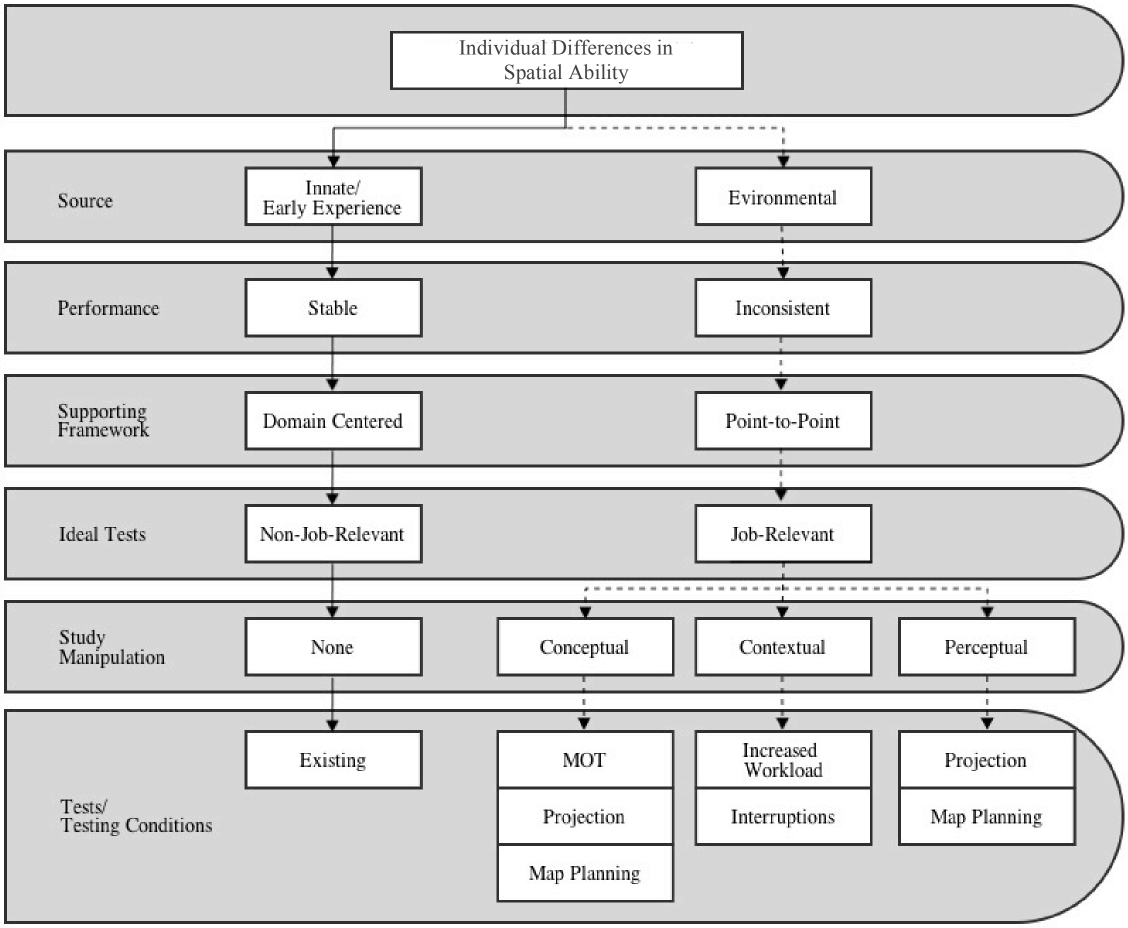


Figure 1. The conceptual framework of the thesis.

If spatial ability is influenced by individuals' environment, then the current use of job-non-specific tests is inadequate. Studies on environmental and experiential influence on spatial ability (e.g., Cherney & London, 2006; Schellenberg, 2004; Terlecki, Newcombe, & Little, 2008) provide strong evidence to indicate that external factors influence this ability. If the spatial ability is shaped by one's environment, then it is possible that the immediate environment in which spatial tasks are performed may also influence performance. If this is true, then individuals' performance on a task would not be consistent under different context. In this case, the Point-to-Point Theory would be most apt for spatial test development and administration and it would advocate the use of job-relevant tests. Applying this theory would result in having tests that are more conceptually, contextually, and perceptually similar to job tasks. In this thesis, the test content, administration context, and test format were made to be similar to what ATCOs experience on the job. Test content was manipulated in Study 2 using tests that assess job-specific spatial abilities. Test context was manipulated in Study 3 by administering tests under increased workload or with interruptions. Perceptual similarity was manipulated in Study 4 by administering two tests of spatial planning ability that were either visually similar or dissimilar to ATC-related tasks (See dashed line on Figure 1). The four studies and their results are presented in Chapters 5 to 8.

Chapter 5: Study 1 – Understanding the Problem and Need in Testing Spatial Abilities of ATCOs

The objectives of this study were to perform a Cognitive Task Analysis in order to identify the spatial tasks performed by ATCOs and to determine the cognitive processes that are required to perform these tasks. Recall that a spatial task was previously defined as the general manipulation or configuration of information in space *and* using this information to perform either a physical or mental task. Cognitive processes are defined as anything an individual does mentally in order to obtain, assess and comprehend spatial information, use reasoning and judgment on it, retain it in memory, and utilize it to make decisions in order to perform a task (Darken & Peterson, 2002; Hintzman, O'Dell, & Arndt, 1981; Thorndyke & Hays-Roth, 1982). Ultimately, the cognitive processes were translated into corresponding spatial abilities and spatial tests.

Method

Participants

Three separate focus groups were conducted for each of the three streams of the ATC occupation within the Canadian Forces. Each focus group had five SMEs: Weapons (5 males), Instrument Flight Rules (IFR; 4 males, 1 female), and Visual Flight Rules (VFR; 3 males, 2 females). Although the 4:1 male to female ratio of participants was unequal, the SMEs stated that it reflected the ratio of ATCOs in the field. Three SMEs in the Weapons division identified themselves as current instructors and two as former instructors (one was a career manager at the time of the focus group and the other was retired). These individuals' experience as Weapons Controllers ranged from 6 to 19 years, with a mean of 12.00 years (*SD* = 6.12 years). All IFR SMEs were employed as

IFR Controllers at the time of the focus group and had 8 to 25 years of experience, with a mean of 13.60 years (*SD* = 7.33 years). Most IFR SMEs stated that they were both IFR and VFR qualified. Additionally, they reported that the IFR Controllers' task is primarily divided between administrative work and controlling air traffic. For the purpose of the focus group, they were asked to discuss only the air traffic controlling aspect of their job in the IFR division. All VFR SMEs were VFR Controllers at the time of the focus group and had 4 to 23 years of experience, with an average of 11.75 years (*SD* = 9.00 years).

Procedure

The same procedure was followed for all three focus groups. Each day was divided into a morning and an afternoon session and lasted 6 hours in total, which included a one-hour lunch break and two scheduled 15-minute breaks. After introductions, a PowerPoint aided presentation was given by the primary researcher at the start of each day to communicate the overall goal of conducting the focus group, as well as to provide the objectives and schedule for the day (See Appendix A for presentation slides). The main purpose of the focus groups was to uncover the appropriate mental model for the ATC occupation. This was achieved by using an open-ended process that is often referred to as an "affinity diagram process" (Spool, 2004). Briefly, in this process, participants work individually and then as two groups to identify as many tasks and elements in their occupation as possible. Once these are exhausted, participants work as a group to cluster and categorize the tasks and create links between the different categories (hence the name *affinity diagram*). The morning and afternoon sessions utilized this approach to ascertain the mental model for the ATC occupation. The PowerPoint aided presentation resumed throughout the day in order to communicate the goals and instructions for the activities performed by the SMEs. Any questions raised by the SMEs were answered prior to administering the informed consent form (see Appendix B) and starting each focus group session. A second researcher moderated the focus groups while the primary researcher took notes of verbal information that was exchanged during the sessions but were not necessarily captured in the spatial and cognitive diagrams the SMEs generated.

Using the affinity diagram method, the morning session required SMEs to identify the spatial tasks they perform on the job. Individuals were first asked to work independently and then as two groups of 2-3 individuals and record each spatial task on separate Post-it® notes. Approximately 45 minutes were dedicated to this activity. The first break occurred once the SMEs exhausted all possible tasks and were no longer able to generate new ones. After the break, the SMEs were asked to work as a group and categorize their notes according to criteria agreed upon by all group members, with clear titles given to each category. The developed diagram was then presented to the researchers, who then summarized the main findings and verified it with the SMEs. This activity lasted approximately 60 minutes and was followed by the lunch break.

The afternoon session focused on ascertaining the cognitive processes required for performing the spatial tasks identified in the morning session. Following the same procedure as the morning session, the SMEs spent 45 minutes working individually, and then as a group, to identify the cognitive processes, after which they had a 15-minute break. After the break, the SMEs spent 45 minutes working as a group to categorize the cognitive processes, before presenting their diagram. Once again, the researchers summarized the main findings to ensure accurate interpretation.

Prior to the completion of each focus group, the SMEs were asked to complete a brief questionnaire on the importance of various spatial ability dimensions for the ATC occupation and the necessity of assessing these abilities during personnel selection (seeAppendix C). These dimensions were taken from a literature review on spatial ability conducted by the researchers for the Department of National Defence (Shanmugaratnam & Parush, 2012a). Participants were debriefed at the end of each focus group (see Appendix D). Each focus group session was audio recorded with the permission of the SMEs. Refreshments were provided throughout the day and breaks outside scheduled the times were permitted on a need basis.

Results

The goal of the focus groups was to extract the spatial tasks performed on the job and their cognitive requirements. This information was collected from both the focus group activities and the discussions generated by the SMEs. For each cognitive task, the corresponding spatial ability was identified and classified it into one or more of the seven spatial ability dimensions according to Shanmugaratnam and Parush (2012a).
All three focus groups (i.e., streams of ATC) were analyzed separately before identifying overlapping KSAOs and the results are presented in this order in this chapter. The following results were presented to the Canadian Forces in a technical report (Shanmugaratnam & Parush, 2012b) and have been reviewed by an SME from the respective ATC stream to ensure accurate interpretation of information.

Weapons Sub-occupation

The role of the Weapons Controllers is to provide sound command and control to all air and ground assets within their jurisdiction. There are three positions within this division: Line Controller, Senior Director, and Mission Crew Commander. Individuals become Line Controllers upon training completion and perform job tasks similar to that of ATCOs, but are restricted to the military. As individuals gain experience, they are typically promoted to be Senior Directors, where they supervise and instruct Line Controllers. Senior Directors may then become Mission Crew Commanders, who are responsible for planning and executing missions.

The SMEs noted that there is currently a 40% failure rate within the Weapons training program, even after selecting individuals based on educational and other entry requirements. They reported that this may be due to the fact that critical abilities are not adequately assessed prior to training. Selecting individuals with job-specific KSAOs was reported as not only important for training, but also later when those individuals become supervisors and trainers because promotions are often based on length of time in a particular division and other criteria such as education level and second language ability, rather than on having KSAOs pertinent to the occupation. The SMEs reasoned that this promotion practice likely influences the failure rate of trainees because trainers may not have the necessary KSAOs themselves in order to supervise and train new individuals.

Cognitive Requirements

The SMEs agreed that their current screening test does not access all cognitive abilities that are necessary for the job. The affinity diagram activity revealed 51 spatial tasks performed by Weapons Officers (See Appendix E). From this, ten cognitive abilities were identified as important for the Officers in the Weapons sub-division: one's ability to (a) think in three-dimensions; (b) apply learned knowledge to three-dimensional tasks; (c) calculate and think in degrees; (d) provide navigational directions to pilots while recognizing that the pilots' perspective will be different from one's own; (e) recognize that objects in space can be identified from multiple reference points; (f) alternate between various points of reference; (g) access and utilize information from multiple sources to devise and execute a plan; (h) maintain an updated mental model of the current situation; (i) change plans when unexpected events occur; and (j) communicate thoughts coherently.

Discussions during the focus group revealed two additional cognitive abilities that are especially important for Controllers. First, individuals need to be able to use the updated mental model of the current situation to predict what will happen in the near and distant future (e.g., where an aircraft will be at a given time). The SMEs reported that while trainees are often good at identifying air-traffic conflict that is about to occur in the immediate future, they find it difficult to project those that will occur in the distant future (e.g., 2 minutes, 5 minutes). The SMEs argued that the ability to generate a mental model of a situation and to use that model to plan ahead and anticipate future events was extremely important and that it should be assessed early in the selection process. Second, all SMEs were in agreement that Controllers need to be able to execute multiple tasks simultaneously. Although most trainees were able to perform basic tasks on the job, the performance of some individuals was impaired when they were required to take in information from multiple inputs and perform multiple tasks simultaneously, especially within restricted timeframes. The SMEs identified that this is an area in which they observed differences between academic training and practical skills; given unlimited time, most individuals were able to execute the necessary job-related tasks but when time was restricted, some individuals were not able to do so.

Spatial Requirements

An analysis of the cognitive requirements revealed that *Verbal Reasoning*, *Spatial Visualization*, as well as spatial and non-spatial *Memory* were the spatially-relevant abilities that were most important for Weapons Controllers, while *Way-finding*, *Spatial Imagery*, *Processing Speed*, and *Mental Rotation* were the least important abilities. Table 2 presents the number of times the SMEs identified cognitive tasks demanding each spatial requirement. The results for all three ATC streams are presented together to facilitate comparison.

Table 2

Frequency of Each Spatial Ability Dimension Identified as Important for Performing ATC-related Spatial Tasks

	Spatial Ability Dimensions						
AECO streams	Verbal Reasoning	Spatial Visualization	Spatial and Non-Spatial Memory	Mental Rotation	Processing Speed	Spatial Imagery	Way-finding
Weapons	29	24	17	5	4	2	0
IFR	24	18	3	3	1	0	0
VFR	29	19	4	7	3	0	0

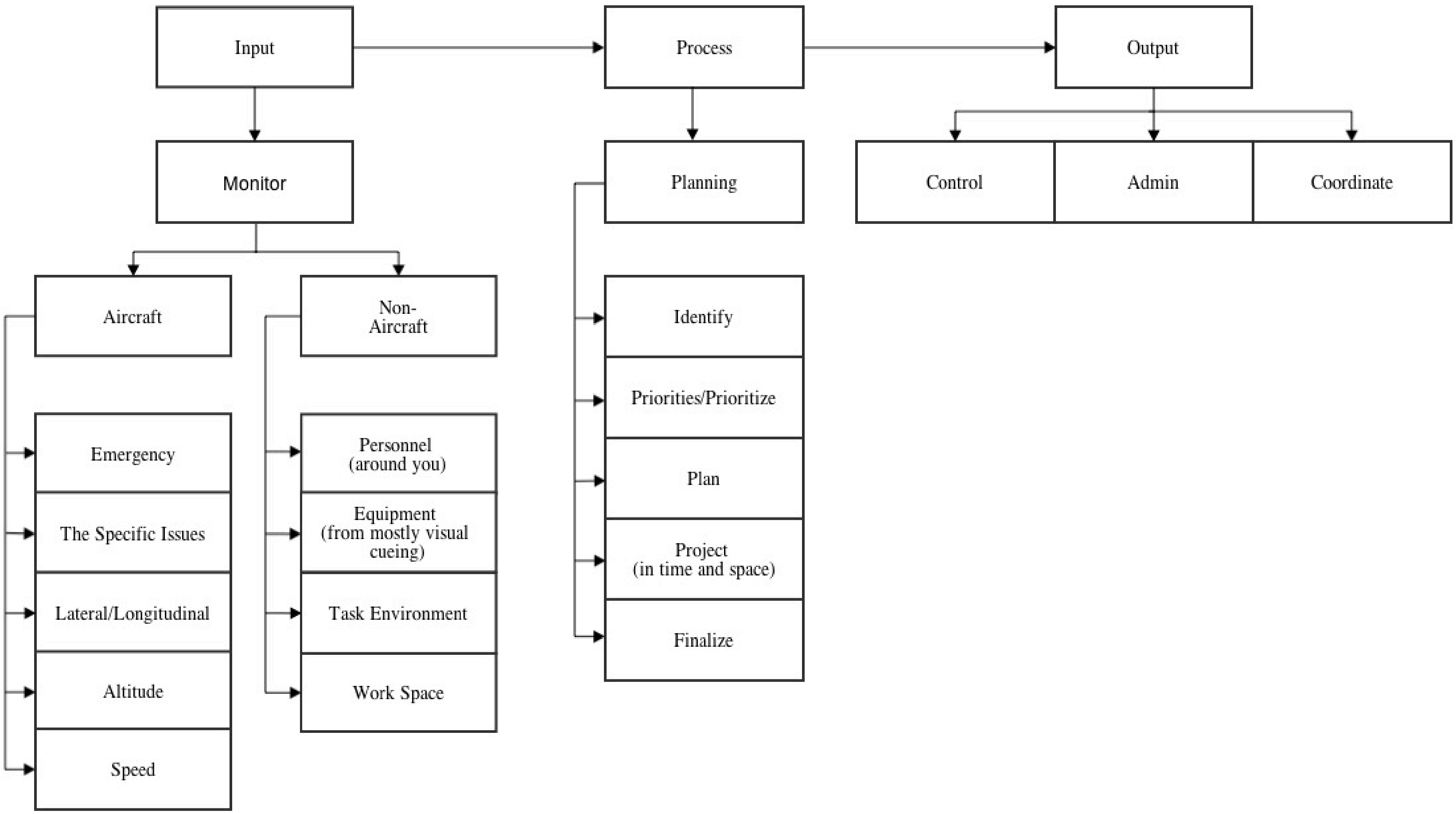
Key Findings

The SMEs agreed that the majority of the problems on the job arise when individuals have difficulty identifying and analyzing problems and then fail to follow-up with the necessary action(s) to rectify it in a complex environment where multiple events often occur simultaneously. More specifically, they disclosed that while most individuals in the training program were generally capable of performing the various job-related tasks in a static environment (e.g., in the classroom setting), difficulties arose when similar tasks were required to be performed in the real-world; this was especially true when time was a factor. Consequently, the SMEs proposed that a screening test that assesses individuals' ability to perform job tasks under realistic time constraints and other job-relevant factors would be ideal to use for identifying qualified ATCO trainees.

Instrument Flight Rules Sub-Occupation

IFR Controllers described their primary responsibility as controlling all assets within 25-60 nautical miles of airspace, depending on their base. These Controllers are required to use instruments, such as radars, to scan their airspace so that they can identify and process tasks that need to be completed in order to produce some form of output. The Controllers referred to these three steps as *input*, *process*, and *output*, respectively (See Figure 2).

Figure 2. Job responsibilities of IFR Controllers



Information at the input stage requires monitoring at the non-aircraft and aircraft level. At the non-aircraft level, the Controllers monitor the surrounding personnel, equipment, task environment, and workspace. At the aircraft level, these individuals monitor all aircraft and all information pertaining to the aircraft. This requires (a) checking for any emergency that may occur by identifying if pilots are following all instructions and determining if any situation requires attention, (b) being aware of and monitoring aircraft-specific characteristics that could affect the type of instructions provided to pilots, and (c) being cognizant of lateral/longitudinal spacing of aircraft, aircraft altitude, and speed.

At the process stage, the previously identified and cognitively registered information regarding the aircraft is processed to formulate an action plan. This requires (a) identify or develop sequencing and instruction plan for each aircraft; (b) identifying aircraft that have prioritized arrivals and departures so that they could be addressed accordingly; (c) formulating and evaluating the plan and deciding whether or not additional Controllers will be needed to deal with the volume of aircraft; (d) projecting where each aircraft in his/her jurisdiction will be in the immediate future and what actions will be necessary to maintain safety; and (e) finalizing the plan.

Once a plan has been formulated and finalized, the output stage requires three activities to be performed simultaneously: (a) taking control of the aircraft by communicating necessary information with the pilot (e.g., climb/decent instructions); (b) performing administrative duties (e.g., inputting range, bearing, and line of sight); and (c) coordinating all aircraft within the controller’s airspace to avoid conflicts (e.g., coordinate with other agencies responsible for safe manoeuvring of the aircraft).

Cognitive Requirements

The IFR Controllers reported that they ask themselves a series of questions in order to guide aircraft: "Where is my aircraft now and where does it want to go? How is it going to get there? Where are the other aircraft in relation to the airfield and this aircraft? And what other factors (e.g., speed and altitude of the aircraft) do I need to take into account?" While each controller must answer these questions prior to executing their action plan, the approaches taken to answer these questions vary for each controller. The SMEs reported that how these questions should be answered is not something that is formally taught during training and each controller learns to devise his or her own set of strategies. This is because their training focuses primarily on what should be achieved, and not how it should be achieved. This flexibility in how problems are addressed posed difficulties in extracting cognitive requirements shared by all Controllers. As an alternative means to arrive at essential cognitive abilities, the SMEs were asked to compare the qualities of good and poor Controllers. This process resulted in the SMEs identifying 32 spatial tasks performed by IFR Controllers (See Appendix F).

When the SMEs were asked, "what separates a good IFR controller from a bad one?" they noted two main abilities as essential for this job: good IFR Controllers are able to (a) successfully perform their job-related tasks under time constraints, and (b) handle a heavy mental workload. When probed further, the SMEs stated that given that a large portion of their job revolves around planning for the future course of aircraft, one's ability to projects events into the future was also a vital ability. Projection requires individuals to take in and cognitively process information form various sources, synthesize it with the current state of affairs, and then formulate a plan to execute. The

SMEs reported that individuals who are good at their job are able project correctly on their first try or are able to learn from their mistakes. Alternatively, some individuals continue to make errors because they adhere to a rule they have learned without considering situation dependent information when making decisions.

Further discussions revealed that IFR Controllers' ability to (a) estimate, rather than calculate, solutions to mathematical problems; (b) compare obtained and expected results; (c) recall basic rules and protocols learned during training and use it in conjunction with contextual information to make decisions; (d) be flexible to change initial plans and adapt as required; (e) visualize airspace in three dimensions; and (f) work quickly and think several steps ahead to detect potential problems and resolve them in a timely manner were also important.

Spatial Requirements

Analysis of the cognitive tasks revealed *Verbal Reasoning* and *Spatial Visualization* as the two main spatially-relevant abilities required for this sub-occupation, while spatial and non-spatial *Memory*, *Mental Rotation*, *Spatial Imagery*, *Processing Speed*, and *Way-finding* abilities as the least relevant (See Table 2 on p. 39).

Key Findings

The primary cognitive task performed by IFR Controllers is projecting the future course of aircraft and revising plans in order to avoid aircraft conflicts. Individuals' ability to work under time pressure, with increased mental workload, and to project events in the future distinguishes good and inept IFR Controllers. Assessing these abilities during selection may reduce training failure rate.

Visual Flight Rules Sub-Occupation

VFR Controllers perform tasks similar to that of IFR Controllers but they rely exclusively on their sight to obtain information, rather than rely on instrumental aids. These SMEs identified that a large portion of their job revolves around air and ground traffic management. More specifically, they are responsible for (a) aerodrome management (e.g., clearing takeoff/landing/close, controlling vehicle movements); (b) runway management (e.g., deciding how many runways to keep open, taking into consideration the ground Controllers' plan); (c) sequencing (i.e., deciding where to fit aircraft in the airspace); (d) monitoring passing traffic (e.g., deciding if the physical separation between aircraft is sufficient); (e) dealing with emergency situations and aircraft; (f) anticipating potential conflicts and preparing resolutions; (g) coordinating the actions performed with other crew members; and (h) providing radio communications.

Cognitive Requirements

The SMEs revealed that VFR Controllers perform 33 spatial tasks (See Appendix G). The SMEs believed that assessing five essential cognitive abilities could reduce the 20-40% VFR training course failure rate. First, individuals' memory capacity was identified as one of the important abilities to evaluate. This is because this job requires remembering and recalling learned materials (e.g., procedures, flight characteristics, radio transmission codes) to perform job tasks. More importantly, good memory capacity is important given that VFR Controllers need to perform visual identification and visual and mental tracking of aircraft. This is because the Controllers often visually identify an aircraft and then lose sight of it for a short time while it flies over their tower before appearing again at a new spot. While the aircraft is out of sight, the Controllers must not only remember that there is an aircraft that is currently not visible to them, but they must also predict where it will reappear after some time has passed. This ability to project events into the future was identified as a second essential ability. The SMEs agreed that projection usually requires one to visualize a situation two or three steps in advance of its occurrence and relies on his or her ability to judge relative speed and distance. Third, the SMEs stated that their job required them to obtain information from multiple inputs and to switch from one task to another very quickly; they are not able to focus on one task and complete it prior to moving on to the next task. For example, a controller must be able to manage the traffic sheet, maintain contact with the ground controller, and communicate with pilots at the same time. Given this, the ability to prioritize duties and multitask was identified as vital for successful job performance. Fourth, since VFR Controllers can work from different towers, their ability to adapt as the situation requires was identified as another essential ability. Often a location change requires Controllers to refresh their plans and approaches because conditions, such as aircraft visibility, change from one location to another. Lastly, the SMEs reported that Controllers need to be able to synthesize large amount of information and make sound decisions in a timely manner.

Spatial Requirements

The spatial tasks reported by the SMEs primarily required *Verbal Reasoning* and *Spatial Visualization* abilities, and relied less on spatial and non-spatial *Memory*, *Mental Rotation*, *Spatial Imagery*, *Processing Speed*, and *Way-finding* abilities (See Table 2 on Page 39).

Key Findings

The VFR Controller position requires individuals to rely primarily on their visual cues to direct air traffic. According to the SMEs, successful job performance relies on (a) having good memory capacity and recall ability; (b) being able to project events into the future; (c) being able to multitask; (d) being adaptable to changing situations; and (e) being able to make timely and accurate decisions.

Cognitive Abilities Common to All Three Sub-Occupations

The Weapons, IFR, and VFR divisions within the ATC occupation share the same primary goal and, because of this, the three divisions have an overlap in the cognitive abilities required to achieve this goal. The primary goal of an ATCO is to identify and track multi-aircraft location, speed, altitude, and direction in order to safely navigate the aircraft within their jurisdiction. In order to achieve this goal, all SMEs agreed that individuals need to be able to (a) identify spatial and temporal relations between multi-aircraft; (b) visualize aircraft in three-dimensional space; (c) think in terms of multi frames of reference; (d) acquire information from multiple sources and use that information in addition to learned protocols and rules to produce an output; (e) plan aircraft sequencing according to priorities and aircraft states; and (f) monitor plan execution, note changes, adapt and change plans as necessary.

Individuals' ability to project in time and space, as well as their ability to perform job-related tasks in realistic job conditions emerged as two other vital factors to consider during selection. An emphasis was placed on both of these elements in all three focus groups and the SMEs believed that differences in these abilities may be what separate good and bad ATCOs. The success of an ATC-related task relies primarily on the

Controllers' ability to correctly project the state of multi-aircraft into the future. Despite this importance, a test of projection ability is currently not included as part of the Canadian Forces' ATCO screening process. The ability to project may utilize one or more of the cognitive abilities listed above (e.g., ability to identify relationship between aircraft, visualize in three-dimensional space). However, it is important to note that isolated assessments of these abilities may not reflect individuals' actual ability to perform a projection task. This notion led to the identification of assessing performance under job-relevant context as the second key ability: individuals must be able to perform multiple tasks simultaneously, effectively, with efficiency, and under time pressure and high workload. While individuals were often able to perform job-related tasks when assessed independently, problems were evident when these same tasks were required to be performed simultaneously and under job-relevant context (e.g., high workload).

Spatial Ability Questionnaire

The SMEs were asked to rate the importance of (1) overall spatial ability, (2) three major areas of spatial ability and (3) the seven spatial ability dimensions identified by Shanmugaratnam and Parush (2012a) on a scale from 1 to 10, with 1 being "Irrelevant" and 10 being "Vital." They then rated the same dimensions on the potential to have negative impact to the occupation if they were ignored during selection, using a scale from 1 to 5, with 1 being "Not at all" and 5 being "Trouble definitely likely if ignored in selection." While the results of significance tests on a small sample of SMEs need to be interpreted with caution as it may not generalize to all ATCOs, these results, especially those with large effect sizes, are invaluable in understanding ATCOs' current problems and needs for further exploration.

Importance Ratings

The SMEs in the three groups identified overall spatial ability as highly and equally important for their job (*M* = 9.80, *SD* = .05) and their ratings did not differ significantly among the three groups ($p > .05$). When asked to rate the importance of the three major areas of spatial ability (i.e., Orientation, Navigation, and Visualization), there was only a significant difference among the three SME groups in the importance rating of Visualization ability. The SMEs in the IFR (*M* = 9.40, *SD* = .89) and VFR (*M* = 9.60, *SD* = .55) groups gave significantly higher rating than SMEs in the Weapons group (*M* = 4.40, *SD* = .55, $F(1.06, 3.2) = 108.5$, $\eta^2 = .96$, $p < .05$), indicating that Visualization ability is more important for the IFR and VFR Controllers than Weapons Controllers. Orientation ability ratings did not differ significantly among the three groups; it was rated as highly and equally important for all Controllers (*M* = 9.40, *SD* = .05, $p > .05$). All Controllers rated navigation ability as the least important and ratings did not differ significantly among the three groups (*M* = 5.60, *SD* = 2.41, $p > .05$).

When the importance ratings were assessed separately for the seven sub-abilities that make up the three major areas of spatial ability, the ratings for each dimension did not differ significantly among the three SME groups ($p >.05$; see Table 3). Across the three focus groups, SMEs rated verbal and reasoning ability (*M* = 8.53, *SD* = 1.30) and processing speed (*M* = 8.53, *SD* = 1.25) as the most important ability for their occupation and spatial imagery (*M* = 7.47, *SD* = 2.03) as least important. When the overall importance of spatial ability was analyzed using the mean rating of all seven dimensions, Weapons, IFR and VFR SMEs provided average ratings of 8.09 (*SD* = 1.34), 7.86 (*SD* = 2.07), and 8.63 (*SD* = 1.31), respectively.

Table 3

Spatial Ability Dimensions, Definitions, and Their Importance Ratings for All Three Sub-Occupations

Ability	Definition	Mean Rating Weapons (n = 5)	IFR (n = 5)	VFR (n = 5)	Cumulative (N = 15)
Spatial					
Mental rotation	Ability to mentally visualize and manipulate things (e.g., maps, images) in 3D.	7.60 (1.14)	6.60 (3.36)	9.00 (1.00)	7.73 (2.22)
Way finding	Ability to mentally navigate around areas.	8.20 (1.10)	7.60 (2.70)	8.60 (0.89)	8.13 (1.68)
Spatial visualization	Ability to visualize how things will appear from various spatial orientations.	8.00 (1.58)	8.00 (1.58)	8.00 (2.00)	8.00 (1.60)
Spatial imagery	Ability to visualize and recall spatial information as they were represented in space at an earlier time (e.g., where and how things were arranged).	7.00 (2.00)	7.00 (2.12)	8.40 (2.07)	7.47 (2.03)
General					
Processing speed	Ability to quickly process information in order to make a decision and/or carryout a task.	8.20 (1.64)	8.20 (1.09)	9.20 (0.84)	8.53 (1.25)
Verbal & reasoning	Ability to use cognitive processing to understand situations and to make decisions.	7.80 (1.48)	8.60 (1.14)	9.20 (1.10)	8.53 (1.30)
Working memory	Ability to temporarily keep visual and spatial information in the forefront of memory so that they can be used in the immediate future (e.g., within 5 minutes).	8.40 (0.55)	9.00 (1.73)	8.00 (0.71)	8.47 (1.13)
Cumulative (n = 5)		8.09 (1.34)	7.86 (2.07)	8.63 (1.31)	

Consequence Ratings

When asked to rate the potential negative consequences if spatial ability was not assessed during selection, there was no significant difference in mean ratings among the three groups ($p > .05$). The SMEs in the Weapons, IFR, and VFR divisions gave mean ratings of 4.40 ($SD = .09$), 4.60 ($SD = .08$), and 5.00 ($SD = 0.0$) out of 5, respectively. There was also no significant difference in the mean consequence ratings between the three groups for the three major areas of spatial ability ($p > .05$). On average, the SMEs rated that a lack of visualization ability would result in the greatest consequence ($M = 4.67$, $SD = .05$), followed by orientation ability ($M = 4.07$, $SD = 0.80$), and navigation ability ($M = 2.53$, $SD = 1.19$).

The SMEs in the three divisions provided a rating of 3 or higher for each of the seven spatial ability dimensions (see Table 4) and the ratings for each dimension did not differ significantly among the three groups ($p > .05$). Overall, SMEs rated processing speed as vital to assess during selection ($M = 4.13$, $SD = 0.74$) while individuals' spatial imagery ability ($M = 3.47$, $SD = 0.99$), mental rotation ($M = 3.53$, $SD = 1.19$) and way finding ($M = 3.53$, $SD = 1.06$) were rated as least important). When the negative consequence of not including measurements of spatial ability was assessed using mean ratings of all seven dimensions, significant differences in ratings were observed between the Weapons and VFR ratings, as well as the IFR and VFR rating, $F(2, 68) = 7.07$, $\eta^2 = .18$, $p < .05$). Not assessing spatial ability may have more dire consequence for VFR Controller selection, compared to the other two groups.

Table 4

Spatial Ability Dimensions Rated on Their Potential to Have Negative Impact to the Occupation if They Were Ignored During Selection the Three Sub-Occupations

Ability	Mean Rating			
	Weapons (n = 5)	IFR (n = 5)	VFR (n = 5)	Cumulative (N = 15)
Spatial				
Mental rotation	3.00 (1.22)	3.20 (1.17)	4.40 (0.55)	3.53 (1.19)
Way finding	3.60 (0.89)	3.00 (1.26)	4.00 (0.71)	3.53 (1.06)
Spatial visualization	3.80 (0.84)	3.80 (0.75)	4.00 (1.00)	3.87 (0.83)
Spatial imagery	3.00 (1.00)	3.20 (0.75)	4.20 (0.84)	3.47 (0.99)
General				
Processing speed	3.80 (0.84)	3.80 (0.40)	4.80 (0.45)	4.13 (0.74)
Verbal & Reasoning	3.60 (1.14)	3.80 (0.98)	4.60 (0.55)	4.00 (1.00)
Working memory	3.60 (0.55)	4.40 (0.80)	3.80 (0.45)	3.93 (0.70)
Cumulative (n = 5)	3.49 (0.92)	3.60 (1.03)	4.26 (0.70)	

Summary and Discussion

The results of the three focus groups revealed four main findings with regards to the ATC occupation and how we can use this information to develop screening tests for ATC and other relevant occupations:

1. Spatial ability is important for performing tasks in the ATCO occupation.
2. ATCOs need to be able to plan aircraft routes and then monitor and make amendments to the plans. Performing these tasks require ATCOs to track multiple aircraft and project in time and space.
3. For selection purposes, spatial ability tests should include more than just measurements of mental rotation ability.
4. Selection tests may need to be administered under job-relevant context.

The results of the focus groups indicate that spatial ability is vital for ATCOs to perform their job tasks. The finding that not assessing this ability may have more dire consequence for VFR Controller selection, compared to the other two groups, makes conceptual sense given that VFR Controllers need to use their spatial ability to monitor aircraft in order to compensate for the lack of technological aids. It is not obvious why the SMEs initially reported Visualization ability is more important for IFR and VFR Controllers than Weapons Controllers. It is possible that the more managerial role that Weapons Controllers adopt as they get promoted require less of this ability as they move from performing active ATC tasks to personnel management and mission planning. However, the SMEs later rated this ability as equally important for all Controllers.

The finding that mental rotation ability may not be as important as other types of spatial ability for ATCOs is important to note. Most tests of spatial ability, including some aspects of the one used by the Canadian Forces as part of the screening process for ATCOs, primarily assess mental rotation ability. However, the focus group discussions revealed that it is more important to consider individuals' ability to plan aircraft-related activities as well as their ability to track and coordinate multiple aircraft. Consequently, administering tests of multi-aircraft tracking and planning abilities may aid in selecting successful ATCO training candidates.

It is important to choose individuals who have the ability to successfully complete the training program because only a small number of applicants can be selected for training. Currently, the course capacity for specialized IFR, Weapons, and VFR training is only 10, 9, and 6 students and students go through 90, 87 and 61 days of coursework, respectively (Canadian Forces, 2013a, b, c). This is in addition to 3 weeks of pre-course training, 7 weeks of Distributive Learning, and additional training at Airspace Control Unit after course training (Canadian Air Division Headquarters, 2004). Furthermore, this training is offered only two or three times per year (Canadian Forces, 2013a, b, c). Given this, a lot of time and financial commitment is made on a small number of individuals in hopes of them passing the training program.

The results of the focus groups make us question the conventional way in which personnel selection tests are selected and administered across many cognitive domains and occupations and indicate possible ways to improve them. The results indicate that the traditional approach to spatial ability testing using any given test of spatial ability may not be sufficient to predict individuals' qualification to perform spatial tasks on the job. Instead, they suggest that the tests may need to be catered to the target job. Furthermore, the results raise questions regarding the way in which personnel selection tests are traditionally administered and suggests that tests may need to be administered in similar context as job tasks since that is where trainees fail to perform. These findings pave the path for the three subsequent laboratory studies that investigate what kind of spatial tests may be appropriate for use in selection (Study 2), the condition under which they should be administered (Study 3), and whether they need to be perceptually similar to the tasks on the job (Study 4).

Chapter 6: Study 2 – Using Tests with Increased Conceptual Similarity to Predict Job-Related Task Performance

The primary purpose of this study was to identify whether job-specific or job-non-specific tests are better predictors of job task performance. This study used nine non-ATC-specific and three ATC-specific spatial ability tests to predict performance on a simplified ATC game that required the safe guidance of aircraft to their predetermined destinations. Given that there is not a single designated test of spatial ability, this study utilized the MRT, the Object Perspective Taking Test, and seven spatial tests from a test battery from ETS as job non-specific tests. These tests have been reported to be reliable measures of spatially-relevant abilities. Validity of a test was not chosen as a criterion for inclusion because it is dependent on the task used. Three ATC-specific tests that assess individuals' ability to either track multiple objects (i.e., MOT Test) or to perform spatial planning (i.e., Projection and Map Planning Tests) were included because Study 1 revealed that these abilities are vital for ATCOs. The ATC-related game used to measure individuals' ATC-related task performance has good face validity and playing the game requires the type of spatial abilities identified in study 1. A pilot study (n = 5) revealed individual differences in performance, thus indicating that it is a good task to use.

Method

Participants

Sixty-five Carleton University undergraduate students (23 male, 42 female) were recruited through the online SONA undergraduate recruiting system to participate in this study. All participants reported that they were fluent in English, had normal or corrected-to-normal vision, and had colour vision. Although some individuals were left handed, they all naturally used their right hand to control the computer mouse. One male participant failed to follow instructions and his data were excluded from analyses. The remaining participants' ages ranged from 17 to 40 years (*M* = 20.50 years, *SD* = 4.30 years). This study lasted approximately 3 hours and each student received three research participation credits.

Stimuli and Apparatus

This study required participants to complete nine job-non-specific spatial ability tests, three job-specific spatial ability tests, and one spatial task. Details pertaining to these tests and task and their administration are discussed below.

Job-Non-Specific Spatial Ability Tests

A previous literature review on spatial ability (See Shanmugaratnam & Parush, 2012a) identified seven spatially-relevant abilities. ATCOs then rated the importance of these abilities for ATC in Study 1. Of the seven abilities, only the abilities that were considered to be uniquely spatial were used in Study 2. Consequently, assessments of *processing speed*, *verbal and reasoning ability*, and *working memory* were not included in this study. Although *spatial imagery* has a spatial component, this ability was not assessed due to support from the literature that this ability is actually comprised of other assessed abilities (i.e., visualization and mental rotation; Thompson, Slotnick, Burrage, & Kosslyn, 2009). Furthermore, the SMEs in Study 1 rated this ability as the least important ability for ATCOs and that omitting assessments of this ability in selection would have the least consequence. The remaining three abilities, namely *visualization, mental rotation*, and *way-finding*, were assessed in Study 2. However, hereafter, they will be referred to as *visualization*, *spatial orientation*, and *spatial scanning* abilities, respectively, in order to utilize the nomenclature used by ETS. Despite the name differences, the corresponding abilities are similar (See Appendix H for definitions).

The job-non-specific spatial ability tests administered in this study were seven tests from the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976a) by ETS, the Revised Vandenberg-Kuse MRT (Peters, 1995), and Hegarty et al.'s (2008) Object Perspective Taking Test. The tests from ETS are organized into three spatially-related areas: Visualization (Form Board, Paper Folding, and Surface Development tests), Spatial Orientation (Card Rotations and Cube Comparisons tests), and Spatial Scanning (Maze Tracing Speed and Choosing a Path tests). These tests have been administered by ETS for over 30 years and ETS reports them to be reliable measures. They were obtained and used under licence from ETS (See Appendix I). The MRT was included in this study given that it is a classic test used to measure spatial ability of individuals. The Object Perspective Taking Test (Hegarty et al., 2008) measures individuals' spatial orientation ability and was included in this study because this test requires individuals to orient themselves in space and visualize information from different perspectives; this ability is similar to what is required by ATCOs when they need to visualize information from a third-person perspective, for example, when providing non-radar descriptions, planning air traffic flow, and determining intercept solutions. This test also has better face validity than the spatial orientation tests from ETS.

All the job-non-relevant spatial ability tests were administered in the paper-and-pencil format and eight of the nine tests were administered in two parts; participants were given the first part and when the time expired, that part was collected and the second part was administered. Participants were instructed to perform their best on the test and to check their answers to ensure accuracy if they completed the given a part before the time expired. Materials were collected only after the time expired. Three different timers were developed using PowerPoint and set depending on the duration of each part of the test (i.e., 3 mins, 6 mins, or 8 mins). An audio recording (i.e., "end of part 1" or "end of test") indicated when the time expired.

After each test, participants were asked to complete a simplified paper-and-pencil version of the NASA-Task Load Index (NASA-TLX) V2.0 (Hart & Staveland, 1988; See Appendix J). These ratings were used to understand how participants felt about each test. Details pertaining to each of the job-non-specific spatial ability tests are discussed below. The statistics reported for each of the ETS test were provided by Ekstrom et al. (1976a) based on two large-scale studies with naval recruits (N = 625-746 for each session in one study and N = 542-574 for the second study), and studies with grade 11 and 12 male (N = 288-300) and female (N = 317-329) students.

Spatial Visualization Tests

Form Board Test. The purpose of this test is to determine individuals' ability to visualize and mentally rotate shapes. This test required individuals to determine how various shapes will look when they are assembled. In this test, individuals were presented with a target shape at the top of the page and several problems below. Each problem required individuals to indicate which of the five small shapes provided were required to construct the larger target shape by placing a "+" under shapes that are required and "-" under those that were not required (See Appendix K). There were two parts for this test with 24 problems each and individuals were provided 8 minutes to complete each part. The score for this test was obtained by tallying the number of correct responses minus the number of answers marked incorrectly, for a maximum score of 240. A higher score indicates better performance. Ekstrom et al. (1976a) report that in a study with male and female college students (N = 46), the mean test score was 124.8 (SD = 38.3) with a reliability of .81.

Paper Folding Test. This test assesses individuals' ability to visualize two-dimensional spatial information. This test presented individuals with images of a square sheet of paper, first folded in one or more ways and then hole punched. Individuals were asked to visualize how the hole punched sheet of paper will appear when unfolded and select an answer from one of the five options provided on the right by marking an "X" through the appropriate figure (See Appendix L). This test was administered in two parts with 10 items each and individuals were provided 3 minutes per section. The score for this test was the number of questions answered correctly minus the number of incorrect answers. The maximum score for this test is 20, with a higher number indicating better performance. Ekstrom et al. (1976a) reported that when this test was administered to male and female 11th and 12th grade students, the mean test score for males was 11.5 (SD =3.7) with a reliability of .75. The mean test score for females in same sample was 10.4 (SD = 3.7) with a reliability of .77. These researchers found that male and female college students (N = 46) had a slightly higher mean score (M = 13.8, SD = 4.5) with reported reliability of .84.

Surface Development Test. This test evaluates individuals' ability to visualize the transformation of a paper cut-out into a three-dimensional folded object presented on paper. The left side of the page presented an unfolded outline of a shape containing marked dotted lines on the inside to indicate where the folding should occur. Five of the outside edges of the shape were marked with numbers. The middle of the page presented a folded image with letters corresponding to each of the visible edges. An "X" was marked on one of the faces on the unfolded image and a corresponding "X" was marked on the same face on the folded image. The right side of the page provided a table with numbers from 1 to 5 on one column and blank spaces beside each number on the next column (See Appendix M). Individuals were required to visualize the image on the left being folded along the dotted lines to make the shape in the middle. Then they were asked to determine which numbers on the unfolded image correspond with which letters on the folded image and mark the answers on the blank space next to each number on the table. This test was administered in two sections with six problems each and individuals are provided 6 minutes to complete each section. A score for this test was obtained by tallying the number of correctly identified number and letter pairs and applying a .25 penalty for the number of incorrect answers, for a maximum score of 60. For example, someone with 40 correct answers and 16 incorrect answers received a score of 36 (i.e., 40 – .25*16). A higher score indicates better performance. Ekstrom et al. (1976a) report that in a study with male and female college students ($N = 46$), the mean test score was 43.68 (SD = 15.1) with reliability of .90.

Spatial Orientation Tests

Card Rotations Test. This test measures individuals' ability to mentally rotate various two-dimensional images and determine whether two images are the same or different. In this test, individuals were presented with a target figure on the left and eight comparison figures on the right. The comparison figures were either planar rotations or reflected images of the target figure. Each comparison figures had two small checkboxes under it with the labels "S" and "D." Individuals were required to mark whether or not the comparison figures were the same as the target figure by placing a check mark inside the box with the corresponding label (S for same or D for different). A figure was considered the "same" if it was rotated on a plane and "different" if it was a mirrored image. This test had two sections with each section containing 10 rows of target figures with 8 response options each (See Appendix N). Participants were permitted 3 minutes per section, for a total of 6 minutes to complete the test. The score for this test was calculated by subtracting the total number of incorrect responses from the total number of correct responses. The maximum possible score for this test is 160 and a higher score indicates better performance. Ekstrom and colleagues (1976a) reported only the mean test score for the first part of this test and stated that male naval recruits received a score of 44.0 (SD = 24.6). When they administered an earlier version of this test to male and female 11th and 12th grade students, the reliabilities were .86 for male students and .89 for female students. In two studies with male and female college students (N = 46 for one study, N = 99 for the other study), they reported reliabilities of .80 and .83, consecutively. The developers did not provide mean scores for the complete test.

Cube Comparisons Test. The purpose of this test is to assess individuals' ability to mentally manipulate three-dimensional images presented on paper (i.e., in two-dimensions) and determine whether two images are similar or dissimilar. In this test, two cubes were rotated differently around the horizontal and vertical axis and were presented side-by-side, with each cube resting a face along the horizontal axis. Although all six faces of the cube were marked with a letter, only three of faces were exposed. Sometimes both cubes shared the same relationship between the various letters marked on the face (e.g., “J” is to the left of “S” and “K” is above both these letters) and other times this relationship did not exist. Beneath each cube were the letters “S” and “D” with checkboxes next to them. Individuals were required to compare the two cubes, with reference to the exposed and concealed letters, and place a mark in the checkbox corresponding to “S” or “D” to indicate whether the cubes were the same or different, respectively (See Appendix O). In total, this test consisted of 42 pairs of cubes, 21 pairs per part, and individuals were permitted 3 minutes per part. A final score for this test was calculated by tallying the total number of correct responses and subtracting the number of incorrect responses. The maximum score for this test is 42, with a higher number indicating better performance. Ekstrom and colleagues (1976a) report that when this test was administered to male and female 11th and 12th grade students, the mean test score for males was 23.5 (*SD* = 6.6) and 21.5 (*SD* = 6.6) for females. Both male and female samples had test reliability of .77. In another study with male and female college students (*N* = 46), the mean test score was 22.7 (*SD* = 9.4) with a reliability of .84.

The Revised Vandenberg-Kuse Mental Rotations Test. The purpose of the revised MRT (Version MRT-A; Peters, 1995), which is essentially an Autocad-redrawn version of the original test by Vandenberg and Kuse (1978), is to determine individuals’ ability to manipulate three-dimensional images presented on paper (i.e., in two-dimension) and judge their similarity. This test presented individuals with drawings of three-dimensional target figures on the left and four comparison figures on the right. Of the four comparison figures, two were target figures rotated around the vertical axis while the other two were distracter figures (See Appendix P). Individuals were required to identify the two comparison figures that were rotated images of the target figure by marking an

"X" over them. This test consisted of two parts with 12 problems each, for a total of 24 questions. Participants were provided 3 minutes to complete each part. Individuals were required to correctly identify both correct answers for each problem to obtain a point; no partial points were awarded. A score for this test was calculated by tallying the total points and higher scores indicated better performance. Maximum score for this test is 24. Vandenberg and Kuse (1978) report that the original version of the test has an internal consistency of .88 and a test-retest validity of .83. Using very similar stimuli from the Shepard and Metzler (1971) version of MRT test, Hooven, Chabris, Ellison, and Kosslyn (2004) have found reliability between .66 and .76 for error rate and reliability between .78 and .83 for response time. Peters and colleagues (1995) report a mean score of 10.8 (*SD* = 5.0) for male and female university students. This test is most commonly used in the literature to measure individuals' spatial ability.

Object Perspective Taking Test. The Object Perspective Taking Test (Hegarty, et al., 2008) measures individuals' spatial orientation ability. Each question on this test presented individuals with a basic town map with only seven objects and no visual cues to represent the relationship between the objects. Individuals were asked to imagine standing by one of the objects while facing another and were asked to indicate the direction of a third object. Each question was presented at the top half of the paper while the bottom half contained a drawing of a circle. Within the circle, the center was marked as the location of the object the individual was asked to imagine standing next to and a vertical arrow pointed up from the center. The tip of the arrow was marked as the location of the object the individual was asked to imagine facing. Individuals were asked to draw a line from the center of the circle to the perimeter of the circle to represent the direction in which they believed the target object was located (See Appendix Q). In total, this test contained 12 items and individuals were provided 5 minutes to complete the test. A score for each question was calculated by measuring how much the marked answer deviated from the correct answer (in degrees). A final score was determined by averaging the 12 scores and a lower score indicated better performance. Hegarty and Waller (2004) have found evidence to indicate that this test assesses something other than mental rotation ability and report that the reliability of this test is .79. The authors also state that this test is a valid measure of spatial orientation ability given that this test loaded on the same factor as Money's Road Map Test (Money, Alexander, & Walker, 1965), which is a measure of perspective-taking ability.

Spatial Scanning Tests

Maze Tracing Speed Test. This test provides individuals with a number of interconnected mazes and requires them to use a pen to draw a path through each maze as quickly as they could without crossing any printed lines within the maze (See Appendix R). This test was administered in two parts with 24 test mazes each, for a total of 48 questions. Individuals were provided 3 minutes to complete each part. The score for this test was the total number of correctly completed mazes, with a higher score indicating better performance. Ekstrom et al. (1976a) report that when this test was administered to male and female 11th and 12th grade students, the mean test score for males was 22.4 (SD = 6.3) with a reliability of .91. The mean test score for females in same sample was 20.7 (SD = 6.0) with a reliability of .89. In a study with male and female college students (N = 46), that the mean test score was 30.3 (SD = 8.3) with a reliability of .94.

Choosing a Path Test. This test assesses individuals' ability to choose a correct path from five path options provided. In this test, individuals were shown five squares that were linked to each other via interconnected vertical and horizontal line segments (i.e., paths). The starting and finishing points of each path were indicated with an "S" and "F," respectively, and were placed at the top within each square. Individuals were required to visually trace the path from the starting location to the finishing location, while ensuring that it passed around a circle at the top. Directional changes along the path were permitted only at intersections indicated with a small black circle. When these rules were followed, only one of the five path options provided met both requirements. Individuals were required to mark their response by blackening the space at the lower right hand corner of the correct square (See Appendix S). This test was administered in two parts with 16 questions each, for a total of 32 questions. Individuals were provided 7 minutes to complete each of the two parts. A score for this test was calculated by tallying the number of correctly solved problems, with higher scores indicating better performance. Ekstrom and colleagues (1976a) report that in a study with college students (N = 46), the mean score on the test was 15.5 (SD = 7.0) and the reliability was .77.

Job-Specific Spatial Ability Tests

Three tests were used to measure individuals' job-specific spatial ability: Map Planning, MOT, and Projection. The Map Planning Test was a licenced paper-and-pencil test from ETS, while MOT and Projection Test were programmed in C++ and were based on a MOT Test used by Hunter (2011) for her doctoral dissertation on multiple object tracking. Hunter based her test on the traditional MOT paradigm. While the MOT and Projection Tests in this dissertation were perceptually identical, the instructions provided to the participants dictated whether it was a MOT or Projection Test. The MOT and Projection Tests used in this study differed from Hunter's (2011) test in four ways:

1. **The number of stimuli**: The number of stimuli was adjusted for two reasons: First, Hunter used 14 stimuli and required participants to track six of them. For the MOT Test, it was desirable to have equal number of targets and distractors (as used by Pylyshyn, & Storm, 1988; Pylyshyn, 2004; Scholl, Pylyshyn, & Feldman, 2001; Thornton, Bülthoff, Horowitx, Rynnin, & Lee, 2014), therefore the total number of stimuli was reduced to 12 for both the MOT and Projection Tests to make both tests comparable. A small-sample pilot study (*N* = 5) revealed that the reduced number of stimuli did not result in a ceiling effect and individual differences were still observed (e.g., Projection Test accuracy ranged from 50% to 79%). Second, the ideal ratio of targets to distractors in the MOT Test for Study 3 was 1:2. Bettencourt and Somers (2009) have demonstrated that increasing the targets to distractors ratio, from a 1:1 to roughly 1:2 ratio resulted in significant performance impairments, thus indicating increase in workload. The 1:2 ratio was also recently used by Thornton et al. (2014) for a study on interactive MOT. The number of stimuli used by Hunter would have resulted in participants tracking seven stimuli from a total of 21, which would have cluttered the screen. Participants in the aforementioned pilot study reported extreme difficulty and frustration when attempting the MOT and Projection Tests with 21 stimuli.
2. **Stimuli identifiers (i.e., call signs)**: While similar alpha-numeric call signs (e.g., AC671) representing aircraft prefix and number were used as stimuli identifiers in both studies, all target stimuli in Hunter's research were marked with the "CA"

prefix and distractor stimuli were marked with one of three other letter prefixes in uneven distribution (2 for “AG”, 5 for “DL,” and 1 for “WF”). This may have made some aircraft more salient than others and could have influenced individuals’ tracking performance. In order to address this problem, six different types of airline prefixes were used in this study, with each airline having two different aircraft on the screen with different aircraft numbers. Individuals were asked to track planes starting with “AC,” “DL,” and “PO” identifiers.

3. **Tracking duration**: Hunter used tracking durations between 25 and 35 seconds to minimize participants predicting the onset of an occlusion in her study. Since this study did not include an occlusion period, tracking durations of 15 and 30, and 45 seconds were used in the pilot study ($N = 5$) with the intention of including all three durations in the study. Participants reported that the 45-second trials were too long and that they started to lose focus. Therefore, only the 15 and 30 second trials were used. 12 randomly distributed trials for the MOT Test lasted 15 seconds while the other 12 trials lasted 30 seconds. All 12 trials for the Projection Test lasted 30 seconds since the pilot study revealed that 15 seconds was too brief to identify and track all six aircraft. Participants predicting the end of a trial was not a factor for this test since all participant responses were made during the trial.
4. **Aircraft speed**: Hunter’s stimuli moved at speeds ranging from 4.8 to 16.8 pixels/second, while the stimuli in this study moved at speeds between 12.3 and 36.2 pixels/second. The speed was adjusted since there were no occlusions in this study and the slow moving stimuli were too easy track.

Multiple Objects Tracking (MOT) Test

The MOT Test presented participants with 12 vertically and horizontally moving white dots (13 pixel in diameter) against a black background on a 17" inch desktop computer screen (See Appendix T). The dots represented aircraft and each dot had a five alpha-numeric call sign located to the upper right corner. The screen was divided into four, equal sized, quadrants represented by white lines and the dots moved to and from these quadrants at speeds ranging from 12.3 pixels/second to 36.2 pixels/second. The speed, movement direction, and starting location of the aircraft were randomized across trials but were the same for each participant.

In order to take this test, participants were seated at a desk and the monitor was placed 17" away from them. The stimuli subtended a visual angle of 0.51 degrees when it was at the center of the screen. Individuals were asked to visually track six predefined aircraft as they moved across the screen, after which the identifiers of the aircraft disappeared and the aircraft stimuli froze on the screen. The identifiers were removed in order to ensure that participants did not use that information to perform the identification task. Individuals were asked to quickly, but accurately, click on the six aircraft that they were asked to track. The six target aircraft were the same across all 24 trials: there were two of each of the aircraft marked by "AC," "DL," and "PO" prefixes. After making their responses, participants were asked to press the 'enter' key on the keyboard when they were ready to move on to the next trial. Participants completed 12 15-second trials and 12 30-second trials but the duration of the trials was not revealed to them. Following Hunter's (2011) procedure, the 24 trials were divided into three blocks to represent early, mid, and late trials (Block 1 = trials 1-8, Block 2 = trials 9-16, and Block = trials 17-24).

Participants were permitted to take a brief break in between trials if they needed to rest their eyes. Participants were asked to keep their hand on the mouse during the visual tracking period but were asked not to track the aircraft with the cursor. If they found this to be difficult, they were permitted to hover their hand over the mouse until it was time for them to make the responses. The program automatically tracked response-related information (e.g., the identification of selected aircraft, when each selection was made).

Projection Test

The Projection Test was visually indistinguishable from the MOT Test and only differed in the instructions provided to participants. Participants were asked to visually track the trajectory of all 12 aircraft and identify the pairs that will collide (i.e., have dots fully or partially overlapping) if they continued along the same path by clicking on all affected aircraft. As with the MOT Test, the 12 trials were divided into three blocks.

Map Planning Test

The goal of this test is to measure individuals' ability to navigate within a paper map. For each set of questions, individuals were presented with a 7X6 grid map that was alphabetically labelled (from A to Z) at each outer intersection. Individuals were asked to visually navigate from one letter to another while ensuring that they take the shortest route possible and pass by only one building. Buildings on the map were indicated with small numbered squares. Certain locations on the grid contained roadblocks, indicated by black circles, and prevented individuals from taking that path (See Appendix U). Each question provided a starting and ending location (e.g., "J" to "S") and individuals were asked to write the number of the building they passed by beside the question. This test had two parts with 20 questions each, for a total of 40 questions. Individuals were provided with 3 minutes to complete each part. The score on this test was the number of correct answers, with higher scores indicating better performance. Ekstrom and colleagues (1976a) report that when this test was administered to male and female 11th and 12th grade students, the mean test score for males was 19.6 (*SD* = 6.0) with a reliability of .80. The mean test score for females in same sample was 19.0 (*SD* = 5.8) with a reliability of .75. In a study with male and female college students (*N* = 46), the researchers reported a mean test score of 25.0 (*SD* = 6.1) with reliability of .79.

Spatial Task

The spatial task in this study was the Plane Control HD V.1.9 game obtained through iTunes. This game was released in July 2011 and it was no longer available for free download while conducting this research, thus reducing the possibility that participants were familiar with it. This game was administered on a Macintosh laptop computer with 15-inch colour LCD monitor with 1680 X 1050 resolution. The game required participants to perform simplified versions of the tasks performed by ATCOs. Participants were presented with a bird’s eye view of a portion of a simulated world map and were asked to direct planes from one location to a designated location, indicated with crosshair-like symbol, by matching the colour of the plane with the destination symbol colour. A new plane appeared on the screen every 5 seconds at random locations and with random destination points. Participants were instructed to route each plane to its destination as soon as it appeared on the screen. If the plane was not routed, it stayed on the same spot for 14 seconds before placing itself in a holding pattern (i.e., it moved on the screen but not towards a particular destination) until it was routed. Directing aircraft required the use a computer mouse to click on a plane and drag it to the destination while ensuring that two planes do not share the same horizontal space at the same time (i.e., collide). The vertical space was not factored into the game. If two aircraft were too close to each other, the game alerted participants via an alarm sound and by visually indicating radiating circles around the affected planes (See Appendix V). Participants were instructed to re-route one or more planes only if they believed a collision would occur. Amendments to the chosen paths were made by clicking on an aircraft and drawing a new path. Each trial of the game ended when the first plane crash occurred. Participants completed three trials of the game and maximum duration of each trial was 5 minutes. Camptasia by TechSmith was used to video capture gameplay.

In order to ensure that this game simulated similar cognitive working conditions as the real ATCO environment, participants were required to perform a spatial comparison task while playing the game. This task was administered to ensure that participants did not have their undivided attention on the primary task, much like in real life ATCO working environment. For this task, individuals were auditorily presented with two times (e.g., 3 o'clock and 9:30) and were required to visualize the smaller angles formed by the hour and minute hands for the two times and determine whether they were the same or different. They were asked to verbally state "same" or "different." The angles formed by the two comparison times were either the same or were different by 30 degrees (e.g., 12:30 and 7 o'clock). Although participants were asked to work quickly and provide an accurate answer, answers were permitted until a new time pair was presented. On average, new comparison questions was administered every 30 seconds, with a range of 20 to 40 seconds. Paivio (1978) found that such clock comparison tasks require spatial imagery processing and similar tasks have been used in literature as a visuospatial task that utilizes individuals' spatial processing ability (e.g., Formisano et al., 2002; Grossi, Modafferi, Pelosi, & Trojano,1989; Johannsdottir & Herdman, 2010). A pilot study (*N* = 5) revealed individual differences for average number of planes landed (11 to 26), game duration (103 to 190 seconds), number of close calls (5 to 22), and number of path changes (<1 to 5).

Procedure

Each participant completed this study individually in a room that had access to a desktop computer with a mouse and keyboard, as well as a Macintosh laptop with a mouse. Any questions raised by the participant were answered prior to administering the informed consent form, demographics questionnaire (See Appendix W), tests and game.

Administering all the tests and the game required roughly 6 hours of each participant's time, therefore participants were randomly assigned to one of two groups and were administered the tests with two non-job-specific tests, both job-specific tests, and the ATC game overlapping. Testing in multiple sessions was not preferred due to high likelihood of participant loss. The tests from the spatial visualization, orientation and scanning categories were distributed between the two groups, with at least one test from each category in each group. Participants were randomly assigned to a group. One group of participants were administered the following tests: MRT, Card Rotations, Maze Tracing, Surface Development, Cube Comparison, and Choosing a Path. The second group of participants were administered the following tests: MRT, Card Rotations, Form Board, Map Planning, Paper Folding, and Spatial Orientation. Administration order of all tests and the task were counterbalanced using the balanced Latin Square approach.

Participants were provided written instructions for all the tests and the game. Practice test items were provided within the instructions for the non-job-specific tests. Two practice trials were provided for the Projection Test and three trials were provided for the MOT Test given that the pilot study ($N = 5$) revealed that participants used the first trial of the MOT Test to fully understand the task and treated the latter two as real practice. Participants were provided two practice questions for the spatial comparison task within the game instructions and three more questions during a one-minute practice game play. All instructions were provided separately from the tests and game and individuals had unlimited time to read and work through the sample problems. The tests and the game were administered once individuals indicated they understood the instructions. Eight of the nine non-job-specific tests had two parts, therefore each part was administered separately to ensure that participants did not go back to the previous part or skip ahead to the second part and use more than the time allotted for each section. Participants were instructed to perform their best on all the tests and the game. Breaks between each test were permitted and individuals were allowed to leave the testing area during their breaks. At the end of each test, participants were given the NASA-TLX (Hart & Staveland, 1988). The instructions for the next test was presented either on the screen or was placed next to the participants and they were asked read the instructions when they were ready. Participants were verbally debriefed at the end of the study and were given a written debriefing sheet for them to take with them (See Appendix X). Participation credits were granted online at the end of each day.

Measures

Job-Non-Specific Spatial Ability Tests and Map Planning Test. A score for each of the nine job-non-specific and the Map Planning tests was calculated according to instructions provided by the test developers (See Stimuli and Apparatus section, p. 39).

MOT Test. Two performance measures were taken from this test: mean correct response and mean response time. A mean correct response score was calculated for each block by averaging the total number of aircraft correctly identified across the eight trials. A higher score indicates better performance. A mean *response time* measure for each block was obtained by averaging the time (in seconds) taken to make each response across the eight trials, with less time indicating better performance.

Projection Test. Three performance measures were taken from this test: total number of correct response, mean lead time, and mean false alarm rate. A correct response score for each block was attained by tallying the number of collisions correctly identified across the four trials. A higher score indicates better performance. Mean lead time, which is a measure of how far in advance collisions were detected, was computed for each block by subtracting the time at which potential collisions were identified from the time at which they actually occurred and by averaging it across four trials. For example, if a potential collision was identified between two aircraft 9 seconds into the trial and the collision was set to occur at 23 seconds, then the score for this single instance was 14 (i.e., 23-9). Longer lead times indicate better performance. Mean false alarm rate for each block was calculated by dividing the total number of incorrect responses (i.e., non-colliding aircraft pairs identified) by the total number of aircraft pairs clicked and averaging it across the four trials. A low score indicates better performance.

Spatial Task. Four performance measures were taken from the ATC game: (1) the number of planes landed; (2) duration of game play; (3) number of path changes made; and (4) the number of close calls (i.e., alarms triggered). The number of planes landed was provided by the game while the other three measures were manually calculated by reviewing the video capturing of game. Game duration was calculated using the time markers provided by the software. Actions were counted as path changes if a routed plane was rerouted using a different path. The game was sensitive regarding maintaining a certain degree of separation between aircraft, therefore only alerts that lasted more than three fast “beeps” were counted as close calls. Additionally, any triggered alerts that were not the fault of the participant, for example due to a new plane appearing at a spot close to another plane, was not counted as close calls. Higher values for the number of planes landed and game duration and lower values for the number of path changes and alerts indicate better performance.

ATC Clock Task. The number of correct and incorrect responses, the number of questions individuals missed, and accuracy rates (correct responses/total responses) for each of the three trials were obtained for each participant. Given that the clock task required participants to report whether two auditorily presented times were the same or different, the chance level for accuracy is 50%.

Results

Data Cleaning and Assumptions. The data was examined using SPSS (Statistical Package for the Social Sciences) V.21. The 64 participants were evenly divided between the two groups (32 per group). Missing data, normality, practice effects, sphericity, multicollinearity, and singularity were examined and dealt with accordingly (see

Appendix Y). Once the initial data cleaning was complete, paired-sample *t*-tests were used to examine whether differences existed between the two groups of participants for any of the overlapping tests and game performance measures. If the Homogeneity of Variance assumption was violated for any of the *t*-tests performed, then the Welch-Satterthwaite method was used for correction. The results revealed no significant differences between the two groups for any of the overlapping measures ($p > .05$). This resulted in conducting three separate regression analyses using the same models (one for overlapping variables, one for Group 1, one for Group 2).

Job-Non-Specific Spatial Ability Tests

The mean scores for the job-non-specific spatial ability tests are provided in Table 5, along with test developers' values reported in the Stimuli and Apparatus section (p. 56). The test developers' scores reported here are for the college student sample, except for the Card Rotation test, which has values for naval recruits. The maximum score for each test is provided in parentheses next to the test name in the first column. The average number of questions completed by participants in this study is provided in the sixth column. In general, participants in this study performed similarly to that of test developers' participants, with the exception of the Surface Development and Object Perspective Taking tests; individuals in this Study performed slightly worse on both tests.

Table 5

Job-Non-Specific Spatial Ability Test Scores Comparing Test Developer and Study 2 Mean Scores

Tests	Test Developer Score		Study 2 Score		Number of Questions Completed[4]	
	M	*SD*	*M*	*SD*	*M*	*SD*
Spatial Visualization						
Form Board (240)	124.80	38.30	127.79	48.78	155.13	58.77
Paper Folding (20)	13.80	4.50	10.09	3.63	15.72	3.21
Surface Development (60)	43.68	15.10	29.30	16.63	48.13	12.55
Spatial Orientation						
Card Rotation (160)	44.00[1]	24.60	98.59	26.55	108.36	31.16
Cube Comparison (42)	22.70	9.40	20.93	7.20	31.81	8.21
MRT (24)	10.80[2]	5.00	9.47	4.72	15.03	5.02
Object Perspective Taking (180)	24.53[3]	14.29	48.59	35.41	10.78	1.81
Spatial Scanning						
Maze Tracing (48)	30.30	8.30	25.72	7.57	24.66	8.22
Choosing Path (32)	15.50	7.00	8.35	5.28	17.43	7.22

[1] Scores reported for only Part 1 of the two part test.
[2] values from Peter et al. (1995).
[3] values from Hegarty and Waller (2004).
[4] Total number of questions for each test is equal to maximum test scores listed in column 1, except for Object Perspective Taking test, which has 12 questions in total.

Job-Specific Spatial Tests

Map Planning Test. Participants in this study obtained a mean score of 21.35 (*SD* = 6.64) out of a possible 40, which was slightly lower than that of the participants in the ETS' sample (*M* = 25.0, *SD* = 6.1). On average, participants completed 24.32 (*SD* = 5.76) out of 40 questions.

MOT Test. A paired sample *t*-test revealed that participants were more accurate in identifying the six target planes in the MOT Test for the 30-second trials (*M* = 64.20, *SD* = 4.90) than the 15-second trials (*M* = 61.66, *SD* = 5.85), $t(63) = -5.07$, $p < .01$ $\alpha = .05$). Despite this difference, all trials were analyzed in blocks in subsequent analyses since overall accuracy was the primary interest in this study. There was no significant difference between the two trial durations in the mean response time taken to make a response ($p > .05$).

Participants' performances across the three blocks were compared. On average, individuals successfully identified five of the six tracked aircraft and they took approximately one second to identify each aircraft. There was no significant difference in performance between the three blocks for the mean number of correct response ($p > .05$). Consequently, a single average score was obtained and used in subsequent analyses. There were significant differences in response time between Blocks 1 and 2 as well as Blocks 1 and 3 (See Table 6). Since there was no difference in performance between Blocks 2 and 3, these reaction times were averaged to create a single variable.

Table 6

MOT Test Performance for the Three Blocks of Trials

	Trial Blocks									
	1		2		3					
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Correct Response ($N = 64$)	5.19	.48	5.28	.48	5.27	.46	2	1.97	.03	.14
Response Time ($N = 64$)	1.20	.33	1.05	.29	1.02	.38	1	34.13	.35	< .01

Projection Test. Table 7 presents the mean scores for the Projection Test. On average, individuals identified 15 of the 24 total plane crashes (5 for each of Blocks 1, 2, and 3) and they were able to identify it between 2.6 and 4.2 seconds prior to the crash. However, roughly 41% of the plane pairs identified were not planes that crashed. There was a significant difference in the mean number of correct responses between Blocks 1 and 3 and Blocks 2 and 3 (no difference between Blocks 1 and 2). There was also a significant difference between all three blocks in lead time. No significant difference was observed between the three blocks for false alarm rates ($p > .05$). For each variable, blocks with significant differences in performances were retained as separate variables in subsequent analyses.

Table 7

Projection Test Performance for the Three Blocks of Trials

	Trial Blocks									
	1		2		3					
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Correct Response (*N* = 64)	5.31	1.36	5.27	1.19	4.80	1.21	2	4.65	.07	.01
Lead Time (Seconds, *n* = 59)	3.09	1.68	4.17	1.69	2.58	1.36	2	38.64	.39	< .01
False Alarm (%, *N* = 64)	43.05	21.00	42.71	19.77	38.68	24.19	2	2.36	.04	.10

Spatial Task

Given that planes appeared at random locations on the screen, sometimes a new plane appeared on top of an existing plane. Participants were instructed that if this occurred, they would only be able to select the top plane until the two planes separated. Despite these instructions, three individuals attempted to repeatedly select the bottom plane and this subsequently lead to a collision between the two planes. Performances on the affected trials, one for each participant, were not included in the analysis. None of the participants stated that they were familiar with this game, although some participants reported playing other ATC-related games on their phone.

There was no significant difference in ATC game performance between the two groups ($p > .05$). For the three trials of the game, participants played the game for 59 to 223 seconds and, in that time, safely landed 7 to 40 planes, received 3 to 23 crash warnings, and made 0 to13 path changes. Participants landed significantly more planes on the third trial, compared to the first two trials. They also played the game for a significantly longer period of time and had significantly more path changes in the third trial, compared to the first trial (no significant difference in performance between Trials 1 and 2 or Trials 2 and 3). There was no significant difference between the three trials for the number of alerts individuals received ($p > .05$; see Table 8).

Table 8

ATC Task Performance for the Three Trials of the Game

	Trials									
	1		2		3					
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Planes Landed (*n* = 58)	15.34	8.55	16.52	8.47	19.67	9.11	2	5.27	.09	< .01
Duration (*n* = 59)	130.07	47.56	140.07	50.85	153.88	46.92	2	4.64	.07	.01
Path Changes (*n* = 59)	1.19	.96	1.25	1.05	1.60	1.17	2	4.31	.07	.02
Alerts (*n* = 59)	9.00	6.02	9.95	6.42	11.44	5.93	2	3.10	.05	.05

Clock Task

Individuals performed above, at, and below chance level for the first, second, and third trials of the game, respectively. The number of incorrect responses and accuracy rates significantly differed between Trials 1 and 2 as well as between Trials 1 and 3 (no significant difference between Trials 2 and 3). Participants had less incorrect responses for the first trial, compared to the latter two, and, consequently, were more accurate for the first trial. There were no significant differences between the three trials for the number of correct responses or number of missed questions (p > .5, see Table 9).

Table 9

Clock Task Performance for the Three Trials of the ATC Game

	Trials									
	1 (N = 64)		2 (N = 64)		3 (N = 64)					
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Correct	2.47	1.35	1.96	1.35	2.04	1.59	2	2.66	.04	.07
Incorrect	1.19	.85	1.89	1.50	2.16	1.10	2	12.49	.17	< .00
Miss	.12	.35	.07	.29	.16	.51	2	.80	.01	.44
Accuracy (%)	63.02	25.96	50.70	30.80	43.64	27.26	2	8.75	.12	< .01

Predictors of ATC-Related Task Performance

Data cleaning resulted in 18 variables (See Appendix Z), which were too many to include in modeling since a reliable model with 18 measures requires 149 participants (assuming R^2 = .13, power = .80, α = .05). Therefore, only the significant predictors of game performance using the "enter" and "stepwise" methods of regression analyses were used for model building. This resulted in excluding Cube Comparisons, Choosing a Path, Paper Folding, and Object Perspective Taking tests. All blocks of Projection or MOT variables were retained even if only one or two blocks appeared as significant predictors (See Appendix Z for information on blocking).

Fourteen regression models were developed using the remaining 14 tests to predict game performance (See Appendix AA for models). For each significant model with significant F_{Change}, the slope and significance of that slope for each test was used to assess the benefit of each test. While 135 participants were required for a reliable model with 14 measures, there were only 58 and 59 participants for the two groups in this study. Yet, all significant models had either a medium (R^2 = .13) or large effect size (R^2 = .26) according to Cohen (1988). Therefore, we can be confident about the results despite the small sample size. The results are presented below according to how they were analyzed.

Analysis Using Overlapping Variables. Models 1 and 3 were significant predictors of the number of planes landed for Trials 1 and 2 combined, the number of path changes made for Trial 1, and the game duration for Trial 1 (See Appendix AB for results of Model 1). No models significantly predicted any of the other performance measures (See Table 10). Based on these results, Card Rotations Test, MOT Response Time, and Projection Correct Response were retained for Study 3.

Table 10

Effect Sizes of Significant Models (R^2, Adjusted R^2) and Significant Predictors of Game Performance Measures for Combined Analysis

	Model														
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Predictor(s)
Planes Landed															
Trials 1 and 2	.19	.19	.34	.29	.25	.19	.34	.19	.27	.34	.27	.21	.34	.34	CR
	(.16)	(.16)	(.20)	(.18)	(.16)	(.16)	(.20)	(.16)	(.18)	(.20)	(.18)	(.12)	(.20)	(.20)	
Trial 3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None
Path Changes															
Trial 1	.16	–	.15	–	–	–	–	–	–	–	–	–	–	–	MOT RT B23
	(.10)		(.12)												
Trial 3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None
Alerts	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None
Duration															
Trial 1	.22	.32	.13	.32	.13	–	–	–	–	–	–	.32	–	–	MOT RT B23*
	(.16)	(.18)	(.10)	(.18)	(.10)							(.18)			Proj. CR B12
Trial 3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None

Note. R^2 values are provided in the table, with adjusted R^2 values in brackets. If there are multiple predictors, then the best predictor, based on higher β weight, is indicated with an asterisk. CR = Card Rotations Test, MOT RT = MOT Test Reaction Time, Proj. CR = Projection Test Correct Response

Analysis Using Group 1 Variables. Table 11 presents the models that significantly predicted various game performance measures along with the significant predictors. While there were several models that significantly predicted game performance, Model 5 predicted the number of planes landed for Trials 1 and 2 combined, number of path changes made in Trials 1 and 3, as well as the number of alerts provided (see Appendix AC for details). No models significantly predicted any of the other game performance measures. In total, seven additional measures were retained for use in Study 3: MRT, Surface Development Test, MOT Correct Response, Projection Correct Response, and Projection Lead Time.

Table 11

Effect Sizes of Significant Models (R^2, Adjusted R^2) and Significant Predictors of Game Performance Measures for Group 1 Analysis

	Model														
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Predictor(s)
Planes Landed															
Trials 1 and 2	–	–	–	–	.53 (.39)	–	–	–	–	–	–	.56 (.37)	–	–	MOT CR* MRT
Trial 3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None
Path Changes															
Trial 1	–	–	.27 (.21)	–	.27 (.21)	–	–	–	–	–	–	–	–	–	MOT RT B23* MOT CR
Trial 3	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.66 (.42)	.67 (.43)	MOT CR*, CR, MRT, Proj. LT 123 Proj. CR B3
Alerts	.51 (.39)	–	–	–	.55 (.39)	–	–	–	–	–	–	.55 (.39)	–	.63 (.37)	SD*, MOT CR
Duration															
Trial 1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None
Trial 3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None

Note. If there are multiple predictors, then the best predictor, based on higher β weight, is indicated with an asterisk. CR = Card Rotations Test, MRT = Mental Rotation Test, MOT CR = MOT Test Correct Response, MOT RT = MOT Test Reaction Time, Proj. LT = Projection Test Lead Time, Proj. CR = Projection Test Correct Response, SD = Surface Development Test

Analysis Using Group 2 Variables. Models that significantly predicted various game performance measures and the significant predictors are presented in Table 12. No single model was a significant predictor of all ATC measures; regression Model 3 and 5 predicted six of the seven ATC variables. These models predicted the number of planes landed for Trials 1 and 2 combined as well as Trial 3, the number of path changes made in Trials 1 and 3, number of alerts, and the game duration for Trial 3 (See Appendix AD for Model 3 details). Group 2 analysis revealed three additional variables to include in Study 3: Form Board Test, Map Planning Test, and Projection False Alarm.

Table 12

Effect Sizes of Significant Models (R^2, Adjusted R^2) and Significant Predictors of Game Performance Measures for Group 2 Analysis

	Model														
Measures	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Predictor(s)
Planes Landed															
Trials 1 and 2	.35 (.24)	.35 (.24)	.64 (.39)	.56 (.34)	.44 (.30)	.35 (.24)	.64 (.39)	.35 (.24)	–	.64 (.39)	–	–	.64 (.39)	.64 (.38)	Proj. FA*, FB, CR, MRT, MP, Proj. LT B3
Trial 3	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.82 (.69)	.77 (.69)	.82 (.69)	.80 (.67)	MRT*, FB, CR, Proj. FA, Proj. CR B12, MOT CR
Path Changes															
Trial 1	–	.64 (.38)	.64 (.38)	.64 (.38)	.23 (.18)	.64 (.38)	.64 (.38)	.64 (.38)	.64 (.38)	.64 (.38)	.64 (.38)	.64 (.38)	–	–	Proj LT B3,* Proj. LT B2, CR, MRT, MP, Proj. FA, MOT RT B23
Trial 3	–	.65 (.39)	.35 (.30)	.65 (.39)	.35 (.30)	.65 (.39)	.61 (.47)	.65 (.39)	.65 (.39)	.61 (.47)	.65 (.39)	–	–	–	MOT CR*, MOT RT B23
Alerts	.44 (.30)	–	.29 (.23)	–	.29 (.23)	–	–	–	–	–	–	–	–	–	MOT RT B23
Duration															
Trial 1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	None
Trial 3	.40 (.50)	.68 (.46)	.39 (.35)	.68 (.46)	.60 (.50)	.40 (.31)	.52 (.34)	.40 (.31)	–	.52 (.34)	–	.65 (.52)	–	.52 (.34)	MOT CR*, FB, MOT RT B23

Note. If there are multiple predictors, then the best predictor, based on higher β weight, is indicated with an asterisk. FB = Form Board, CR = Card Rotations, MP = Map Planning, MRT = Mental Rotation Test, MOT CR = MOT Correct Response, MOT RT = MOT Reaction Time, Proj. FA = Projection False Alarm, Proj. LT = Projection Lead Time, Proj. CR = Projection Test Correct Response

Subjective Workload Ratings

The ratings for the six NASA-TLX dimensions for the overlapping tests between the two groups were compared and revealed no significant difference in scores for MRT, Card Rotation, and MOT Tests ($p > .05$). Compared to Group 1, participants in Group 2 reported the Projection Test as significantly less mentally and temporally demanding, less frustrating, and requiring less effort (See Table 13). Overall, participants gave the highest mean NASA-TLX rating to the Surface Development ($M = 11.30$, $SD = 3.12$) and Choosing a Path ($M = 11.57$, $SD = 2.45$) tests, and lowest mean rating to the Maze Tracing test ($M = 8.16$, $SD = 3.17$), indicating they resulted in the most and least workload, respectively.

Table 13

Mean NASA-TLX Ratings for the Projection Test

	Group 1 ($n = 32$)		Group 2 ($n = 32$)				
Dimension	*M*	*SD*	*M*	*SD*	95% CI	*t*	*df*
Mental	12.22	4.74	11.28	4.53	[.35, 3.52]	2.49	31*
Physical	6.06	5.17	5.63	4.93	[-.87, 1.74]	.68	31
Temporal	11.47	5.61	9.16	4.75	[.28, 4.34]	2.32	31*
Performance[1]	7.75	4.10	8.25	3.53	[-2.23, 1.23]	-.59	31
Effort	13.56	4.18	11.84	4.10	[.38, 3.06]	2.62	31*
Frustration	9.69	4.68	7.56	4.26	[.52, 3.73]	2.70	31*

* $p < .05$

[1] This dimension was reverse scored; a lower number indicates better performance.

Spatial Planning Ability Measures

Both the Map Planning and Projection Tests were used as measures of spatial planning ability. If both tests measure this ability, then performance on both tests should be highly correlated. However, Pearson correlation analyses (1-tailed) revealed no significant, positive correlations; correct responses in block 3 of the Projection Test approached significance ($p = .06$). This finding was unexpected, however, these results do not necessarily indicate that these tests are not measuring aspects of spatial planning; it is possible that these tests measure different aspects of spatial planning ability or that one is more successful at measuring this ability than others. A factor analysis is a more appropriate method to determine whether these tests measure the same construct, however, this study does not have sufficient sample size to conduct such analysis.

Summary and Discussion

The primary goal of this study was to determine whether job-non-specific or job-specific tests are good predictors of job-related performance. According to the Domain Centered Framework, and the traditional use of spatial ability tests, scores from the non-specific tests should be good predictors of game performance. However, according to the Point-to-Point Theory, scores from the job-specific tests should be better predictors of game performance, provided that appropriate job-specific tests were used. The results revealed that five of the nine non-job-specific and all three job-specific tests were good predictors of various game performance measures. Given this, there was no clear and conclusive evidence to indicate that one approach to test development is superior to the other when tests are administered under job-non-specific context.

The inconclusive findings from this study regarding which approach to testing is better may be due to the context in which the spatial ability tests were administered. If an increase in job and test similarity is an important consideration for testing, then administering tests under more job-relevant context could reveal which type of test is better. This notion is examined in Study 3 by administering both types of tests under a control and two different job-relevant contexts (i.e., with increased mental workload and with interruptions). The two potential concerns with this study are discussed below.

Computerized Test Difficulty. It is possible that the two computerized job-specific tests may have been qualitatively different from the job-non-specific tests and the Map Planning job-specific test. The two computerized tests were developed so that they would be comparable to the other tests in terms of difficulty and effort required to complete the tests; this was to ensure that the findings could be attributed to a difference in similarity between job tasks and tests, rather than another factors such as workload. Since mean NASA-TLX ratings for the two computerized tests fall within the mean scores for the other tests, participants perceived the computerized tests to require no more and no less workload than the others, thus making them comparable to other tests.

ATC Clock Task Performance. With each succeeding trial, there were increased number of incorrect responses, and subsequently reduced accuracy rate, for the ATC clock task. The clock task was integrated with the ATC game to more realistically simulate the working conditions of ATCOs. The at and below chance level performance for trials 2 and 3 of the game may mean that participants paid less attention to the clock task and more attention to the game for those trials. Participants' ATC game performance indicates that they performed better on the third trial, lending some support to this possibility. Collectively, these findings suggest that contextualizing the ATC game to replicate similar working conditions of ATCOs was not successful on the third trial of the game. Interestingly, if participant's success on the ATC game was judged solely on the number of planes landed, then, for the third trial, the MRT was the only predictor for participants in Group 2 (no significant predictors for Group 1 or combined analyses). This suggests the possibility that a job-non-specific test, such as MRT, may be a good predictor of job performance if (1) the job tasks are performed out of job-context and (2) only the primary job goal (i.e., landing planes) was used as performance measure. Other factors that indirectly affected the number of planes landed (i.e., number of path changes made, how long individuals were able to play the game, and the number of alerts provided by the game) were primarily best predicted using job-specific tests (the number of alerts in Group 1 was the only measure predicted by a job-non-specific test). This is similar to Lievens and Patterson's (2011) finding that while tests with low and high similarity to tasks on the job predicted primary task performance, the high similarity test was also able to predict other abilities that contributed to primary task performance.

Chapter 7: Study 3 – Using Tests with Increased Contextual Similarity to Predict Job-Related Task Performance

The goal of this study was to administer job-specific and job-non-specific tests under job-relevant contexts and assess how that impacted how well tests predicted job-related task performance. Obtaining test performance under realistic working conditions may be important because, as the SMEs in Study 1 identified and as Study 2 alluded, it is possible that while job-non-specific tests measure individuals' ability to perform job tasks in isolation, those tests may not sufficiently predict job performance in realistic job context. An additional purpose of this study was to further investigate whether job-specific tests are better predictive of job performance than job-non-specific tests.

Although the SMEs in Study 1 acknowledged that multiple environmental factors could affect job-task performance, the ability to handle heavy mental workload and manage time efficiently was a common theme amongst all three focus groups. Given this, testing context in this study was manipulated in two distinct ways: by increasing mental workload of participants and by introducing interruptions during testing. A control condition, which was essentially a replication of Study 2 with reduced number of tests, was also employed. If testing under job-relevant context is important, then tests administered under higher mental workload and/or with interruptions should be better predictors of ATC-related game performance than tests in the control condition.

Method

Participants

One hundred and eighty Carleton University undergraduate students were screened and recruited according to procedure outlined in Study 2. Individuals who participated in Study 2 were not eligible to participate in this study. This study lasted approximately 2.5 hours and participants were compensated with 2.5 research participation credits. The data for four individuals (two from increased workload condition and two from interruption condition) were excluded from analyses because one participant did not complete the study and three participants exhibited behaviour that indicated they were not following instructions (e.g., not looking at the screen during computer tests, selecting response to paper and pencil tests without reading the questions). Given that exposure to a test affects performance on subsequent administration of the test, each participant was assigned to one of the three conditions: control, increased mental workload, or interruption. Fifty-eight participants (30 male, 28 female) were assigned to the Control condition and their ages ranged from 17 to 30 years (M = 19.50 years, SD = 2.20 years). Fifty-nine individuals (23 male, 36 female) were in the Workload condition and their ages ranged from 17 to 48 (M = 20.56 years, SD = 4.39 years). The remaining 59 participants (33 male, 26 female) were in the Interruption condition and their ages ranged from 17 to 53 years (M = 20.98 years, SD = 5.24 years). Initially all participants were randomly assigned to one of the three conditions and towards the end of the study, assignments were made to ensure roughly equal distribution of participants between the three conditions.

Stimuli and Apparatus

Job-Non-Specific Spatial Ability Tests

The following four job-non-specific spatial ability tests that predicted ATC-related task performance in Study 2 were used for this study: MRT, Card Rotations, Surface Development, and Form Board. All tests were scored according to instructions provided in Study 2. After each test, in addition to the NASA-TLX, a second Subjective Workload Rating (SWR) rating sheet that was developed for this research was administered to all participants (See Appendix AE). Participants performed the tests under one of the three conditions: control, increased mental workload, or with interruption. Participants in the control condition performed the tests according to the instructions provided in Study 2.

Participants in the increased workload condition were required to perform the test under time constraints, which is one of the ways in which workload has been manipulated in the literature (Gonzalez, 2005, Hendy, Liao, & Milgram, 1997; Hertzum & Holmegaard, 2013). Time pressure was induced in three ways: First, individuals were instructed to complete *all* the test questions within the allotted time, while ensuring that they were not guessing since a penalty would be assigned for wrong answers. Recall that in the control condition, individuals were asked to do their best, without an emphasis on the need to complete all the questions. The requirement to complete all questions in the allotted time should increase participants' workload since Study 2 revealed that participants only competed between 60.81% and 80.22% (M = 67.20%) of the questions for each of the tests that are used in this study. Second, a full circle timer was visually presented on the computer monitor that was located directly in front of the participants. The elapsed time filled the circle in red colour. In addition, every 30 seconds, participants were provided verbal reminders of the time remaining to complete each part of the test. Other researchers have found that simply presenting a countdown of the time remaining to perform a task was sufficient to induce time pressure (Hertzum & Holmegaard, 2013; Wang & Hu, 2011), which contributes to overall mental workload

(Hertzum & Holmegaard, 2013). Lastly, participants were presented with a soft, but detectable, metronome sound in the background. The sound was similar to that of a slightly fast ticking clock. This manipulation was added to increase participants' sense time passing by quickly. All three aspects of the workload manipulation were administered concurrently using a PowerPoint presentation. As a further attempt to increase individuals' sense of mental workload, when participants were ready to begin each part of the tests, they were asked to use the computer mouse to click at the center of a circle to start the timer, which activated the timer, verbal countdown, and metronome sound. The timers had the same durations (i.e., 3 mins, 6 mins, or 8 mins) and time expired indication (i.e., audio recording of "end of part 1" or "end of test") as Study 2.

The four attempts to increase the amount of mental workload experienced by participants should be successful, provided that the individuals were motivated to complete the tasks. A pilot study ($N = 5$) revealed that participants found the tests to be engaging and individuals reported that they attempted to do their best to get high scores. Therefore, it was assumed that participants in this study would also be intrinsically motivated to do their best. If the workload manipulation was successful, then ratings should be reflected in individuals' self-report ratings.

Participants in the interruption condition performed each test with multiple interruptions. Individuals always started with the primary test and were asked to switch between the primary test and the interruption task every time they heard "switch task" from the speakers. The interruption required individuals to visually remove themselves from the test and look at a sheet of paper that was located slightly to the right of them on the table. This sheet of paper contained simple addition questions with one or two carry over number (e.g., 46 + 18; See Appendix AF). Addition questions with carryovers have been used in surgery-related literature as a cognitive distractor task (e.g., Hsu, Man, Gizicki, Feldman, & Fried, 2008; Park et al., 2011). Individuals' ability to perform arithmetic questions have been found to require spatial ability, and consequently their spatial recourses (see Shanmugaratnam & Parush, 2012a), thereby reducing the likelihood of rehearsing information from the primary test during the interruption period. The audio instructions to switch task were presented via PowerPoint presentation.

The findings from the literature are inconsistent with regarding to the appropriate amount of time to interrupt individuals from their primary task. For example, Hodgetts and Jones (2006) have found that both 6 and 18-second interruption tasks significantly interrupted individuals' primary task performance. Furthermore, Monk, Trafton, and Boehm-Davis (2008) have demonstrated that interruptions lasting 3, 8, and 13 seconds equally affected task performance. A pilot study ($N = 5$) revealed that a 3 second interruption was too brief for participants to solve the math questions and resulted in frustration. The same pilot study revealed that both 5 and 9 second interruptions were sufficient for participants to attempt one or more math questions, thus successfully taking their attention away from the primary test. Therefore, both 5 and 9 second interruptions were employed at random points throughout each test, with each 3, 6, and 8 minute segments of the test having 3,4, and 4 interruptions, respectively, in order to have similar subjective density of interruption in all three time segments. Given that all job-non-specific spatial ability tests had two parts, one part contained 5-second interruptions while the other part contained 9-second interruptions. The order of interruption duration was counterbalanced across participants. The time lost due to performing the interruption task was compensated by providing the same amount of extra time to complete each part of the test. For example, if 3 minutes were provided for the first part of a test and interruptions lasted 15 seconds, then the total allotted time was 3 minutes and 15 seconds. The simplified version of the NASA-TLX and the SWR ratings were administered after each test. If the manipulation was successful, then the ratings for interruption condition should be higher than that of the control condition. The percent of correct answers was used as performance score for the arithmetic questions, with high scores indicating successful interruption.

Job-Specific Spatial Ability Tests

Participants performed the three job-specific tests under one of three conditions: control, increased mental workload, or with interruptions. Participants in the control condition were presented with the same instructions and tests as in Study 2, with the exception that only the 12 30-second trials were used for the MOT Test. The 15 second trials were not included for two reasons: First, introducing interruptions within a 15 second trial would limit when interruptions could occur, making it possible for participants to predict when they will occur. Second, increase in mental workload in the Workload condition would likely hinder performance, so it was ideal to use trials where participants performed better in Study 2 (i.e., 30 second trials) to see the extent to which increased workload affects performance.

The same workload manipulation used for the job-non-specific tests was used for the Map Planning Tests. For the MOT and Projection Tests, participants in the workload condition performed the tests with an increased number of distractor stimuli. Both tests had 6 additional randomly placed and randomly moving distracting stimuli in each of the

12 trials, for a total of 18 stimuli. Ho and colleagues (2004) have used similar number of stimuli (approximately16) to manipulate high workload. Eighteen stimuli were preferred in order to have 1:1 and 1:2 target:distractor ratios for the control/interruption and workload conditions, respectively. Increasing air traffic density has been previously used to increase mental workload (e.g., Galster, Duley, Masalonis, & Parasuraman, 2001; Hendy et al., 1997; Ho, Mikolic, Walters, & Sarter, 2004; Stein, 1985).

The interruption task for both the MOT and Projection Tests was the same arithmetic task administered for the non-job-specific tests. Interruptions were presented once per trial and occurred somewhere during the trial, with the exception that no interruptions occurred during the first and last 5 seconds of each trial. Six of the 12 trials had 5-second interruptions while the other six trials had 9-second interruptions. Interruption durations were randomized throughout the test. The measurements obtained for the MOT and Projection Tests were the same as those in Study 2. The workload and interruption manipulations were checked using the NASA-TLX and SWR ratings.

Spatial task. The ATC game, clock task, and the measurements taken from them were same as in Study 2. An independent rater analyzed the game performance and inter-rater reliability for the four game measures was assessed for the first trial of the game for the first 45 participants.

Procedure

This study employed the same procedure as Study 2, with the exceptions that all tests were administered under one of three conditions: control, increased mental workload, or interruptions. See Appendix AG for informed consent and debriefing.

Results

Data Cleaning and Assumptions. The same general procedure employed in Study 2 was used to clean and test assumptions, with the exception that the initial data cleaning procedure was more stringent and thorough than Study 2 (See Appendix AH).

Job-Non-Specific Spatial Ability Tests

There were no significant differences in performance between the three conditions for any of the job-non-specific tests; participants performed similarly on each tests regardless of the condition (See Table 14). The mean scores for the job-non-specific spatial ability tests in Study 3 were comparable to Study 2 findings, albeit slightly higher for all tests. The only significant difference in performance between participants in Study 2 and those in the Control condition in Study 3 was for the Surface Development test. Individuals in Study 3 performed significantly better (M = 37.83, SD = 14.41) than those in Study 2 (M = 29.30, SD = 16.62, $t(87)$ = -2.52, p < .05, CI [-15.26, -1.80], d = -0.55).

Table 14

Job-Non-Specific Spatial Ability Test Scores for the Three Conditions

	Condition									
	Control (N = 58)		Workload (n = 58)		Interruption (n = 58)					
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Spatial Visualization										
Form Board (240)	136.42	49.06	146.30	48.40	132.00	45.33	2	1.28	.02	.28
Surface Development (60)	37.83	14.41	33.34	15.06	34.88	15.86	2	1.20	.02	.30
Spatial Orientation										
Card Rotation (160)	106.71	34.04	109.54	32.02	103.81	30.89	2	.53	.01	.53
MRT (24)	11.12	4.74	10.01	5.21	9.19	4.90	2	2.16	.04	.12

Note. Maximum score for each test is presented in parentheses beside the test name.

Job-Specific Spatial Ability Tests

Map Planning Test

There were no significant differences in performance between the three conditions for the Map Planning Test ($p > .05$). On average, participants obtained a mean score of 23.72 ($SD = 6.64$), 26.28 ($SD = 6.75$) and 22.86 ($SD = 7.54$) for the Control, Workload, and Interruption conditions, respectively. Participants in the Control, Workload, and Interruption conditions completed 25.29 ($SD = 6.41$), 28.34 ($SD = 6.74$), and 24.97 ($SD = 6.39$) out of the 40 questions, respectively. Participants in Study 2 and those in the Control condition in this study did not differ significantly in their performance ($p > .05$).

MOT Test

When performances were compared across conditions, significant differences were observed in the average number of correct responses across all three conditions. The participants in the Control condition identified the most planes correctly, followed by those in the Interruption condition, and those in the Workload condition identified the least number of planes correctly. For Block 1, individuals in the Workload condition took significantly more time to make a response than those in the other two conditions. For Blocks 2 and 3 combined, individuals in the Workload condition took significantly more time to make a response than those in the Control condition (see Table 15).

Table 15

MOT Test Performance for the Three Conditions

	Condition									
	Control (*n* = 58)		Workload (*n* = 58)		Interruption (*n* = 58)					
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Correct Response	5.29	.57	4.49	.50	4.83	.82	1.78	19.66	.26	< .01
Response Time (seconds)										
Block 1	1.57	.40	1.81	.48	1.56	.43	2.00	5.43	.09	.01
Blocks 2 and 3	1.33	.38	1.53	.41	1.39	.38	2.00	3.71	.06	.03

One difference between Studies 2 and 3 is that only the 30-second durations for the MOT trials were used in Study 3. There was no evidence to indicate that having 30 second trials lead to participants predicting the end of each trial in such a way that it influenced performance; there was no significant difference for accuracy rates between Study 2, which had a mixture of 15 and 30 second trials, and Study 3 for any of the three blocks ($p > .05$). This also indicates that having a reduced number of trials in Study 3 (12 compared to 24 trials in Study 2) did not influence how individuals performed on the test.

Projection Test

While the mean number of correct responses per block was used as a measure of performance in Study 2, the unequal distribution of the number of plane crashes in each block (nine in Block 1, eight in Block 2, seven in Block 3) could have affected the results of practice effect analysis. This potential problem was remedied by using accuracy rate.

Performances were significantly different between Control and Workload conditions as well as between Workload and Interruption conditions for the following three measures: false alarm rate for Blocks 1 and 2 combined, lead time for Block 2, and accuracy rate for Block 2. Individuals in the Workload condition detected crashes later, had higher false alarm rate, and were least accurate, compared to those in the Control or Interruption conditions. Individuals in the Workload condition also had significantly higher false alarm rate for Block 3 than those in the Control condition. The accuracy rate for Block 1 was significantly higher for those in the Interruption condition, compared to those in the Workload condition. Block 3 accuracy rates were significantly higher for participants in the Control condition than those in the Interruption condition, who in turn were better than individuals the Workload condition (see Table 16).

Table 16

Projection Test Performance for the Three Conditions

	Condition									
	Control (n = 58)		Workload (n = 58)		Interruption (n = 58)					
Performance	M	SD	M	SD	M	SD	df	F	η	p
False Alarm										
Blocks 1 and 2	.29	.15	.43	.23	.25	.17	2.00	20.03	.26	< .01
Block 3	.37	.17	.46	.23	.43	.24	2.00	4.06	.07	.02
Lead Time (seconds)										
Blocks 1 and 3	2.65	1.16	2.47	1.22	2.77	1.77	1.73	.53	.01	.56
Block 2	3.95	1.32	3.08	1.26	4.43	1.68	2.00	11.95	.17	< .01
Accuracy										
Block 1	.67	.16	.62	.14	.68	.14	2.00	4.64	.08	.01
Block 2	.75	.19	.62	.17	.69	.18	2.00	3.19	.05	.04
Block 3	.68	.18	.49	.13	.57	.18	2.00	29.77	.34	< .01

Spatial Task

Despite clear verbal and written instructions, the same clicking-related difficulty experienced by some participants in Study 2 was also seen in this study, resulting in 38 such instances. In study 2, if a plane crash occurred due to such difficulty, then the entire trial was discounted. In hindsight, it is possible that an individual could have performed better on that trial than on the other trials. Given this, trials that were terminated due to clicking-related difficulty were retained if the number of planes landed on that trial was grater than one or both of the other trials.

There was no significant difference in performance between the three trials of ATC game for the four performance measures taken ($p > .05$). Given this, a single mean score for was obtained for each of the four measures. Performance between the three conditions was also not significantly different ($p > .05$). Participants' game duration lasted between 74 to 261 seconds, and during that time, they landed 6 to 46 planes, made 0 to 17 path changes, and received 1 to 22 alerts. Compared to Study 2, participants in this study landed significantly more planes and played the game for longer duration (see Table 17). The inter-rater reliability was assessed using SPSS and it ranged from .99 to 1.0 for the four performance measures.

Table 17

ATC Game Performance Comparison Between Studies 2 and 3

	Study			
	2[1] (*N* = 64)		3 (*N* = 176)	
Performance	*M*[1]	*SD*	*M*	*SD*
Planes Landed	17.01	5.96	19.34	6.13**
Duration (Seconds)	141.26	33.55	152.15	31.15*
Alerts	10.20	4.32	9.55	3.73
Path Changes	3.06	2.77	3.48	3.11

[1] In order to facilitate comparison, the values for Study 2 were averaged across blocks.
** $p < .01$ * $p < .05$

Clock Task

Significant differences in the number of correct and incorrect responses, as well as accuracy rates were observed between Trials 1 and 2 and between Trials 1 and 3 of the game (See Table 18). There were no significant differences between Trials 2 and 3 for these performance measures. Participants had more correct responses and less incorrect responses in Trial 1, resulting in better accuracy rates for the first trial of the game. Their performance then dropped to chance level and slightly below chance level for trials 2 and 3, respectively. There was no significant difference between trials for the number of questions missed ($p > .05$). This pattern of results is the same as what was observed in Study 2. See Appendix AI for a comparison of clock task accuracy for Studies 2 and 3.

Table 18

Clock Task Performance Scores for the Three Conditions

	Trials									
	1 (N = 64)		2 (N = 64)		3 (N = 64)					
Performance	M	SD	M	SD	M	SD	df	F	η	p
Correct	2.76	1.14	2.03	1.42	2.07	1.72	1.94	18.23	.09	< .01
Incorrect	1.25	1.13	1.94	1.37	2.11	1.33	1.94	30.91	.15	< .01
Miss	.13	.41	.21	.52	.20	.53	2.00	1.54	.01	.22
Accuracy (%)	71.87	23.02	50.84	30.10	46.05	33.25	2.00	51.22	.23	< .01

Predictors of ATC-Related Task Performance

The data were analyzed separately for each condition using the 14 regression models used in Study 2. There were no significant models with significant F_{Change} that predicted any of the ATC-related performances in either the Control or Interruption condition. Significant predictors were found only for the Workload condition and are presented below. A reliable model with 15 variables requires 139 participants (assuming $R^2 = .13$, power = .80, and $\alpha = .05$); however, the observed medium to large effect sizes indicates that the results are reliable despite the small sample size.

Number Of Planes Landed. While all 14 models were significant predictors of the number of planes landed, eight models accounted for the greatest variance, 33%, by three predictors, $R^2 = .50$, $R^2_{adjusted} = .33$, $F(15,43) = 2.86$, $p < .01$: Map Planning Test ($\beta = .45$, $p < .01$), Projection Lead Time for Block 2 ($\beta = .40$, $p < .01$) and Projection False Alarm Rate for Blocks 1 and 2 combined ($\beta = -.54$, $p < .05$).

Game Duration. All 14 models were significant predictors of game duration and five of these models indicated that 32% of the variance was explained by the same three predictors as for the number of planes landed, $R^2 = .50$, $R^2_{adjusted} = .32$, $F(15,43) = 2.86$, $p < .01$: Map Planning Test ($\beta = .49$, $p < .01$), Projection Lead Time for Block 2 ($\beta = .47$, $p < .01$) and Projection False Alarm Rate for Blocks 1 and 2 combined ($\beta = -.44$, $p < .05$).

Number of Alerts. Models 4 and 6 were significant predictors of the number of alerts provided by the game and Projection Lead Time for Blocks 1 and 3 combined ($\beta = .35$, $p < .05$) and Projection False Alarm Rate for Blocks 1 and 2 combined ($\beta = -.65$, $p < .01$) explained 16% of the variance, $R^2 = .26$, $R^2_{adjusted} = .16$, $F(7,51) = 2.60$ $p < .05$.

Number of Path Changes. There were no significant predictors for the number of path changes individuals made while playing the game.

Subjective Workload Ratings

Participants' NASA-TLX and SWR ratings did not differ significantly among the three groups for any of the job-non-specific tests. The following differences in ratings occurred for the job-specific tests among the three groups (see Table 19):

- For the Map Planning test, participants in the Workload condition believed they performed better than those in the Control condition.
- For the MOT Test:
    - Participants in the Interruption condition reported feeling the most rushed, followed by those in the Workload condition, who subsequently felt more rushed than those in the Control condition.
    - Participants in the Control condition believed they performed better than those in the Workload condition.
    - NASA-TLX ratings indicated participants in the Interruption condition reported exerting more effort than those in the Workload condition, who subsequently exerted more effort than those in the Control condition. SWR ratings revealed that the individuals in the Interruption condition felt they exerted more effort than those in the Control condition.
    - Participants in the Interruption condition reported experiencing more frustration than those in the Control condition.
- For the Projection Test, participants in the Interruption condition felt more rushed than those in the Control condition.

Table 19

Significant Differences in Subjective Workload Ratings Provided by Participants in the Three Conditions for the Spatial Ability Tests

	Condition									
	Control (n = 58)		Workload (n = 59)		Interruption (n = 59)					
Ratings	*M*	*SD*	*M*	*SD*	*M*	*SD*	*df*	*F*	η	*p*
Map Planning										
Performance	9.60	4.68	7.68	3.87	8.12	4.07	2, 173	3.34	.04	.04
MOT										
Rushed	8.79	5.60	10.23	5.08	13.05	5.07	2, 173	9.87	.10	< .01
Performance	7.62	3.72	10.20	3.33	8.90	3.28	2, 173	8.21	.09	< .01
Effort (NASA-TLX)	11.53	5.28	13.71	3.99	14.02	3.58	2, 173	5.69	.06	.01
Frustration	8.14	5.53	8.88	5.33	10.49	4.51	2, 173	3.21	.04	.04
Effort (SWR)	5.09	2.36	5.66	2.23	6.10	1.79	2, 172	3.28	.04	.04
Projection										
Rushed	10.05	5.45	10.19	5.62	12.42	4.54	2, 173	3.82	.04	.02

Note. Performance dimension was reverse scored; a lower number indicates better performance.

Spatial Planning Ability Measures

In contrast to Study 2, Pearson correlation analyses (1-tailed) for Study 3 provided some support to indicate that Map Planning and Projections tests assess spatial planning. For participants in the Control condition, both the Projection Test lead time for block 2 and false alarm rates for blocks 1 and 2 combined correlated significantly with Map Planning Test scores (r = .29 and -.23, respectively, $p < .05$). For participants in the Workload condition, individuals' Map Planning tests scores were significantly correlated with their false alarm rate for block 3 of the Projection Test (r = -.29, $p < .05$). For participants in the Interruption condition, the Projection Test lead times for blocks 1 and 3 combined, as well as for block 2 were significantly correlated with their Map Planning test scores (r = .23 and .33, respectively, $p < .05$).

Manipulation Check

The success of the manipulations was examined in three distinct ways: By using NASA-TLX and SWR ratings, the number of questions completed for job-non-specific spatial ability tests, and by using arithmetic task accuracy.

NASA-TLX and SWR Ratings

Higher mean scores in the Workload condition for each of the NASA-TLX and SWR ratings would indicate higher workload and difficulty compared to the Control condition, and lend support to successful workload manipulation. Ratings for the Interruption condition should also be higher than that of the Control condition since individuals in the Interruption condition should have experienced increased workload as a result of the interruptions. However, there was only one significant difference in workload ratings between the conditions: for the MOT Test, participants in the

Interruption condition reported higher NASA-TLX workload ratings (M = 8.79, SD = 3.53) than those in the Control condition (M = 8.79, SD = 3.53), $F(1.85, 105.33) = 6.49$, p = < .01. Yet, mean NASA-TLX and SWR ratings for each test indicate that individuals gave higher subjective workload scores for the Workload and Interruption conditions, compared to the Control condition, with the exception of NASA-TLX rating for the Form Board Test and the SWR rating for Surface Development test, where participants in the Control condition gave slightly higher rating that those in the Workload condition (the Interruption condition had the highest rating).

Number of Questions Completed

Given that individuals in the high Workload condition were asked to complete all the questions on the job-non-specific and Map Planning tests, while the individuals in the Control and Interruption conditions were told to do their best, individuals in the Workload condition should have completed more questions than those in the other two conditions. Furthermore, individuals in the Control and Interruption conditions should complete roughly the same amount of questions since they were given the same instructions and the same amount of time to complete each test.

The number of questions completed for each test were in the anticipated pattern, with more questions completed in the Workload condition and lower and comparable number of questions completed in the Interruption and Control condition (See Appendix AJ). However only the Map Planning Test and MRT showed significant difference in the number of questions completed between Workload and the other two conditions. Individuals in the Workload condition completed significantly more Map Planning test questions (M = 28.31, SD = 6.69) than those in the Interruption (M = 24.90, SD = 6.37) and Control, $M = 25.29$, $SD = 6.41$, $F(2, 114) = 4.38$, $p < .05$, conditions. Similarly, individuals in the Workload condition completed significantly more MRT questions ($M = 18.38$, $SD = 5.21$) than those in the Interruption ($M = 15.36$, $SD = 4.69$) or Control ($M = 15.48$, $SD = 4.81$, $F(2, 112) = 6.53$, $p = < .01$) conditions.

Arithmetic Task Accuracy

High accuracy rates for the arithmetic task administered during the interruptions would indicate successful interruption manipulation. Participants' mean arithmetic question accuracy ranged from 52.94% to 100%, with a mean of 89.96% (SD = 8.96%). The frequency of score distribution is as follows: 50% - 59% = 1, 60% - 69% = 1, 70% - 79% = 4, 80% - 89% = 17, 90% - 99% = 34, 100% = 1.

Factor Analysis

Factor analysis was conducted to examine whether job-specific and job-non-specific spatial ability tests would be identified as two separate factors, indicating that they are measuring different aspects of spatial ability. An additional interest was to observe whether spatial planning would emerge as a separate factor. This analysis was conducted using combined data from all three conditions in order to have a larger sample size. While the factor analysis revealed three different factors and the factors were grouped based on the type of tests (i.e., job-specific and job-non-specific), the observed Cronbach's Alpha values were lower than the traditional acceptance level of .70 (Nunnally, 1978); this suggests that these factors may not be reliable. This analysis did not identify spatial planning as a separate factor (see Appendix AK for more detail).

Summary and Discussion

The primary purpose of this study was to examine whether tests administered under conditions that individuals experience on their job would be better predictors of their performance on job-related tasks than tests administered under the standard testing condition. An additional goal of this study was to further investigate whether job-specific or job-non-specific tests are better predictors of job-related performance.

The results of this study provide a definite answer to the first inquiry: only the tests administered under the Workload condition predicted any of the game performances. Tests administered in the Control and Interruption conditions did not predict any of the game performances. This finding is important for two reasons:

1. It provides support that there is benefit to increasing test and job-task similarity and, more specifically, testing under similar job-context. However, simply testing under any job-specific context was not sufficient since tests administered under the Interruption condition were not good predictors of any performance measures. The Workload condition may have been successful in predicting performance because being able to manage high workload may influence ATC-related performance more than being able to handle interruptions. This study's results support the finding from the focus group that ATCOs need to be able to manage high workload.
2. The results also raise questions regarding the conventional way in which selection tests are administered. Often tests are administered in a situation similar to what participants experienced in the Control condition. Since neither the job-non-specific nor the job-specific tests administered under this condition predicted performance on the game, personnel selection testing may need to occur under relevant job-context.

With regards to the second inquiry, only two of the three job-specific tests (i.e., Map Planning and Projection) were good at predicting game performance. These were tests that assessed spatial planning ability. Once again, this indicates that not all attempts to increase test and job-task similarity will result in better tests; the MOT Test was not a good predictor of game performance despite MOT being an important aspect of what ATCOs do on the job. This finding could be due to using an unvalidated test of MOT. It could also be because, while it is important to track aircraft, it may be more important for ATCOs to plan ahead and project the trajectory of aircraft. It is also possible that spatial planning requires aspects of MOT; therefore while MOT Test alone may not be sufficient to predict performance, this ability may contribute to overall spatial planning. The findings from this study confirm observations from the focus groups and indicate that spatial planning ability of individuals may need to be assessed as part of ATCO selection.

In summary, the results of this study revealed that the testing context influences what the tests reveal regarding an individual's ability to perform job-related tasks. Additionally, not all tests of spatial ability assess the same type of ability since some tests were better predictors of game performance than others. Therefore we should exercise caution when choosing selection tests since different tests will reveal different results regarding individuals' ability to perform spatial tasks on the job. Potential concerns regarding the study are discussed below.

Manipulation Success. While the results of this study appear to be promising, it is important to address the findings that there were limited statistically significant results to indicate that the Workload and Interruption manipulations were successful since participants' subjective ratings were similar across all three conditions, especially for the job-non-specific tests. This finding may raise questions regarding the success of the two contextual manipulations. While the results were not statistically significant for all tests, the number of questions completed for the job-non-specific and Map Planning tests indicate that participants in the Workload condition attempted more questions than those in the Interruption or Control conditions. There is evidence to indicate that the workload manipulation was successful for the MRT and Map Planning Test. If individuals in the Workload condition attempted to complete all questions for these tests as instructed, then it is expected that they would have exerted similar efforts with other paper and pencil tests. Additionally, mean NASA-TLX and SWR ratings were in the anticipated direction, with individuals in the Workload and Interruption conditions indicating that they experienced more workload than those in the Control condition.

The workload manipulation for the MOT and Projection tests was inherent in the design of the tests, therefore it is expected that individuals in the Workload condition experienced more workload than those in the Control condition. Support for successful workload manipulation for the MOT Test comes from the finding that individuals in the Workload condition identified the least number of planes and took longer to respond compared to those the Control condition. For the Projection Test, the fact that individuals in the Workload condition detected plane crashes later, had higher false alarm rates, and lower accuracy rates compared to those in the Control and Interruption condition indicates that the workload manipulation was successful. A 90% mean accuracy rate for the interruption task (i.e., arithmetic questions) provides support that the interruption was successful, for both the job-specific and job-non-specific tests.

ATC Clock Task Performance. Participants' performance on the ATC clock task was the same pattern as that of Study 2, with participants having above, at, and below chance accuracy levels for Trials, 1, 2, and 3 of the game, respectively. It is unclear why participants' performance on the clock task deteriorated since there appears to be no trade-off between the clock task performance and ATC game performance. Some participants verbally reported that performing both tasks simultaneously was difficult for them. Therefore, they may have, either consciously or subconsciously, chosen to give more weight to the ATC task than the clock task despite being instructed to pay equal importance. If this is the likely explanation, then the finding that game performance did not benefit from this preferential attention indicates that the clock task still successfully demanded participants' attention away from game and made participants play the game in realistic job context.

Computerized Test Difficulty. The mean NASA-TLX and SWR ratings for the computerized tests fall within the mean scores for the existing, paper-and-pencils tests, which indicates that the two types of tests produced comparable workload. Therefore, observed performance differences between the two types of tests are not due differences in test difficulty.

Chapter 8: Study 4 – Using Tests With Increased Perceptual Similarity to Predict Job-Related Task Performance

The results from Study 3 indicated that an increase in both conceptual and contextual similarities between tests and job tasks are important considerations when developing and administering personnel selection tests. This study examined whether spatial planning tests also need to be perceptually similar to job tasks by administering two tests of spatial planning ability: one that is visually job-specific (i.e., Projection Test) and the other that is visually job-non-specific (i.e., Tower of London Test; TOL). If perceptual similarity is important, then the Projection Test should be the better predictor of ATC-related game performance. However, if perceptual similarity is not important, then either one or both tests should predict game performance.

An additional purpose of this study was to administer a validated test of spatial planning ability in order to confirm findings from Study 3. Ekstrom et al. (1976b) claimed that despite the name "Map Planning" and other researchers using this test to measure planning ability, this test was not intended measure planning ability. Given this, the present study utilized the TOL Test, which has been used as a measure of planning ability, especially in neuroscience research (e.g., Albert & Steinberg, 2011; Baker et al., 1996; Dagher, Owen, Boecker, & Brooks, 1999; Lazeron et al. 2000; Newman, Carpenter, Varma, & Just, 2000) and the Projection Test to examine whether tests of spatial planning are still good predictors of ATC-related performances. Both of these tests were administered to two different groups of participants, one in the increased Workload condition and the other under the Control condition, and the same four game performance measures used in Studies 2 and 3 were obtained.

Method

Participants

Seventy-seven Carleton University undergraduate students (39 males, 38 females) were screened and recruited using the same procedure as in Studies 2 and 3. Participants were randomly assigned to either the Control or Workload condition. Four participants' data (three from Workload condition and one from Control condition) were excluded from analysis: two participants reported they are colour-blind (which was an exclusion criterion), one participant detected plane crashes for the Projection Test only after they had occurred and resulted in no usable data being recorded for that test, and one participant's game data was lost due to technical difficulty. Of the remaining participants, 35 (18 males, 17 females) were assigned to the Control condition and their ages ranged from 17 to 45 years (M = 20.26 years, SD = 5.88 years). Thirty-eight participants (19 males, 19 females) were assigned to the Workload condition and their ages ranged from 17 to 40 years (M = 20.39 years, SD = 5.00 years). Individuals who participated in either Study 2 or 3 were not eligible to participate in this study. This study lasted approximately 1 hour and participants received 1% research participation credit.

Stimuli and Apparatus

Tower of London Test

The TOL Test is conceptually and perceptually similar to that of the Tower of Hanoi problem and requires individuals to use online strategic planning to complete each question (Phillips, Wynn, McPherson, & Gilhooly, 2001). The standard three pegs TOL Test (Shallice, 1982) was obtained from Millisecond Software. User license was not required since researchers are permitted to conduct their research during the 30-day trial period without licensing. While this test was originally developed to be more planning oriented and less spatial, (Shallice, 1982) the presentation and response has strong visual and spatial components (Phillips, Wynn, Gilhooly, Della Sala, & Logi, 1999). Furthermore, the task's reliance on spatial working memory has also been demonstrated in the literature (e.g., Morice & Dehaunty, 1996; Owen, Downes, Sahakian, Polkey, & Robbins, 1990; Welsh, Cicerello, Cuneo, & Brennan, 1995), thus making it a spatial test.

In this test, three differently coloured beads were arranged in a formation on three pegs. Participants were asked to rearrange the configuration in the middle of the screen to match the target arrangement presented at the top (See Figure 3). Participants were instructed to use a computer mouse to pick up and move one bead at a time. Only three beads could be placed on the left peg, two on the middle peg, and one on the right peg. Each question required a certain number of moves (i.e., 2, 3, 4, or 5) for successful completion and this number was displayed on the screen along with the number of completed moves. There were 12 problems in total and participants were provided three attempts to solve each problem. Participants were allowed to use the "reset" button if they wished to restart a trial, as long as it did not exceed 3 attempts. If participants were stuck on a question, they were permitted to use the "Next" button to skip to the next question. However, they were instructed to do their best and complete each question in one attempt and to try and not any skip questions. Successful solution to a question or pressing the "next" button revealed the next question. Given this, participants were not permitted to take a break once they start the test.

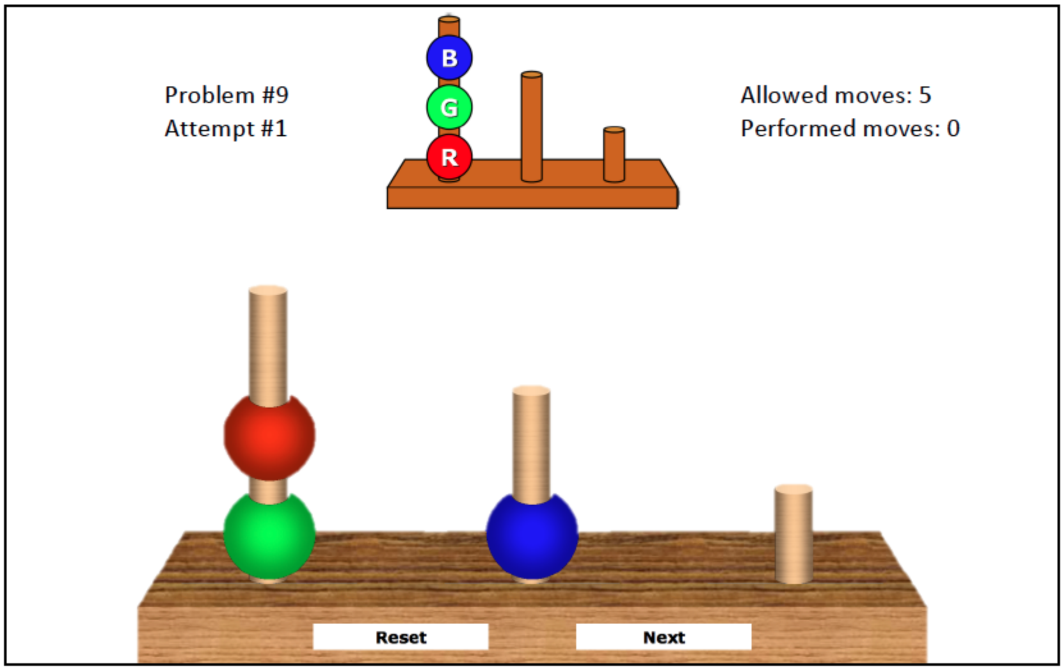


Figure 3. Tower of London Test.

The participants in the Workload condition completed the same test but they were permitted a maximum of 4000ms per move. Therefore, questions that required 2, 3, 4, and 5 moves had a maximum allowed solution time of 8,000 ms, 12,000 ms, 16,000 ms and 20,000 ms, respectively. This duration was deemed the best to induce time pressure after attempting the questions with various durations. The software recorded the time taken to solve each problem, the number of attempts made for each problem, and a test score for each participant.

Projection Test

The same Projection Test from Studies 2 and 3 was used for this study. This test was administered according to the procedure for the Control and Workload conditions in Study 3 and same measurements were taken.

Spatial Task

The same ATC game and clock task used in Studies 2 and 3 were used for this study and the same performance measures were taken.

Procedure

This study used the same general procedure as Study 3, with the exceptions that two job-specific spatial ability tests were administered under either the Control or Workload condition (See Appendix AL for informed consent and debriefing). Additionally, the demographic questionnaire included a question regarding participants' previous exposure to the Tower of Hanoi game or The Tower of London Test.

Results

Data Cleaning and Assumptions. Data cleaning and assumption checking were conducted according to the procedures outlined in Study 3 (See Appendix AM for detail).

Tower of London Test

Eight of the 73 participants reported having previous exposure to the Tower of Hanoi game or the Tower of London Test. Participants in the Workload condition took significantly less time (M = 2678.56 ms, SD = 427.42 ms) to make a move than those in the Control condition, M = 3822.68 ms, SD = 1110.72 ms; $t(43.17) = 5.72$, $p < .01$ (two-tailed), 95% CI [740.54, 1547.69], d = 1.34. Participants in the Workload condition scored significantly lower on the test (M = 27.24, SD = 3.37) than those in the Control condition, M = 31.31, SD = 3.11; $t(71) = 5.35$, $p < .01$ (two-tailed), 95% CI [2.56, 5.60], d = 1.25. There was no significant difference between the two conditions in the mean number of resets ($M_{\text{Control}} = 3.37$, $SD_{\text{Control}} = 2.51$, $M_{\text{Workload}} = 4.00$, $SD_{\text{Workload}} = 2.96$) and skipped questions ($M_{\text{Control}} = .20$, $SD_{\text{Control}} = .63$, $M_{\text{Workload}} = .21$, $SD_{\text{Workload}} = .62$; $p > .05$).

Projection Test

Overall, participants in the Workload condition had a significantly higher false alarm rate for Block 2, and had significantly smaller overall lead time and accuracy than those in the Control condition (see Table 20).

Table 20

Projection Test Performance for Participants in Control and Workload Conditions

	Condition							
	Control (n = 35)		Workload (n = 38)					
Performance	*M*	*SD*	*M*	*SD*	95% CI	*t*	*df*	*d*
False Alarm (%)								
Block 2	26.65	22.33	37.80	23.20	[-.22, -.01]	-2.09	71*	-0.49
Block 3	40.09	23.95	41.96	23.77	[-.13, .09]	-0.34	71	-0.08
Lead Time (seconds)								
Blocks 1 and 3	2.89	.99	2.45	.80	[.02, .86]	2.09	71*	0.49
Block 2	4.36	1.61	3.21	1.26	[.47, 1.82]	3.39	71**	0.79
Accuracy (%)	74.29	9.85	57.87	14.43	[10.69, 22.16]	5.72	66**	1.34

** p < .01, ** p < .05

Spatial Task

There was no significant difference in game performance between the three trials (p > .05). The game performance between individuals in the Control and Workload conditions was also not significant (p > .05). The game duration ranged from 90 seconds to 250 seconds, and in that time participants landed 7 to 41 planes, made 0 to 15 path changes, and received 3 to 24 alerts. The game performance, collapsed across the three trials, was not significantly different across Studies 2, 3, and 4 (See Appendix AN).

Clock Task

Significant differences in the number of correct and incorrect answers and accuracy rates were observed between Trials 1 and 2, as well as between Trials 1 and 3 (no significant difference between Trials 2 and 3; see Table 21). The pattern for the three trials of the game mimics the findings from Studies 2 and 3 and indicates that individuals' had more correct answers and less incorrect answers in Trial 1, compared to Trials 2 and 3, and consequently performed above, at, and below chance levels for Trials 1, 2, and 3, respectively. There was no significant difference in the number of questions missed across the three trials ($p > .05$). See Appendix AO for a comparison of clock task performance across Studies 2, 3, and 4.

Table 21

Clock Task Performance for the Three Trials of the Game

	Trials									
	1 (N = 64)		2 (N = 64)		3 (N = 64)					
Performance	M	SD	M	SD	M	SD	df	F	η	p
Correct	2.86	1.42	1.97	1.27	1.92	1.62	2	17.59	.20	< .01
Incorrect	1.18	1.02	1.92	1.20	2.23	1.37	2	19.20	.21	< .01
Miss	.18	.48	.25	.60	.26	.58	2	.49	.01	.62
Accuracy (%)	71.86	22.21	49.58	26.90	43.69	33.06	2	1.84	.31	< .01

Predictors of ATC-Related Task Performance

Six models were developed based on the models used in Studies 2 and 3, and were tested separately for Control and Workload conditions (see Appendix AP for models). New models were required given the reduction in the number of tests and the introduction of a new test. A reliable model with 7 variables requires 103 participants (assuming $R^2 = .13$, power = .80, and $\alpha = .05$). However, there were only 35 and 38 participants in the Control and Workload conditions, respectively. As with Studies 2 and 3, the observed medium to large effect sizes indicate that the results are reliable.

There were no significant models with significant F_{Change} that predicted any of the ATC performances in the Control condition. The results for the Workload condition are reported below. In general, there were no significant models with Projection Test variables as significant predictors. The mean time taken to make a move for the TOL Test was the only significant predictor of number of planes landed and game duration. TOL Test score was the only significant predictor of the number of path changes made.

Number of Planes Landed. While four of the six models were significant predictors of the number of planes landed, Model 6 accounted for the most variance at 20%, $R^2 = .32$, $R^2_{Adjusted} = .20$, $F(5,32) = 2.89$, $p <. 05$. TOL Test solution time was the only significant predictor of the number of planes landed ($\beta = -.39$, $p < .05$).

Game Duration. All six models were significant predictors of game duration, but Models 2 and 4 accounted for the most variance at 23%, $R^2 = .37$, $R^2_{Adjusted} = .23$, $F(7,30) = 2.54$, $p < .05$. The TOL Test solution time was the only significant predictor of game duration ($\beta = -.46$, $p < .01$).

Number of Alerts. There were no significant models with significant predictors for the number of alerts individuals received during the game.

Number of Path Changes. Both Models 2 and 6 were significant predictors of the number of path changes made, accounting for 19% and 11% of the variance, respectively. For both of these models, TOL Test score was the only significant predictor, Model 2: $R^2 = .23$, $R^2_{\text{Adjusted}} = .19$, $F(2,35) = 5.36$, $\beta = .42$, $p < .01$; Model 6: $R^2 = .17$, $R^2_{\text{Adjusted}} = .12$, $F(2,35) = 3.56$, $\beta = .41$, $p < .05$.

Subjective Workload Ratings

The only significant difference in subjective workload rating for the Projection Test was that participants in the Workload condition believed they performed better (M = 9.06, SD = 3.41) than those in the Control condition, $M = 11.94$, $SD = 3.70$; $t(70) = -3.44$, $p < .01$, 95% CI [0.84, -4.57], $d = -0.87$. For the TOL Test, participants in the Workload condition indicated that they felt significantly more rushed, that their test required more effort according to the NASA-TLX, and was difficult, stressful, frustrating, and complex compared to individuals in the Control condition (See Table 22).

Table 22

NASA-*TLX and SWR Ratings for the TOL Test*

	Condition						
	Control (n = 35)		Workload (n = 38)				
Rating dimensions	*M*	*SD*	*M*	*SD*	95% CI	*t*	*df*
NASA-TLX							
Mental Demand	1 0.40	4.60	12.34	4.63	[-4.10, 0.21]	-1.80	71
Physical Demand	3.09	3.49	3.47	4.94	[-2.40, 1.62]	-0.39	71
Temporal Demand	5.94	4.35	13.08	4.58	[-9.23, -5.05]	-6.81	71**
Performance[1]	5.91	4.15	7.05	4.01	[-3.04, 0.77]	-1.19	71
Effort	9.54	4.30	11.53	4.11	[-3.95, -0.02]	-2.01	71*
Frustration	4.89	4.19	6.47	4.91	[-3.71, 0.53]	-1.49	71
SWR							
Easy - Difficult	4.03	2.16	5.55	1.95	[-4.85, -0.99]	-3.16	71**
Relaxing - Stressful	3.83	1.52	5.39	1.73	[-2.33, -0.80]	-4.09	71**
Minimal Effort – Labour Intensive	3.86	1.50	4.66	2.13	[-1.66, 0.06]	-1.89	67
Pleasant - Annoying	3.63	2.06	4.47	1.52	[-1.70, 0.01]	-1.98	62
Calming - Frustrating	3.97	1.69	5.24	1.65	[-2.05, -0.49]	-3.24	71**
Simple - Complex	3.74	1.88	4.82	2.05	[-1.99, -0.15]	-2.32	71*

Spatial Planning Ability Measures

For individuals in the Control condition, the Projection Test lead time for blocks 1 and 3 combined were significantly correlated with TOL Test score ($r = -.31$, $p < .05$). While this indicates that these two tests may be assessing a similar construct, the negative correlation was unexpected; participants who were able to detect planes crashes in advance (i.e., have higher lead time) should have also scored higher on the TOL Test. This finding may be explained by looking at the relationship between lead time, accuracy, and false alarm rates for the Projection Test. The lack of significant, and positive, correlation between lead time and accuracy for the Projection Test rate ($p > .05$, 1-tailed) indicates that early identification did not translate to accurate identification of crashes; this is supported by the significant and positive correlations between lead time (blocks 1 and 3 combined, block 2) and false alarm rate (block 2, block 3) in both the Control and Workload conditions (correlations ranged from .35 to .73, $p < .05$, 1-tailed).

For individuals in the Workload condition, a significant and positive correlation was observed between Projection Test lead times for blocks 1 and 3 combined as well as block 2 and TOL score ($r = .28$ and .29, respectively, $p < .05$). This indicates that the Projection and TOL Tests may be measuring a similar construct.

Manipulation Check

The mean NASA-TLX and SWR ratings were used to obtain overall workload scores for both the Projection and TOL Tests. Independent samples *t*-tests revealed that, for the TOL Test, individuals in the Workload condition gave significantly higher mean NASA-TLX rating ($M = 8.99$, $SD = 3.$ 81) than those in the Control condition, $M = 6.63$, $SD = 2.81$), $t(71) = -3.35$, $p = <.01$, two-tailed. Similarly, mean SWR rating was higher for the Workload group (*M* = 5.02, *SD* = 1.55) than the Control group (*M* = 3.84, *SD* = 1.53), $t(71) = -3.26$, $p = <.01$, two-tailed. There were no significant differences in mean ratings were observed between the two conditions for the Projection Test ($p > .05$).

Summary and Discussion

Study 4 found that, of the two tests used to assess spatial planning ability, only the TOL Test administered under high workload predicted ATC-related game performance. The results from this study are important for the following four reasons:

1. They support the SMEs' statement in Study 1 regarding the importance of spatial planning ability for ATCOs.
2. They validate the results from Study 3 and indicate that spatial planning ability is important for performing ATC-related tasks.
3. They validate the findings from Study 3 regarding the importance of administering spatial ability tests under high workload context to predict ATC-related performance.
4. They suggest that tests need not be perceptually similar to the job tasks in order to predict job-related performance, provided they assess abilities that are conceptually essential to the job and are administered under job-relevant context.

The results of this study support the findings from Study 3 and further defines the extent to which selection test and job-task similarity should be established in order to best predict ATC-related performance. The results indicate that personnel selection tests should consider not only the content of the tests, but also the context in which tests are administered, which is currently not a common practice. The finding that tests need to assess job-specific abilities but those abilities need not be visually similar to tasks in the job is important because this may mean that a number of occupations with conceptually, but not necessarily perceptually, similar job tasks can utilize the same tests.

The results of this study, along with Study 3, indicate that spatial ability of individuals is influenced by the context in which the ability is applied, and therefore we must be cautious about inferring global spatial ability levels of individuals for personnel selection purposes; we may not be able to accurately predict how individuals will perform on spatial tasks without knowing the context in which the tasks will be performed. With regards to the ATC occupation, these results suggest that current ATCO selection could be improved by assessing spatial planning ability while individuals experience high mental workload. Future researchers could use this information to validate the findings from this study with ATCO candidates. The potential concerns with this study are discussed below.

Manipulation Success. The finding that, for the Projection Test, participants in the Workload condition believed they performed better than those in the Control condition could be explained by looking at the NASA-TLX ratings. While all the other five rating dimensions on the questionnaire ranged from “low/poor” on the left to “high/good” on the right, the ratings for the performance dimension was reversed and went from “perfect” on the left to “failure” on the right. While each participant was made aware of this difference, it is possible that they may have forgotten as they performed the ratings. If this is the likely explanation, then participants in the Control condition intending to indicate they performed better than those in the Workload condition makes sense. Yet, this pattern of results was not observed in Study 3. It is possible that any confusion participants had regarding the ratings had more of an impact in Study 4 than in Study 3 given the smaller sample size in Study 4 (58 for Control and 59 for Workload in Study 3 versus 35 for Control and 38 for Workload in Study 4). Another possibility is that participants believed they performed better simply due to the amount of effort they exerted in performing the task. Previous research has also indicated that individuals tend to be over-confident about their performance (e.g., Dunning, Heath, & Sus, 2004; Metcalf, 1998).

The workload manipulation for the Projection Test was inherent in the design. The finding that individuals in the Workload condition performed worse on the Projection Test compared to those in the Control condition indicates that increased workload may have affected their performance.

Unlike the subjective ratings for the Projection Test, participants did show differences in their subjective ratings for the TOL Test, indicating that the workload manipulation for this test was successful. Although the time restriction of 4000ms per move was chosen based on experimentation with various durations, the finding that participants in the control condition spent, on average, 3823 ms/move indicates that while participants in the Workload had sufficient time to make each move (i.e., 4000 ms/move), these individuals felt increased workload due to the time pressure placed on them.

ATC Clock Task Performance. Perhaps participants chose to focus more on the game than the clock task since there was no penalty for performing poorly on the task. However, having to respond to the task may have produced sufficient amount of interruption and increase in workload to not benefit game performance.

The key findings from Studies 1, 2, 3, and 4 are presented in Table 23.

Table 23

Summary of Key Findings From Studies 1, 2, 3, and 4

Study	Research question(s)	Research method	Key finding(s)	Implication(s)	
				Theoretical	Practical
1	What is the current problem with testing SA in ATCOs?	Use ATCOs as SMEs to understand current problem and needs.	Current SA tests are poor predictors of successful trainees.		Identify how to improve prediction.
			Current tests do not assess job-relevant SA.	Does test content affect prediction of task performance?	Assess job-relevant SA.
	How can current ATCO SA tests be improved?		Spatial planning is essential for ATCOs.		Assess spatial planning ability.
			Selection tests need to consider context in which job tasks are performed.	Does testing context affect prediction of task performance?	Administer SA tests under job-relevant context.
2	What should be the conceptual similarity between personnel selection tests and the job task for which they are selected?	Administer tests with high and low conceptual similarity to ATC-related job tasks.	Inconclusive evidence to indicate whether tests with high or low conceptual similarity to job tasks best predict performance.		Further examine the role of conceptual similarity for test development.

3	What should be the conceptual and contextual similarity between personnel selection tests and the job task for which they are selected?	Administer tests with high and low conceptual and contextual similarity to ATC-related job tasks.	Only spatial planning-related tests were predictive of job-task performance when administered under increased workload context. No tests predicted task performance when administered with interruptions or without consideration for job-task context.		Assess job-relevant aspect of an ability using tests with high conceptual similarity to job tasks. Administer tests under job-relevant context that impact task performance.
4	What should be the perceptual similarity between personnel selection tests and the job task for which they are selected?	Administer tests with high and low perceptual similarity to ATC-related job tasks.	Only the spatial planning test with increased conceptual similarity and administered under high workload predicted ATC-related task performance.		May not need to consider perceptual similarity of tests and job tasks for test development.
2 – 4	Is SA a construct that can be understood and assessed independently of the context in which it will be used?	Use SA tests to conduct research.	SA is a context dependent ability; the environment affects whether individuals are able to utilize this ability to their fullest.	Assessment and interpretation of SA should consider the context in which it will be used.	.
	How should SA be tested to be effective predictors of ATC-related performance?	Use ATC-related job tasks to conduct research.	Spatial planning ability tests administered under high workload predicted ATC-related task performance.		ATC-selection test should include test(s) of spatial planning ability under high workload.

Note. SA = spatial ability

Chapter 9: General Discussion

Collectively, the results of the four studies in this thesis offer insight into our improved understanding of (1) what should be the conceptual-, contextual-, and perceptual-similarity between personnel selection tests and the job task for which they are selected; (2) whether spatial ability is a construct that can be understood and assessed independently of the context in which it is used; and (3) how spatial abilities should be tested to be effective predictors of ATC-related performance. This concluding chapter of the thesis provides discussions on the findings and their significance, limitations of the studies, and recommendations for future studies.

Summary of This Research and Findings

This research was conducted to understand how personnel selection tests could be improved so that they can be used to identify qualified job candidates. The studies in this thesis used a variety of spatial ability tests and administered them under different contexts in order to assess how well the tests predicted individuals' ability to perform ATC-related tasks. The purpose of Study 1 was to better understand the ATC occupation and the conceptual, contextual, and perceptual similarities required between ATC job tasks and selection tests. The results revealed that (1) spatial ability is important for ATCOs; (2) ATCOs primarily deal with planning, executing, and monitoring aircraft activities; (3) the context in which ATCOs perform their task may need to be considered when testing; and (4) mental rotation ability, which is the aspect of spatial ability that is commonly used to assess spatial ability, may not be as important for ATCOs as other types of spatial ability.

Based on these findings, three subsequent laboratory studies were conducted to examine the extend to which personnel selection test need to be similar to job tasks in order to successfully identify suitable job candidates. In Study 2, participants were administered nine existing, job-non-specific, spatial ability tests, along with three tests that measured job-specific spatial ability. While the results narrowed down the number of tests, it did not provide conclusive evidence regarding whether job-non-specific or job-specific tests were better predictors of ATC-related task performance.

The purpose of Study 3 was to (1) examine whether selection tests need to be contextualized so that they are administered under conditions similar to what individuals experience on the job and (2) to replicate Study 2 with a reduced number of tests and a larger number of participants. This study utilized the tests that were significant predictors of ATC task performances in Study 2 and administered them under either a control, interruptive, or increased workload context. The results revealed that only the spatial planning tests administered under the high workload context were significant predictors of various ATC-related performance measures. The results of this study support Study 1's finding that spatial planning ability is important for ATCOs and that personnel selection tests should be administered under high workload context.

The goals of Study 4 were to (1) determine whether personnel selection tests need to be perceptually similar to job tasks, and (2) confirm the findings of Study 3 using a validated test of spatial planning ability, along with the Projection Test. The results revealed that spatial planning tests should be administered under high workload, but the tests do not need to be visually similar to job tasks (See Figure 4).

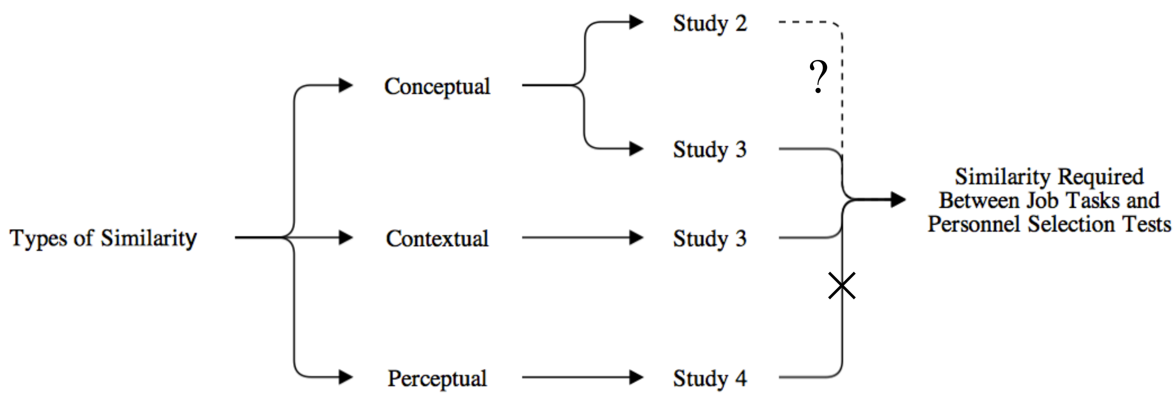


Figure 4. Similarity required between job tasks and personnel selection tests.

Contributions

Theoretical Aspects

Spatial Ability Is Context Dependent

Research Question 2: Taking spatial abilities as a test case, is it a construct that can be understood and assessed independently of the context in which it is used?

Many researchers (e.g., Kaufman, 2007, Terlecki et al., 2008) have discussed the contribution of innate ability and early childhood environmental influence in shaping individuals' spatial ability. Findings from the training literature indicate that spatial skills can also be improved with adequate training (e.g., Feng, Spence, & Pratt, 2007), thus contributing to improved overall spatial ability of individuals. Therefore spatial ability may be a construct that develops early in childhood and continues to be influenced by life experiences. Researchers also agree that performance on spatial tasks can be affected by various factors within the environment. For example, factors such as increased stress (Motowidlo et al., 1986; El Shikieri & Musa, 2012) and time constraints (Hertzum & Holmegaard, 2013; Wang & Hu, 2011) have been found to impact one's performance. With that said, most often spatial ability is viewed and discussed independently of context in which it is utilized. For example, a recruiter may identify that spatial ability is required for a particular job but not further investigate what type of spatial ability is required and/or under which conditions those abilities will be used.

The results of the studies in this thesis indicate that spatial abilities themselves may be contextualized such that one's ability to orient, navigate, and spatially plan, for example, depends on the context and situation in which the tasks are performed. Consequently, an individual performing the same task under two different contexts could have different outcomes despite having same ability level. For example, while an individual may have the necessary KSAOs to drive his car on a clear day, his driving performance will likely differ if the weather condition changes drastically, such as during heavy rain. This individual's performance is now dependent not only his driving ability but also his ability to compensate for the weather condition. Viewing spatial abilities from this perspective aids in better defining this ability. The following definition of spatial ability is proposed based on Linn and Peterson's (1985) as well as Lohman's (1993) explanation of the ability and incorporates findings from this research:

> *Spatial ability is a collection of spatially-relevant aptitudes that aid in performing tasks that require mental formulation, representation, transformation, retention, and/or retrieval of non-linguistic information in order to understand visual and/or spatial relation between objects in space. Furthermore, this ability is context dependent such that one's proficiency at accomplishing these mental tasks is dependent not only on his or her spatial abilities but also the ability to compensate for any constrains imposed by the environment in which the tasks are performed.*

Spatial Ability is Multifaceted

The findings from this research imply that spatial ability is a multifaceted ability. The finding that not all valid tests of spatial ability were good predictors of spatial task performance in Studies 2 and 3 indicate that these tests may be assessing different aspects of spatial ability; the tests that were good predictors assessed ATC-relevant spatial ability while other tests assessed spatial ability not relevant to ATC. In support of this view, factor analytic studies have indicated that multiple factors make up the spatial ability construct and those studies have used different tests to assess those factors (see Hegarty & Waller 2005; Mohler, 2006). Yet, some researchers continue to treat a single measure of spatial ability as a global index of individuals' ability to perform a variety of spatial tasks. The MRT (Vandenberg & Kuse, 1978) is one of the most commonly used tests to infer overall spatial ability levels of individuals; however this test only measures spatial orientation aspect of spatial ability. Therefore it is important to remember that the results of tests can only tell us about the aspect of spatial ability the test measures, rather than about overall spatial ability. It may be necessary to use a battery of spatial tests to obtain a more comprehensive understanding of an individual's overall spatial ability level. With that said, current findings also indicate that even when spatial abilities are measured and inferred correctly, test scores may not accurately reflect individuals' ability to perform tasks that require those abilities because the context in which the ability is used also affects task performance.

Assessing Other Abilities

Although the studies in this thesis utilized spatial ability as the domain in which the research questions were explored, the findings suggest that we should re-evaluate how we view and assess all abilities. Other concepts that we view as a single ability may in fact be comprised of multiple abilities and consequently may limit what we can infer from a single test of that ability. As an example, while we view intelligence as a measure of individuals' intellectual capacity, this concept is comprised of multiple types of intelligence (See Gardner, 1983). Even in cases where we have a good understanding of the composition of an ability, the findings from this research highlight the need to closely examine the impact of context in which the abilities are used. The influence of context is not something that is normally emphasized in assessments of ability. While we acknowledge that environment affects individuals' performance on tasks, this understanding is often not related back to our view of ability and what it means for understanding that ability.

Qualifying Similarity Needs Proposed by the Point-to-Point Theory

Although the Point-to-Point Theory advocates the use of tests with increased similarity to job tasks, it does not state the extend to which this similarity should be achieved. Furthermore, this theory does not specify how many aspects of the tests and tasks need to be similar. Therefore utilizing this theory for selection test development and administration poses difficulty to researchers due to lack to specificity. The findings from this research indicate not only that tests should assess job-specific aspects of an ability but also that these tests need to be administered under job-relevant contexts that have the potential to influence job-task performance. However, these tests need not be perceptually similar to job tasks (see Figure 4 on p. 140). These findings guide researchers in developing and administering improved selection tests.

Practical Aspects

Point-to-Point Similarity in Selection Tests Should Include Both Conceptual and Contextual Similarity

Research Question 1: What should be the conceptual, contextual, and perceptual similarity between personnel selection tests and the job task for which they are selected?

The results from this research indicate that selections tests that incorporate conceptually and contextually defining characteristics of a target job are better predictors of task performance. This finding could be vital in improving current personnel selection tests because current tests seldom assess job-specific aspect of a given ability under job-relevant context. Instead, even when personnel selection tests assess job-relevant ability, they are almost always administered under standard testing conditions. One of the difficulties with administering tests in such ideal condition is that job tasks are often not performed under ideal conditions. As such, the test results may not accurately reflect individuals' ability to perform job-related tasks under realistic job context during training. The main contribution from this research to personnel selection is the finding that when we assess ability levels of individuals, we should not detach it from the context in which the tests are administered. Consequently, it is necessary to increase similarity between testing and job context in order to have test scores that more accurately reflect individuals' ability to perform job-related tasks. Consider the following analogy to illustrate the importance of context. The current approach to ability testing is akin to evaluating an individual's qualification to be a race car driver by assessing her performance on a driving simulator or solo go-cart race. While these activities require driving ability, the driving tasks are performed under different environmental pressures.

It may be better to assess performance on something more similar to a go-cart race with multiple participants because these two activities share more similarities in terms of the abilities required to perform the driving task and the challenges imposed by environment.

The finding that none of the tests, whether job-specific or job-non-specific, administered using the standard testing condition were good predictors of performance raises questions regarding the current use of non-contextualized tests in personnel selection. Additionally, it challenges the use of Domain Centered Theory for personnel selection test development. Instead, the findings indicate that tests need to be developed and administered using the Point-to-Point Theory. The implication of this finding, on a broader scale, is that occupations that experience difficulties in selecting suitable job candidates may benefit from tailoring their personnel selection tests to a target job, rather than using a job-non-specific selection test developed based on the Domain Centered Framework.

ATCO Selection Tests Need to Assess Spatial Planning Under High Workload

Research Question 3: Taking ATC as an occupation where spatial abilities are critical, how should spatial abilities be tested to be effective predictors of ATC-related performance?

The results from this research indicate that individuals' ability to perform spatial planning tasks while experiencing high mental workload is a defining characteristic of ATCOs. Consequently, successful identification of qualified ATCO candidates may be improved by administering test(s) of spatial planning ability while candidates experience high mental workload. This finding is important because it indicates the type of spatial ability test to administer and how to administer it in order to improve ATCO selection.

Application of Results to Solve Real-World Problem

Recently, the Canadian Forces, within the Department of National Defence, have recognized that their trainees experience difficulties in performing a number of spatial tasks despite attaining passing scores on screening tests. They have considered the possibility that this may be due to having one or more inadequate screening test component(s) in the Canadian Forces Aptitude Test (CFAT). Research evidence supports the Canadian Forces' observation that the CFAT is not a good predictor of ATCO training success (Ebel-Lam, & Tanguay, 2011). The Canadian Forces believe that the weakness may be due to their spatial ability test, which has low face validity because it assesses an aspect of spatial ability that is not directly relevant to ATC; it assesses individuals' ability to mentally visualize the transformation of a cardboard pattern into a shape, and vice versa (See Appendix AQ). Furthermore, this test does not assess other job-relevant aspects of spatial ability.

Interviews with ATCOs within the Canadian Forces revealed that 20-40% of the individuals selected for training fail the program. Similar failure rate has also been found in civil aviation (Eurocontrol, 2014a). More alarmingly, failure rates between 35-60% have been observed in the Canadian Forces even after specialized training (Ebel-Lam, & Tanguay, 2011). A better selection test will aid in reducing training failure rate.

The results of this research indicate that a test of spatial planning ability administered under high workload could potentially improve the Canadian Forces' current screening test. Blanc (2003) has previously identified the importance of spatial planning ability for Canadian Forces' Maritime Surface and Sub-surface (MARS) Officers. MARS Officer selection utilizes the Maritime Officer Selection Test (MOST).

Although the intent of MOST was to assess individuals' memory, selection attention, and decision-making ability, Blanc concluded that this test assesses spatial scanning and general reasoning ability. MOST is conceptually and perceptually similar to the Map Planning Test (Ekstrom et al., 1976a) and inquires about the best route between two points on a grid route map (Blanc, 2003). Blanc found that spatial scanning ability, assessed using the same Map Planning Test used in the present research, along with general reasoning ability explained 23% of the variance in aptitude test performance. Furthermore, Blanc questions the reliability of MOST and suggests that the Map Planning Test, along with Mathematics Aptitude Test and Auditory Number Span Test from the Kit of Factor Referenced Cognitive Tests (Ekstrom et al., 1976a) could be used until a more reliable test for Maritime Officers is established. The present research indicates that using the TOL Test, rather than the Map Planning Test, and contextualizing the test may further improve the validity of the test for MARS Officer selection.

The findings from this research revealed that it is important to assess job-specific spatial ability and to administer tests of this ability under job-specific context. Organizations with ATC-relevant occupations could utilize this information to further investigate the role of spatial planning under high workload in order to improve their ATCO selection and reduce their training failure rate. These findings may also extend to other highly spatial and non-spatial occupations by instructing researchers how to develop and administer tests that are better predictive of job-task performance. The key to improving personnel selection tests may rely on assessing abilities that are essential to the job under context(s) that have the most potential to influence job-task performance.

Open Issues for Future Research

Validity and Reliability

Validity of the ATC Game. The validity of the ATC game for measuring spatial abilities required by ATCOs needs to be established. While this game has face validity and the content validity is accounted for since the game requires abilities that the ATCOs in the focus groups identified, other criterion-related validities (i.e., predictive, concurrent, and convergent) need to be established.

Convergent Validity of Spatial Planning Tests. Convergent validity of the Projection, Map Planning, and TOL Tests used to measure spatial planning ability should be examined. Despite the visual and spatial presentation and responses of these tests, there was limited evidence to indicate that these test scores are highly correlated, thus indicating that these tests may not be measuring a similar construct. The small sample size in this research did not allow for a proper factor analysis, which would have been an ideal method to use. Despite the limited evidence to indicate whether these tests were measuring spatial planning ability, the use of a validated test of planning ability in Study 4 allows us to be more confident in the final results regarding the importance of spatial planning ability for the ATC occupation.

External and Ecological Validities. The external and ecological validities of the findings should be established, especially if these tests are to be used with the target population. Future researchers may wish administer the TOL Test to ATCO candidates under realistic work conditions to determine whether the findings from this research can be replicated. If similar findings are observed, then the results generalize to real-world personnel selection context.

Validity and Reliability of Subjective Ratings. Participants' NASA-TLX and SWR ratings did not differ between the three conditions, despite there being evidence to indicate that the manipulations were successful. Validity and reliability can be potential problems for self-report measures (Barker, Pistrang, & Elliott, 2005). While the NASA-TLX is a valid measure of workload (Hart & Staveland, 1988), the validity of the SWR rating sheet has not been established. The reliability of individuals' self-report is also unclear since while both NASA-TLX and SWR had an overlapping effort-related question, the correlations on these two measures were low to medium, depending on the test (r = .21 to .68). This suggests that individuals may not be self-aware of their performance. A non-self-report based assessment of workload may be more appropriate.

Discrepancy in Findings Between Studies

Studies 3 and 4 revealed that tests administered under the standard testing condition (i.e., Control condition) did not predict game performances. However, tests were administered only under this condition in Study 2 and the results revealed several predictors. One possible explanation is that participants in Study 2 may have incidentally performed the tests under higher mental workload than participants in the Control condition in Study 3 since Study 2 had more tests and took longer to complete than Study 3 (approximately 3 and 2.5 hours, respectively). The slight increase in mental workload in Study 2 may have resulted in some tests appearing as predictors of game performance. These predictors did not emerge in Study 3 for the Control condition because of the reduced workload compared to Study 2. Most of these tests did not showed up as predictors under the Workload condition in Study 3 due to much higher levels of workload experienced in Study 3.

Participant Motivation

Although the use of undergraduate students from a university may be viewed as a potential limitation, in actuality, these students represent the type of individuals who apply to become ATCOs. ATCO applicants' minimal requirement is a high school diploma (Air Services Australia, 2014; Eurocontrol, 2014b; NATS, 2014; NavCanada, 2014) and they are not provided with any special training prior to taking the screening tests. However, one thing that may differ between ATCO applicants and undergraduate students is the motivation to do well on the tests; individuals applying to be ATCOs may be more motivated to perform well on the tests and therefore may be better prepared to take the test. Such preparation may influence how individuals perform on the test and, consequently, how well the test scores predict ATC-relevant task performance. The results of this thesis need to be validated with real ATCO candidates.

Moderating Variable

The findings in this research alluded to job context as a moderating variable that may influence the strength of the relationship between selection tests and prediction of job-task performance. However, moderation analysis was not performed due to limited sample size. Future studies with larger sample sizes could investigate this possibility.

Application of Findings

Other Occupations Similar to ATC. The findings from this research may also be applicable to other occupations similar to ATC (e.g., pilots, boat and ship captains). The role of spatial planning under high workload for predicting job-task performance in different occupations can be assessed using the TOL Test under high and low workload with simplified job tasks from target occupations.

Other Domains. Other occupations that have a critical spatial ability component could benefit from better understanding the specific spatial and contextual needs of the occupation and by examining whether they should be considered during selection. In order to accomplish this, the procedures used for the studies in this thesis could be replicated with individuals in the target occupations. Other non-spatial domains can use a similar approach to investigate which tests of abilities in a particular domain are best for personnel selection use and under what condition(s) they should be administered.

Other Occupations' Contextual Needs. While performing job tasks under high workload was found to be important for ATCOs, other occupations may have different requirements imposed by the environment in which the job tasks are performed. Therefore, while most personnel selection tests may benefit from having tests with increased conceptual similarity, the benefit of increased contextual similarity during test administration would depend on the jobs' contextual requirements and the correct identification of those requirements. Future researchers may wish to identify other environmental factors that impact job-task performance and examine whether it is beneficial to incorporate those factors during selection test administration.

Conclusion

Selection tests are commonly developed using the Domain Centered Framework and usually only resemble the target job tasks at a broad, conceptual level. This thesis examined whether tests with increased similarity to job tasks, based on the Point-to-Point Theory's recommendation, would result in test scores that are better predictive of individuals' ability to successfully complete job-related tasks during training. Test and job-task similarity was examined at the conceptual, contextual, and perceptual levels.

The results broadened our understanding of the spatial ability construct and the development and administration of personnel selection tests. The results of this research reveal the benefit of using personnel selection tests with increased conceptual and contextual similarity to job tasks and prompt further research in order to improve how well personnel selection tests identify qualified job candidates.

References

Air Services Australia (2014). *Air traffic control in training program*. Retrieved from http://www.airservicesaustralia.com/careers/air-traffic-controller/air-traffic-control-in-training-program/

Albert & Steinberg (2011). Age differences in strategic planning as indexed by the Tower of London. *Child Development*, *82*, 1501-1517.

Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology*, *25*, 251-269.

Asher, J. J., & Sciarrino, J. A. (1974). Realistic work sample tests: A review. *Personnel Psychology*, *27*, 519-533.

Austin, J. T., & Villanova, P. (1992). The criterion Problem: 1917-1992. *Journal of Applied Psychology*, *77*, 836-874.

Bailey, M. (1999, November). *Evolution of aptitude testing in the RAF*. Paper presented at the RTO HFM Workshop on "Officer Selection." Monterey, USA.

Baker, S. C., Rogers, R .D., Owen, A. M., Frith, C. D., Dolan, R. J., Frackowiak, R. S. J., & Robbins, T. W. (1996). Neural systems engaged by planning: A PET study of the Tower of London task. *Neuropsychologica*, *34*, 515-526.

Barker, C., Pistrang, N., & Elliott, R. (2005). Self-report Methods. In *Research Methods in Clinical Psychology: An Introduction for Students and Practitioners, Second Edition*. (pp. 94-118). Chichester, UK: John Wiley & Sons, Ltd.

Barbarino, M., Clark, C. P., & Phillip, W. (1998). *Integrated task and job analysis of Air Traffic Controllers. Development of Methods*. European Organisation for the Safety of Air Navigation.

Becton, J. B., Carr, J. C., & Judge, T. A (2011). Is the past prologue for some more than others? The hobo syndrome and job complexity. *Journal of Vocational Behavior*, *79*, 448-460. doi:10.1016/j.jvb.2011.04.001

Becton, J. B., Matthews, M. C., Hartley, D. L., & Whitaker, D. H. (2009). Biodata as a predictor of turnover, organizational commitment, and job performance in the healthcare industry. *International Journal of Selection and Assessment*, *17*, 189–202.

Bernardin, H. J. (1987). Development and validation of a forced choice scale to measure job-related discomfort among customer service representatives. *Academy of Management Journal*, *30*, 162–173.

Bennett, G., Seashore, H., & Wesman, A. (1972). Manual of Differential Aptitude Test. New York: The Psychological Corporation.

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*, 387-409. doi:10.1348/096317905X26994

Bettencourt, K. C., & Somers, D. C. (2009). Effects of target enhancement and distractor suppression on multiple object tracking capacity. *Journal of Vision*, *9*, 1-11. doi: 10.1167/9.7.9

Blanc, S. J.-R. (2003). *Naval officer selection in Canada: An evaluation of the Maritime Officer Selection Test (MOST)*. (Master's thesis, Saint Mary's University, Halifax, Canada). Retrieved from http://library2.smu.ca/bitstream/handle/01/22430/blanc_sebastien_j-r_masters_2003.PDF?sequence=1

Borich, G. D., & Bauman, P. M. (1972, September). Convergent and discriminant validation of the French and Guilford-Zimmerman spatial orientation and spatial visualization factors. In *Proceedings of the Annual Convention of the American Psychological Association (Vol. 7, Pt. 1),* Honolulu, Hawaii (pp. 9-10).

Bureau of Labor Statistics (2014). *Occupational outlook handbook, 2010-2011 edition*. Retrieved from http://www.bls.gov/ooh/transportation-and-material-moving/air-traffic-controllers.htm

Canadian Air Division Headquarters (2004). *Qualification Standard. Aerospace Control*. Ottawa: Author.

Canadian Forces (2013a). *Controller course – Instrument Flight Rules*. Retrieved from http://www.forces.gc.ca/en/training-establishments/international-training-programs-courses/controller-course-instrument-flight-rules.page

Canadian Forces (2013b). *Controller course – Air Weapons*. Retrieved from http://www.forces.gc.ca/en/training-establishments/international-training-programs-courses/controller-course-air-weapons.page

Canadian Forces (2013c). *Controller course – Visual Flight Rules*. Retrieved from http://www.forces.gc.ca/en/training-establishments/international-training-programs-courses/controller-course-visual-flight-rules.page

Carpenter, P. A., & Just, M. A. (1986). Spatial ability: An information processing approach to psychometrics. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3). Hillsdale, NJ: Earlbaum.

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

Cascio, W. F., & Aguinis, H. (2005). *Applied Psychology in Human Resource Management* (6th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

Cattell R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement*, *15*, 139-164.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54,* 1-22.

Cherney, I. D., & London, K. (2006). Gender-linked differences in toys, television shows, computer games, and outdoor activities of 5- to 13- year-old children. *Sex Roles, 54*, 717- 726. doi: 10.1007/s11199-006-9037-8

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63,* 83–117. doi: 10.1111/j.1744-6570.2009.01163.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Colom, R., Contreras, M. J., Botella, J., & Santacreu, J. (2002). Vehicles of spatial ability. *Personality and Individual Differences, 32*, 903-912.

Colom, R., Contreras, M. J., Shih, P. C., & Santacreu, J. (2003). The assessment of spatial ability through a single computerized test. *European Journal of Psychological Assessment*, *19*, 92-100. doi: 10.1027//1015-5759.19.2.92

Contreras, M. J., Colom, R., Hernández, J. M., & Santacreu (2003). Is static spatial performance distinguishable from dynamic spatial performance? A latent-variable analysis. *Journal of General Psychology*, *130*, 277-88.

Cox-Fuenzalida, L. E. (2007). Effect of workload history on task performance. *Human Factors*, *49*, 277-291. doi: 10.1518/001872007X312496

D'Oliveria, T. C. (2004). Dynamic spatial ability: An exploratory analysis and a confirmatory study. *The International Journal of Aviation Psychology*, *14*, 19-38. doi: 10.1207/s15327108ijap1401_2

Department of National Defence. (2007). *Canadian Forces Aptitude Test practice version.* Retrieved from http://cdn.forces.ca/_PDF2010/preparing_for_aptitude_test_en.pdf

Dagher, A., Owen, A., Boecker, H., & Brooks, D. (1999). Mapping the network for planning: A correlational PET activation study with the Tower of London. *Brain*, *122*, 1973–1987.

Darken, R. P., & Peterson, B. (2002). Spatial orientation, wayfinding, and representation. In K. M. Stanney (Ed.), *Handbook of virtual environments: Design, implementation, and applications* (pp. 493-518). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Dean, M. (2004). An assessment of biodata predictive ability across multiple performance criteria. *Applied H. R. M. Research*, *9*, 1-12.

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

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest,* *5*(3), 69-106.

Ebel-Lam, A., & Tanguay, D. (2011, October). *Assessing the predictive validity of selection tests among Canadian Forces Aerospace Control Officer candidates*. Presented at the International Military Testing Association Conference, South Kuta, Indonesia.

El Shikieri A. B., & Musa, H. A. (2012). Factors associated with occupational stress and their effects on organizational performance in a Sudanese university *Creative Education*, *3*, 134-144.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976a). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976b). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Eurocontrol (2014a). *Frequently asked questions: What happens if I fail training?* Retrieved from http://www.eurocontrol.int/faq/air-traffic-controller-jobs

Eurocontrol (2014b). *Apply for an air traffic controller job*. Retrieved from http://www.eurocontrol.int/articles/apply-air-traffic-controller-job

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18,* 850-855.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.

Formisano, E., Linden, D. E. J., Di Salle, F., Trojano, L., Esposito, F., Sack, A. T., Grossi, D., Zanella, F. E, & Goebel, R. (2002). Tracking the mind's image in the brain. Time-resolved fMRI during visuospatial mental imagery. *Neuron, 35*, 185-194. doi:10.1016/S0896-6273(02)00747-X

Furr, R. M., & Funder, D. C. (2004). Situational similarity and behavioural consistency: Subjective, objective, variable-centered, and person-centered approaches. *Journal of Research in Personality*, *38*, 421-447. doi:10.1016/j.jrp.2003.10.001

Galster, S. M., Duley, J. A., Masalonis, A. J., & Parasuraman, R. (2001). Air Traffic Controller performance and workload under mature free flight: Conflict detection and resolution of aircraft self-separation. *The International Journal of Aviation Psychology*, 11*, 71-93.*

Gardner, H. (1983). Frames of Mind: Theory of Multiple Intelligences. New York: Basic Book Inc.

Goldberg, J., & Meredith, W. (1975). A longitudinal study of spatial ability. *Behavior Genetics, 5*, 127-135.

Goldstein, I. L., & Ford, J. K. (2002). *Training in Organizations* (4th ed.). Belmont, CA: Wadsworth Group.

Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors*, *47*, 92-101. doi: 10.1518/0018720053653767

González, V. M., & Mark, G. (2004). "*Constant, constant, multi-tasking craziness": Managing multiple working spheres*. Paper presented at the proceedings of the SIGCHI conference on Human factors in computing systems, Austria: CHI.

Gordon, M. E., & Kleiman, L. S. (1976). The prediction of trainability using a work sample test and an aptitude test: A direct comparison. *Personnel Psychology*, *29*, 243-253.

Gregory, R. J. (2011). *Psychological testing: History, principles, and applications* (6th ed.). Boston, MA: Pearson Education Inc.

Girden, E. R. (1992). *ANOVA: repeated measures*. Newbury Park, CA: Sage.

Grossi, D., Modafferi, A., Pelosi, L., & Trojano, L. (1989). On the different roles of the cerebral hemisphere in mental imagery: The "o'Clock Test" in two clinical cases. *Brain and Cognition*, *10*, 18-27. doi:10.1016/0278-2626(89)90072-9

Halpern, D. F. (2001). *Sex differences in cognitive abilities.* New York: Lawrence Erlbaum Associates.

Han, T. H., Chien, P. C., Chen, H. C., Chao, S. M., & Wu, J. T. (2001, March). *Developing an effective measurement of visual–spatial ability in pilot selection battery from a view point of cognitive approach.* Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Hartfield, K. N., & Conture, E. G. (2006). Effects of perceptual and conceptual similarity in lexical priming of young children who stutter: Preliminary findings. *Journal of Fluency Disorder*, *31*, 303-324.

Hegarty, M., & Waller, D. (2004). A Dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*, 175-191. doi:10.1016/j.intell.2003.12.001

Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah, & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121-169). Cambridge University Press.

Hegarty, M., Kozhevnikov, M., & Waller, D. (2008). *Perspective Taking/Spatial Orientation Test*. Santa Barbara, CA.

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*, 151-176. doi:10.1016/j.intell.2005.09.005

Hendy, K. C., Liao, J., & Milgram, P. (1997). Combining time and intensity effects in assessing operator information-processing load. *Human Factors*, *39,* 30-47.

Hertzum, M., & Holmegaard, K. D. (2013). Perceived time as a measure of mental workload: Effects of time constraints and task success. *International Journal of Human-Computer Interaction, 29,* 26-39.

Hintzman, D. L., O'Dell, C. S., & Arndt, D. R. (1981). Orientation in cognitive maps. *Cognitive Psychology*, *13*, 149–206.

Ho, C.-Y., Mikolic, M. I., Walters, M. J., & Sarter, N. (2004). Not now! Supporting interruption management by indicating the modality of urgency of pending tasks. *Human Factors*, *46*, 399-409.

Hodgetts, H. M., & Jones, D. M. (2006). Interruption of the tower of London task: Support for a goal-activation approach. *Journal of Experimental Psychology: General*, *135*, 103-115. doi: 10.1037/0096-3445.135.1.103

Hooven, C. K., Chabris, C. F., Ellison, P. T., & Kosslyn. M. (2004). The relationship of male testosterone to components of mental rotation. *Neuropsychologia*, *42*, 782-790. doi: 10.1016/j.neuropsychologia.2003.11.012

Hsu, K. E., Man, F.-Y., Gizicki, R. A., Feldman, L. S. & Fried, G. M. (2008). Experienced surgeons can do more than one thing at a time: effect of distraction on performance of a simple laparoscopic and cognitive task by experienced and novice surgeons. *Surgical Endoscopy*, *22*, 196-201. doi: 10.1007/s00464-007-9452-0

Human Resources and Skills Development Canada (2013). *Air Traffic Control and related occupation*s. Retrieved from http://www30.rhdcc.gc.ca/CNP/English/NOC/2006/Profile.aspx?val=0&val1=227

Hunt, E., & Waller, D. (1999). *Orientation and wayfinding: A review* (ONR technical report N00014-96-0380). Arlington, VA: Office of Naval Research.

Hunt. E. B. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York, NY: Russell Sage Foundation.

Hunter, A. C. (2011). *How Do We Track what We Don't See?: Spatiotemporal Information in Multiple Object Tracking* (Doctoral dissertation). Retrieved from Carleton University Research Virtual Environment (CURVE).

Johannsdottir, K. R., & Herdman, C. M. (2010). The role of working memory in supporting drivers' situation awareness for surrounding traffic. *Human Factors*, *52*, 663-673. doi:10.1177/0018720810385427

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*, 137-172.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence*, *35*, 211-223. doi:10.1016/j.intell.2006.07.009

Klirs, E. G., & Revelle, W. (1986). Predicting variability from received situational similarity. *Journal of Research in Personality*, *20*, 32-50.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature, 378*, 496-498.

Kozhevnikov, M., Hegarty, M., & Mayer. (2002). Revising the visualizer/verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction, 20*, 47-77.

Larson, G. E. (1996). Mental rotation of static and dynamic figures. *Perception and Psychophysics*, *58*, 153-159.

Lazeron, R. H. C., Rombouts, S. A. R. B., Machielsen, W. C. M., Scheltens, P., Witter, M. P., Uylings, H. B. M., & Barkhof, F. (2000). Visualizing brain activation during planning: The Tower of London Test adapted for functional MR imaging. *American Journal of Neuroradiology*, *21*, 1407-1414.

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology*, *96*, 927-940. doi: 10.1037/a0023496

Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: a comparison in terms of predictive validity. *Journal of Applied Psychology*, *91*, 1181-1188. doi: 10.1037/0021-9010.91.5.1181

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.

Lohman, D. F. (1986). The effect of speed-accuracy trade off in sex differences in mental rotation. *Perception and Psychophysics, 39*, 427-436.

Lohman, D. F. (1988). Spatial abilities as traits, processes and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, (pp. 181-248). Hillsdale NJ: Erlbaum.

Lohman, D. F. (1993). Spatial ability and G. *Paper presented at the first Spearman seminar,* Plymouth, UK.

Lord, C. G. (1982). Predicting behavioural consistency from an individual's perception of situational similarities. *Journal of Personality and Social Psychology*, *42*, 1076-1088.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, *86*, 889-918.

Melsom, A. M. (2009). Individual differences in spatial abilities: Evidence for different strategies in an indoor navigation task (Unpublished thesis). *University of Tromsø*. Tromsø: Norway.

Metcalfe, J. (1998). Cognitive Optimism: Self-Deception or Memory-Based Processing Heuristics? *Personality and Social Psychology Review, 2*(2), 100-110.

Micheal, W. B., Guilford, J. P., Fruchter, B. & Zimmerman, W. S. (1957). The description of spatial-visualization abilities. *Educational and Psychological Measurement, 17*, 185-199. doi: 10.1177/001316445701700202

Mohler, J. L. (2006). *Examining the spatial ability phenomenon from the student's perspective*. Unpublished doctoral dissertation, Purdue University. Indiana, United States of America.

Money, J., Alexander, D., & Walker, H.T. (1965). *Manual: A standardized road-map test of direction sense*. Baltimore, MD: Johns Hopkins Press.

Monk, C. A., Boehm-Davis, D. A., & Trafton, J. G. (2004). Recovering from interruptions: Implications for driver distraction research. *Human Factors*, *46*, 650-663. doi: 10.1518/hfes.46.4.650.56816

Monk, C. A., Trafton, J. G., & Boehm-Davis, D. (2008). The effect of interruption duration and demand on resuming suspended goals. *Journal of Experimental Psychology*, *14*, 299-313. doi: 10.1037/a0014402

Morice, R., & Delahunty, A. (1996). Frontal/executive impairments in schizophrenia. *Schizophrenia Bulletin*, *22*, 125-137.

Motowidlo, S. J., Packard, J. S., & Manning, M. R. (1986). Occupational stress: Its causes and consequences for job performance. *Journal of Applied Psychology*, *71*, 618-629.

Mumwa, R. J., Pellegrino, J. W., Kali, R. V., & Carter, P. (1984). Different slopes for different folks: Process analysis of spatial aptitude. *Memory and Cognition, 12*, 515-521.

NATS (2014). *Eligibility*. Retrieved October 11, 2014 from http://www.nats.aero/careers/trainee-air-traffic-Controllers/eligibility/.

NavCanada (2014) *Selection requirements*. Retrieved from http://takecharge.navcanada.ca/en/content/training/requirements

Newman, S. D., Carpenter, P. A., Varma, S., & Just, M. A. (2003). Frontal and parietal participation in problem solving in the Tower of London: f MRI and computational modeling of planning and high-level perception. *Neuropsychologica*, *41*, 1668, 1682.

Nori, R., Grandicelli, S., & Giusberti, F. (2006). Visuo-spatial ability and wayfinding performance in real-world. *Cognitive Processing, 7*, S135-S137. doi: 10.1007/s10339-006-0104-4

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*, 673-690.

O*NET (2010a). *Details report for Air Traffic Controllers*. Retrieved from http://www.onetonline.org/link/details/53-2021.00

O*NET (2010b). *Abilities – Spatial orientation*. Retrieved from http://www.onetonline.org/find/descriptor/result/1.A.1.f.1?a=1

Owen, A. M., Downes, J. J., Sahakian, B. J., Polkey, C. E., & Robbins, T. W. (1990). Planning and spatial working memory following frontal lobe lesions in man. *Neuropsychologia*, *28*, 1021-1034.

Park, J., Waqar, S., Kersey, T., Modi, N., Ong, C., & Sleep, T. (2011). Effects of distraction on simulated anterior segment surgical performance. *Journal of Cataract Refractive Surgery*, *37, 1517-1522*. doi:10.1016/j.jcrs.2011.01.031

Paivio, A. (1978). Comparisons of mental clocks. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 61-71.

Parush A., & Berman, D. (2004). Orientation and navigation in 3D user interfaces: The impact of navigation aids and landmarks. *International Journal of Human Computer Studies*, *61*, 375-395. doi:10.1016/j.ijhcs.2003.12.018

Pellegrino, J. W., & Kail, R. (1982). Process analyses of spatial aptitude. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 311-366). Hillsdale, NJ: Lawrence Erlbaum Associates.

Peters, M. (1995). *Revised Vandenberg & Kuse Mental Rotations Tests: Forms MRT-A to MRT-D*. Guelph (ON), Canada: Department of Psychology, University of Guelph.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: different versions and factors that affect performance. *Brain and Cognition*, *28*, 39–58.

Phillips, L. H., Wynn, V., Gilhooly, K. J., Della Sala, S., & Logi, R. H. (1999). The role of memory in the Tower of London Task. *Memory*, *7*, 209-231.

Phillips, L. H., Wynn, V. E., McPherson, S. & Gilhooly, K. J. (2001). Mental planning and the Tower of London task. *Quarterly Journal of Experimental Psychology A, 54,* 579-598. doi:10.1080/02724980042000237

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179-197.

Pylyshyn, Z. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*, 801-822. doi: 10.1080/13506280344000518

Robertson, I. T., & Kandola, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. *Journal of Occupational Psychology*, *55*, 171–183.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology*, *88*, 174-188.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*, 511-514.

Scholl, B. J., Pylyshyn, Z. W., & Feldman, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition*, *80*, 159-177.

Schmidt, F. L. & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology*, *62*, 529-540.

Seamster, T. L., Redding, R. E., & Kaempf, G. L. (1997*). Applied Cognitive Task Analysis in Aviation*. Aldershot: Avebury Ashgate Publishing Ltd.

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *298*, 199-209.

Shanmugaratnam, S., & Parush, A. (2012a). Spatial Ability: Its role in the Canadian Forces and proposal for a comprehensive test battery. (Technical Report).

Shanmugaratnam, S, & Parush, A. (2012b). The spatial and cognitive tasks of the Aerospace Control Officer. (Technical Report).

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Sherman, R. A., Nave, C. S., & Funder, D. C. (2010). Situation similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, *99*, 330-343. doi: 10.1037/a0019796

Sjölinder, M. (1998). *Spatial cognition and environmental descriptions*. Retrieved from http://www.sics.se/humle/projects/persona/web/littsurvey/ch4.pdf

Spearman, C. (1904). The abilities of man. The *American Journal of Psychology, 15*, 201-292.

Spool, J. M. (2004). *The KJ-Technique: A group process for establishing priorities*. Retrieved from http://www.uie.com/articles/kj_technique/

Stankov, L. (1978). Fluid and crystallized intelligence and broad perceptual factors among 11 to 12 year olds. *Journal of Educational Psychology, 70*, 324-334.

Stein, E. (1985). *Air Traffic Controller workload: An examination of workload probe* (DOT/FAA/CT-TN84/24). Atlantic City: US Department of Transportation.

Tabachnick, B. G. & Fidell, L. S. (2006). Using multivariate statistics. Boston, MA: Pearson Education, Inc.

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, *22*, 996-1013. doi: 10.1002/acp.1420

Thomas, J. C., & Hersen, M. (2003). Comprehensive handbook of psychological assessment, volume 4: Industrial and organizational assessment. New Jersey: John Wiley & Sons Inc.

Thomson, G.H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, *8*, 271-281.

Thompson, W.L., Slotnick, S.D., Burrage, M.S., & Kossyln, S.M. (2009). Two forms of spatial imagery. *Psychological Science, 20*, 1245-1253.

Thorndike E. L. (1903). *Educational Psychology*. New York: Lemcke & Buechner.

Thorndyke, P. W., & Hayes-Roth, B. (1982). Differences in spatial knowledge acquired from maps and navigation. *Cognitive Psychology*, *14*, 560-589.

Thornton, I. M., Bülthoff, H. H., Horowitx, T. S., Rynnin, A. & Lee, S.-W. (2014). Interactive Multiple Object Tracing (iMOT). *PLOS ONE*, *9*, 1-19. doi:10.1371/ journal.pone.0086974

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotation, a group test of three-dimensional spatial visualization. *Perceptual & Motor Skills, 47*, 599-604.

Vinchur, A.J., Schippmann, J.S., Switzer, F.S. & Roth, P.L. (1998). A meta-analytic review of predictors of job performance for salespeople. *Journal of Applied Psychology*, *83*, 586-597.

Waller, D. A. (1999). A*n assessment of individual differences in spatial knowledge of real and virtual environments* (unpublished doctoral dissertation). University of Washington, USA.

Wang, D., & Hu, Y. (2011). The effect of time constraint on information processing decision making: A computer stimulate. *Journal of Convergence Information Technology, 6*, 311-317. doi:10.4156/jcit.vol6.issue3.36

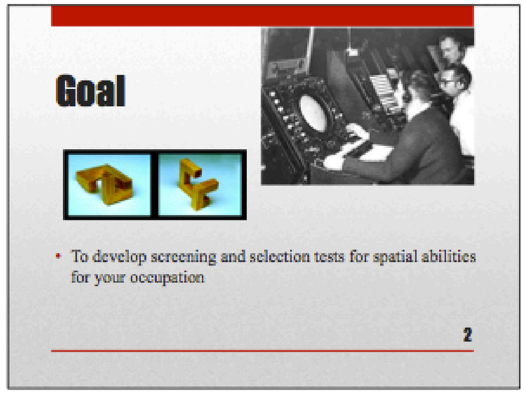
Welsh, M., Cicerello, A., Cuneo, R., & Brennan, M. (1995). Error and temporal patterns in Tower of Hanoi performance: Cognitive mechanisms and individual differences. *Journal of General Psychology*, *122*, 69-81.

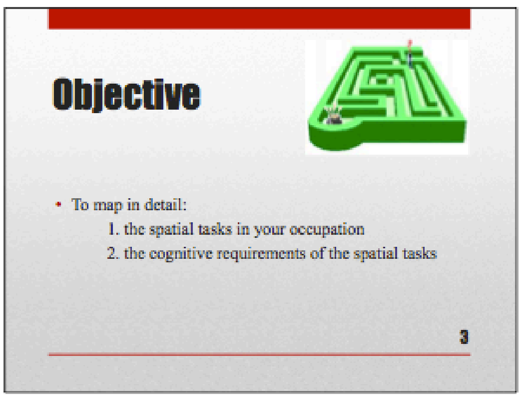
Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *The Journal of Applied Psychology*, *52*, 372–376.

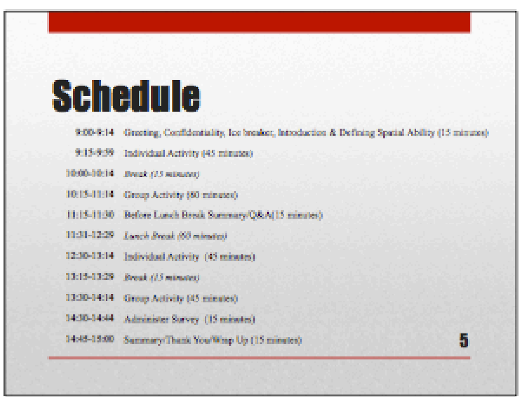
Wyatt, M., Pathak, S. B., & Zibarras, L. D. (2010). Advancing selection in an SME: Is best practice methodology applicable*? International Small Business Journal*, *28*, 258–273. doi: 10.1177/0266242609350815

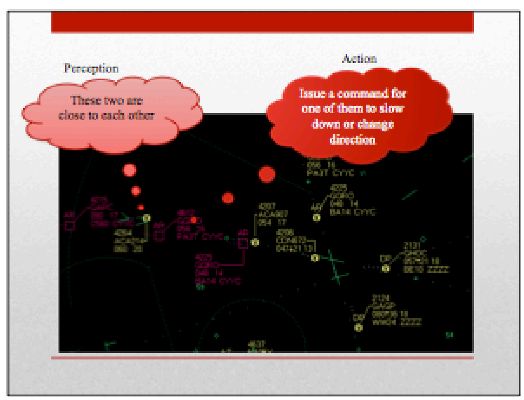
Appendix A

Study 1: Presentation Slides

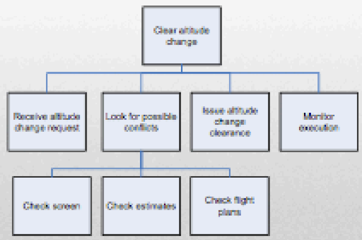








Task Structure Example



7

Spatial Tasks

- Perception:
  - Getting spatial information about elements in the space you are controlling
- Action:
  - Performing actions that can impact the spatial configuration of elements in the environment you are controlling

8

Individual Activity

- Write down all the spatial tasks in your occupation
- One spatial task per post-it note
- Spatial Tasks: any tasks that require you to get information and perform actions related to the space and the relations between elements (e.g., of aircraft) in the space you are controlling/monitoring

Examples of spatial tasks:
- Maintain vertical separation
- Clear altitude change
- Clear approach
- Determine sequencing

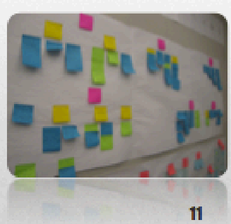
9

Break

10

Group Activity

- Sort the notes into different categories.
- Use new post-it notes to write down category/sub-category names and to help you with the categorization task.



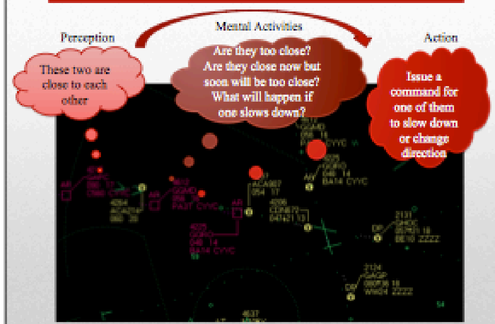
11

Summary

- Group presentation of diagram developed

12

13



Spatial Cognition

- The simultaneous, image-like or graphic or verbal mental representation ("mental map") of the entire space you are controlling/ monitoring, including the layout of all elements and the spatial relationships among them.
- Spatial cognitive tasks:
  - Anything you do mentally in order to get spatial information, assess it, reason and judge, retain it, and make decisions related to it

15

Constructing / Maintaining Spatial Cognition

- Anything you do mentally in order to construct, update and maintain the "mental map" in a way that will enable you to make inferences and decisions and carry them out.
- Examples for spatial cognitive tasks:
  - Extrapolate
  - Assess distance or direction
  - Compare
  - Retrieve from memory
  - Pay attention to…

16

Individual Activity

- For each spatial task in your diagram, write down all cognitive tasks/processes required to perform them
- One cognitive task/process per post-it note
- Cognitive processes = the thought process/what kind of thinking occurs in you head when you perform the tasks.

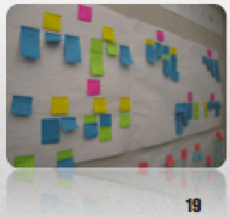
17

Break

18

Group Activity

- Sort the notes into different categories.
- Use new post-it notes to write down category/sub-category names and to help you with the categorization task.



19

Summary

- Group presentation of diagram developed

20

Spatial Ability

It is "*the **ability** to generate, retain, retrieve, and transform well-structured visual images*"

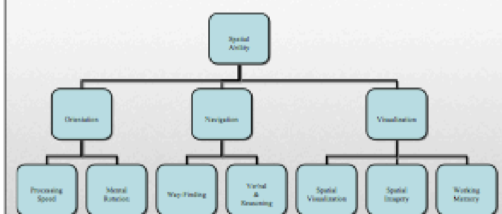
Lohman, 1993, p. 13

In addition to requiring mental visualization, we propose that spatial ability also includes "*individuals' **ability** to prepare for performing spatial tasks (e.g., through orienting and attending to information) and carrying out the spatial task (e.g., navigation)*"

Shanmugaratnam & Parush, 2010, p. 24

21

Spatial Ability



22

Survey

23

Thank You!

24

Appendix B

Study 1: Informed Consent

This informed consent provides information on the purpose of this study and the nature of your involvement. This informed consent must provide sufficient information for you to make an informed decision as to whether or not you would like to participate in this study.

Study Title: Spatial Ability: Cognitive Task Analyses for Aerospace Control Officers

Study Personnel:

Sharmili Shan, Tel. ***-****; EMAIL: sshanmu5@connect.carleton.ca
Dr. Avi Parush, Tel. 520.2600 ext. 6026; EMAIL: avi_parush@carleton.ca
Dr. Cheryl Burgess, Tel. 996.3392; EMAIL: Cheryl.Burgess@forces.gc.ca
John Johnson; EMAIL: John.Johnston5@forces.gc.ca

If you have any ethical concerns about this study, please contact either Dr. Monique Sénéchal (Chair, Carleton University Ethics Committee for Psychological Research, 613-520-2600, ext. 1155) or Dr. Janet Mantler (Chair, Department of Psychology, Carleton University, 613-520-2600, ext. 4173).

Purpose and Task Requirements: A comprehensive literature review conducted by the researchers revealed that spatial ability can be divided into three main sub-categories and these can further be divided into seven subcomponents. The purpose of the focus group is to understand the cognitive, and more specifically the spatial components, in each of the three occupations by discussing the on-the-job tasks with experienced individuals. Additionally, the objective is to obtain sufficient information during the focus groups to identify which of the seven spatial ability subcomponents correspond to each of the three occupations (i.e., Do these three occupations require all seven subcomponent abilities or just some?). In order to determine this, you will be first asked to discuss the various tasks performed in your occupation and later asked to expand on this by discussing the spatial ability relevance for each task. At the end of the day, you will be provided with a brief survey on the relevance of the seven identified spatial subcomponents to your occupation. This focus group will last approximately eight hours, but we will have a 30 minute lunch break and two 15 minute breaks (one in the morning, one in the afternoon). Should the need arise, we can also take additional brief breaks during the day. All aspects of the focus group will take place in this room.

Potential Risk and Discomfort: This focus group does not pose any risk or discomfort. However, if you have any concerns, please address it to the researchers.

Anonymity/Confidentiality: Although the focus group sessions will be audio-recorded, the information collected during the session will not be shared with others (i.e., anyone other than the researchers) and will be aggregated so that will not be traceable to a particular individual. This information will be stored in a locked cabinet and be used only by the researchers. Additionally, the audio recordings will be deleted at the end of the project.

Right to Withdraw: Your participation in this focus group is voluntary and you have the right to withdraw from this focus group at any time without penalty.

I have read the above description and consent to participate in this focus group by providing the following information. My signature indicates my consent to participate in the study and does not constitute a waiver of my rights.

Full Name (please print): ____
Participant Signature: ____
Date: ____
Researcher Signature: ____
Date: ____

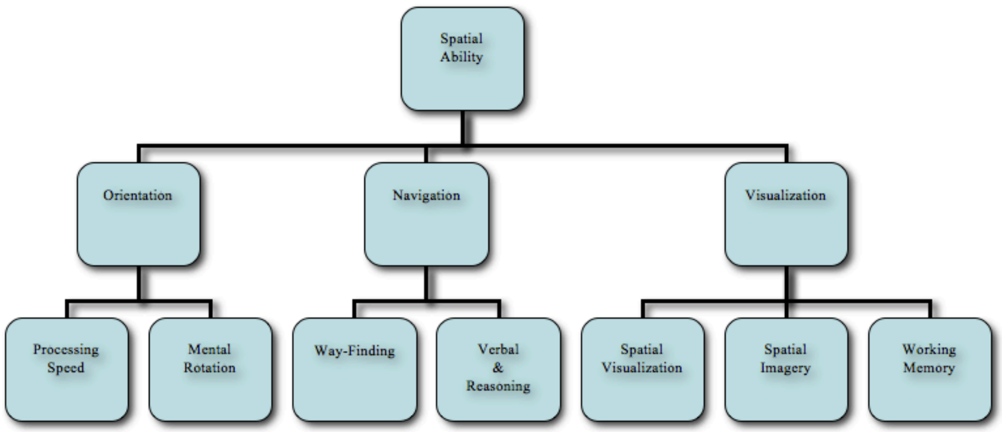
This study has received clearance by the Carleton University Psychology Research Ethics Board (11-061).

Appendix C

Study 1: Questionnaire

Spatial Ability Relevance Survey

In our literature review we identified three components and seven subcomponents of spatial ability (see below). These categorizations were made based on review of all spatial ability materials (i.e., not specifically related to your occupation), therefore, on the next pages, we would like you to rate the relevance of each them to your occupation.



Please rate the importance each of the following at your job on a 1 (this dimension is irrelevant to my job) to10 (this dimension is vital for performing tasks in my job) scale and the importance at selection on a 1 (not at all) to 5(trouble definitely likely if ignored at selection) scale. The rating dimensions are indicated in **bold** and the definitions for each of the dimensions are provided in *italics*.

1. **Spatial Ability**: *One's ability to create, maintain, recall, and mentally manipulate visual images. Also includes individuals' ability to prepare for performing spatial tasks (e.g., through orienting and attending to information) and carrying out the spatial task (e.g., navigation).*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if spatial ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

2. **Orientation**: *Determining where one is located with respect to nearby objects and the target location.*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if orientation ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

a. **Processing Speed**: *Ability to quickly process information in order to make a decision and/or carryout a task.*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if processing speed ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

b. **Mental Rotation**: *Ability to mentally visualize and manipulate things (e.g., maps, images) in 3D.*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if mental rotation ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

3. **Navigation**: *The use of physical motor movements to get from one point to another and involves both locomotion and way-finding components.*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if navigation ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

a. **Way-Finding**: *Ability to mentally navigate around areas.*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if way-finding ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

b. **Verbal & Reasoning:** *Ability to use cognitive processing to understand situations and to make decisions.*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if verbal and reasoning ability is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

c. **Working Memory**: *Ability to temporarily keep visual and spatial information in the forefront of memory so that they can be used in the immediate future (e.g., within 5 minutes).*

1	2	3	4	5	6	7	8	9	10
Irrelevant				Moderately Important					Vital

To what extent are negative consequences likely if working memory capacity is ignored at selection?

1	2	3	4	5
Not at all	Very Little	To Some Extent	To a Great Extent	Trouble Definitely Likely if Ignored at Selection

Appendix D

Study 1: Debriefing

Thank you for participating in this focus group aimed to identify spatial ability components in your occupation. Your involvement has helped us to further understand the tasks in your occupation and have helped us to determine the role spatial ability plays in it. Recently, the Canadian Forces has recognized that some aspects of spatial ability may be important for successful task performance in cognitively complex occupations, such as the Forward Air Controller and Aerospace Controller. As a general screening tool, the Canadian Forces currently uses the Canadian Forces Aptitude Test (CFAT) for assessing cognitive ability. As a subcomponent of the general cognitive ability measure, this test also includes a spatial ability component. However, the Canadian Forces has identified that spatial ability components other than the ones assessed using the CFAT may be important for success in their occupations. In particular, it has been identified that some individuals within the Canadian Forces experience difficulty in performing spatial tasks using egocentric and allocentric perceptions. In a previous examination we utilized findings from classic and current literature to understand how spatial ability can be best assessed and identified seven factors (i.e., spatial subcomponents) that may be related to spatial task performance. During the focus group we obtained on-the-job task information from you that we will attempt to categorize into the seven spatial subcomponents to determine which aspects of spatial ability are relevant to your occupation. We believe that the results of this focus group can serve as a foundation for further research on spatial ability relevance your occupation and for development of a more comprehensive, occupation-relevant, spatial ability test should it become necessary.

Research personnel:

Principle Investigators:
Sharmili Shan, Tel. ***-****; EMAIL: sshanmu5@connect.carleton.ca
Dr. Avi Parush, Tel. 520.2600 ext. 6026; EMAIL: avi_parush@carleton.ca

Other Research Personnel:
Dr. Cheryl Burgess, Tel. 996.3392; EMAIL: Cheryl.Burgess@forces.gc.ca
John Johnson; EMAIL: John.Johnston5@forces.gc.ca

Ethical concerns:
Dr. Monique Sénéchal - Chair, Carleton University Ethics Committee for Psychological Research, 613-520-2600, ext. 1155

Any other concerns:
Dr. Janet Mantler - Chair, Department of Psychology, Carleton University, 613-520-2600, ext. 4173

Appendix E

Cognitive and Spatial Requirements of Weapons Controllers

Table A1

Cognitive and Spatial Requirements of Weapons Division ATCOs

Spatial Task	Cognitive Ability	Spatial Ability
Routing: Where do I have to be & When?		
Where is my mission?	Locate, recall	Verbal & Reasoning, working memory
What is the situation now?	Analyze	Verbal & Reasoning, working memory
Predict: See the future	Project, compute, compare	Spatial Visualization, Mental rotation
Situation in 1 hour	Assess, project	Spatial Visualization, Verbal & Reasoning
Choose scramble heading/alternative	Think in degrees	Verbal & Reasoning
Assess enemy	Memory, take input from multiple inputs, have updated mental model	Verbal & Reasoning, Processing Speed
Assess friendly forces	Memory, take input from multiple inputs, have updated mental model	Verbal & Reasoning, Processing Speed
Verify position (heading, altitude, speed) of fighters	Compare	Mental rotation, Spatial Visualization
WD Considerations: Assess the 3D environment so as to be able to offer meaningful info to upper management		
Visualize track history	Visualize, memory	Spatial Imagery
Where are my assets (2D)	Locate	Spatial Visualization
Range	Compare	Mental rotation, Spatial Visualization

Whose airspace am I in? (3D)	Recall, Compare, visualize	Spatial Visualization
Time/Distance we can fly	Recall	Memory (but not working memory)
Gauge speed/time/distance	Estimate	Verbal & Reasoning
Target heading, altitude, speed (3D)	Estimate, compare, calculate	Verbal & Reasoning
Check for conflicting traffic (3D)	Project, compare, calculate	Spatial Visualization, Verbal & Reasoning
Climbing		
Descending		
Vectoring		
Time	Project, calculate, compare	Verbal & Reasoning, Spatial Visualization
To intercept		
SD Considerations		
Assess need for more assets	Assess situation	Verbal & Reasoning
Keep track of asset locations (airborne and ground based)	Visualize	Spatial Imagery, Working memory
Other assets coming into intercept from another area	Take info from multiple input, visualize	Spatial Visualization
Assess terrain (i.e., refueling, radar picture, radios)	Take info from multiple input, visualize	Spatial Visualization, Memory
Assign altitude in keeping with asset strengths	Compare, recall	Verbal & Reasoning, Memory
Distance to target (go higher if far, save fuel)	Estimate, calculate, compare	Verbal & Reasoning
WD PRO Active Control		
Measure a bearing and range	Measure	Verbal & Reasoning
Vector assisting assets IE, Tankers, RJ, SEAD	Compare, Measure, Recall, Manipulate frames of reference	Verbal & Reasoning, Visualization, Mental rotation
WD Reactive Control		
Target manoeuvres ((Do I have to manoeuvre)	Think in degrees, Compare, calculate	Verbal & Reasoning

Relative motion to target motion analysis (am I on the right vector)	Think in degrees, compare, calculate	Verbal & Reasoning, Spatial Visualization
Turn radius and speed relationship	Compare, Measure	Verbal & Reasoning, Spatial Visualization
Climb or decent required to get target	Judge relative distance	Verbal & Reasoning, Spatial Visualization
Trigonometry (intercept solutions)	Think in degrees, visualize, calculate	Spatial Visualization, Verbal & Reasoning
Uncontrollable Factors		
Go over, around, clearance to cross	Think in 3D, Manipulate frames of reference	Verbal & Reasoning, Mental rotation
Utilize compact airspace (ACMs)	Compare, compute, recall	Verbal & Reasoning, Spatial Visualization
HIDACZ		
ROZ		
Airspace boundaries	Visualize, memory	Spatial Visualization, Memory (but not working memory)
Traffic	Visualize, memory	Spatial Visualization, Working Memory
Areas	Visualize, memory	Spatial Visualization, Memory (but not working memory)
Danger	Recall	Memory
Restricted	Recall	Memory
Parachute	Recall	Memory
Glider areas	Recall	Memory
How big, how high	Recall	Memory
Control Mission Outcome		
Determine Conflict	Visualize, think in degrees, compute	Spatial Visualization, Verbal & Reasoning
Determine Boundaries or Restrictions	Visualize, memory	Spatial Visualization, Memory (but not working memory)

Target Handling Scenario

Target detection	Spatial awareness	Verbal & Reasoning
Observe route/tracks	Visual tracking	Working memory, Spatial Visualization
Identify target location	Compute	Processing Speed, Verbal & Reasoning
Visualize where target will be	Visualize, project	Spatial Visualization, Verbal & Reasoning
Assess if target is a threat	Compare, memory, take in info from multiple inputs	Verbal & Reasoning, Processing Speed
Assign mission		
Identify	Take info from multiple input, visualize	Spatial Visualization
Monitor	Visual tracking	Working Memory
Escort	Project	Verbal & Reasoning
Divert	Think in degrees	Verbal & Reasoning

Appendix F

Cognitive and Spatial Requirements of IFR Controllers

Table A2

Cognitive and Spatial Requirements of IFR Division ATCOs

Spatial Task	Cognitive Ability	Spatial Ability
Monitor Aircraft Emergency		
Aircraft doing what is supposed to be doing	Compare	Verbal & Reasoning
Monitor Aircraft Specific Issues		
Monitory Aircraft Lateral/Longitudinal		
Distance between aircrafts	Compute, think in degrees	Verbal & Reasoning
Direction of flight	Project	Spatial Visualization
Monitor heading/track	Project	Spatial Visualization
Monitor spacing on approach	Compute	Working memory
Monitor separation of aircraft	Think in 3D	Spatial Visualization
Monitor distance/range from a point (airport, facility)	Compute, think in degrees	Spatial Visualization
Monitor Aircraft Altitude		
Compare aircraft altitude to min IFR altitudes	Compare	Verbal & Reasoning
Monitor aircraft altitudes	Visualize, compare, evaluate/judge	Spatial Visualization, Verbal & Reasoning
Just right		
Too low		
Too high		

Monitor altitude versus terrain	Visualize, project	Spatial Visualization, Verbal & Reasoning
Monitor altitude	Visualize, project	Spatial Visualization, Verbal & Reasoning
Monitor Aircraft Speed		
Aircraft estimates	Compute	Verbal & Reasoning
Just right		
Too fast		
Too slow		
Monitor aircraft speed	Visualize, compare	Verbal & Reasoning
Look for aircraft on converging headings	Visualize, compare, think in degrees	Verbal & Reasoning, Spatial Visualization
Watch for aircraft not under your control and plan for when they call you	Project	Verbal & Reasoning
Identify		
Plan disposal of aircraft		
Project from active airspace	Project	Verbal & Reasoning
Build spacing for departures	Compute, compare	Spatial Visualization, Verbal & Reasoning
Judging separation	Compare, think in degrees	Verbal & Reasoning
Deconflict altitude allocation	Compare, visualize	Spatial Visualization, Verbal & Reasoning
Before issuing heading/altitude, integrate this action into other traffic – if it will not work adjust and adapt	Compare, visualize	Spatial Visualization, Verbal & Reasoning
Plan aircraft sequencing for approach	Project, distance estimation	Verbal & Reasoning

Plan		
Develop a plan	Visualize, compute, project, take info from multiple input	Spatial Visualization, Verbal & Reasoning
Plan traffic	Calculate, project	Verbal & Reasoning
Project		
Projecting flight path	Project	Spatial Visualization
Compare action of aircraft to your plan	Compare	Verbal & Reasoning
Appreciate/recognize VFR traffic	Identify	Spatial Visualization, working memory
Identify conflict	Compare, project, compute	Verbal & Reasoning
Finalize		
Not focusing on too small of an area on scope-need to look at big picture	Visualize	Spatial Visualization
Control		
Climb/descend aircraft	Visualize, compare, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Clear aircraft for approach	Visualize, compare, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Vector aircraft	Visualize, compare, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Resequencing traffic	Compute, compare	Verbal & Reasoning, Working memory, processing speed

Appendix G

Cognitive and Spatial Requirements of VFR Controllers

Table A3

Spatial and Cognitive Requirements of VFR Division ATCOs

Spatial Task	Cognitive Ability	Spatial Ability
Clearance		
Clear takeoff	Visualize, compare	Verbal & Reasoning, Spatial Visualization
Clear landing	Visualize, compare	Verbal & Reasoning, Spatial Visualization
Initial clearance	Visualize, compare	Verbal & Reasoning, Spatial Visualization, Mental rotation
Route over flights	Visualize, compare, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Control vehicle movements	Compare	Spatial Visualization
Taxi aircraft	Compare, compute	Verbal & Reasoning
Landing IFR/VFR aircraft	Compare, compute, Manipulate frames of reference	Verbal & Reasoning, Mental rotation
Miscellaneous Higher Level Tasks		
Resolve conflicts	Compare, project	Verbal & Reasoning, Processing speed
Determine if what you are being told reflects reality	Compare	Verbal & Reasoning, Processing speed
Radio Telecommunications		

Resolve conflicts	Compare, project	Verbal & Reasoning, Processing speed
Determine if what you are being told reflects reality	Compare	Verbal & Reasoning
Determine priority of tasks	Visualize, compare	Verbal & Reasoning, Working memory
Airdrome Management		
Clear takeoff	Visualize, compare	Verbal & Reasoning, Spatial Visualization
Clear landing	Visualize, compare, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Initial clearance	Visualize, compare, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Route over flights	Visualize, compare	Verbal & Reasoning, Spatial Visualization
Control vehicle movements	Compare	Spatial Visualization
Taxi aircraft	Compare, compute	Verbal & Reasoning
Landing IFR/VFR aircraft	Compare, compute, Manipulate frames of reference	Verbal & Reasoning, Mental rotation
Runway Management		
Control helo pads	Visualize, compare	Verbal & Reasoning
Decide where to put cable engagements	Visual, recall, compare	Verbal & Reasoning
Sequencing		
Decide where to fit aircraft (sequencing)	Compare, compute, Manipulate frames of reference	Verbal & Reasoning, Spatial Visualization, Mental rotation
Locate aircraft on radar	Identify	Spatial Visualization
Overhead vs. initial	Identify	Spatial Visualization
Project aircraft/vehicle paths forward in time	Project, calculate	Verbal & Reasoning, Spatial Visualization

Evaluate flight path profiles	Visualize, evaluate, project	Verbal & Reasoning, Spatial Visualization
Evaluate aircraft characteristics (speeds, profiles)	Identify, Recall	Verbal & Reasoning, Memory
Determine if spacing is enough for event to occur	Compute, visualize	Verbal & Reasoning, Spatial Visualization
Judging spacing/timing of events	Visualize, compare	Verbal & Reasoning, Spatial Visualization
Create windows for other events	Visualize, compute	Verbal & Reasoning, Spatial Visualization
Scan airspace to make sure everyone is where they should be/doing what you told them to do	Compare	Verbal & Reasoning, Working memory
Traffic		
Pass traffic	Compute	Verbal & Reasoning
Deciding if passing traffic is enough or if you need to physically separate aircraft	Compute, visualize	Spatial Visualization, Verbal & Reasoning, Working memory

Appendix H

Definitions

Table A4

Definition Comparison

Terminology	Definition
Study 1	
Visualization	Perceiving and brining into consciousness of spatial information from the environment. Ability to visualize how things will appear from various spatial orientations.
Mental Rotation	Determining where one is located with respect to nearby objects and the target location. Ability to mentally visualize and manipulate things (e.g., maps, images) in 3D.
Way-Finding	The use of physical motor movements to get from one point to another and involves both locomotion and way-finding components. Ability to mentally navigate around areas.
ETS (Ekstrom et al., 1976b)	
Visualization	“The ability to manipulate or transform the image of spatial patterns into other arrangements.” p. 173
Spatial Orientation	“The ability to perceive spatial patterns or to maintain orientation with respect to objects in space.” p. 149
Spatial Scanning	“Speed in exploring visually a wide or complicated spatial field.” p. 155

Appendix I

ETS Licensing

LICENSING AGREEMENT

THIS AGREEMENT, entered into as of December 15, 2011 between Educational Testing Service (hereinafter called "ETS"), a nonstock, nonprofit corporation organized and existing under the Education Law of the State of New York, with offices at Princeton, New Jersey 08541, and

Sharmili Shanmugaratuam
Carleton University
1125 Colonel By Drive
Ottawa, Ontario
Canada

(Hereinafter called "Licensee"),

WITNESSETH:

WHEREAS, ETS is the publisher and copyright owner of certain test materials; and

WHEREAS, Licensee wishes to produce editions of:

VZ-1 Form Board Test	VZ-2 Paper Folding Test
VZ-3 Surface Development Test	S-1 Card Rotations Test
S-2 Cube Comparisons Test	SS-1 Maze Tracing Speed Test
SS-2 Choosing A Path	SS-3 Map Planning Test

NOW, THEREFORE, ETS agrees that Licensee may reproduce and distribute 90 copies of each of the above editions for use in a research study, subject to the following terms and conditions:

1. Each copy of any edition produced under the Agreement shall bear a copyright notice exactly as it appears on the original test, followed by the statement: *Reproduced under license.*

2. ETS agrees to waive the licensing fee. Licensee will be responsible for any costs involved in the composition, reproduction, and distribution of the editions licenses herein.

LICENSING AGREEMENT (Continued)

3. Test copies or portions thereof are not to be included in research reports or other publications resulting from this use.

4. This license applies to use of the KIT material for a two-year period and will terminate on December 15, 2013, unless renewed in writing by ETS.

5. The license granted herein is nonexclusive and nontransferable to any third party without permission in writing by ETS.

6. Licensee warrants that the research materials licensed hereunder will be used under supervision of Licensee, for research purposes only, and not for counseling, or other operational purposes.

7. ETS warrants that it has the unrestricted right to grant the license specified herein. Licensee will hold ETS harmless from any liability or claims resulting from Licensee's use of these materials, including consequential damages.

8. It is agreed that any changes in the terms and conditions of this Agreement must be in writing and approved by an authorized representative of ETS.

9. This Agreement shall be considered null and void if it is not signed within 30 days.

LICENSEE	EDUCATIONAL TESTING SERVICE
Name: SHARMILI SHANMUGARATNAM	Name: Danelly A. Diaz
Title: RESEARCHER	Title: Copyright Admin., Copyright Group
Date: Dec 22, 2011	Date: December 15, 2011

ETS®

LICENSING AGREEMENT

THIS AGREEMENT, entered into as of August 5, 2013 between Educational Testing Service (hereinafter called "ETS"), a nonstock, nonprofit corporation organized and existing under the Education Law of the State of New York, with offices at Princeton, New Jersey 08541, and

Sharmili Shanmugaratnam
1125 Colonel By Drive, Psychology Dept, Loeb Building
Ottawa, Ontario K1S 5B6
Canada

(Hereinafter called "Licensee"),

WITNESSETH:

WHEREAS, ETS is the publisher and copyright owner of certain test materials; and

WHEREAS, Licensee wishes to produce editions of:

Form Board Test – VZ-1; Surface Development Test – VZ-3;
Card Rotations Test – S-1 (Rev.); Map Planning Test – SS-3

NOW, THEREFORE, ETS agrees that Licensee may reproduce and distribute 155 copies of each of the above editions for use in a research study, subject to the following terms and conditions: plus 40 additional copies of the MAP PLANNING TEST *FA 8/20/13

1. Each copy of any edition produced under the Agreement shall bear a copyright notice exactly as it appears on the original test, followed by the statement: *Reproduced under license.*

2. Licensee agrees to pay ETS a licensing fee of $93.00. Payment shall be sent to Educational Testing Service to the attention of:

Faiza Aziem M/S 42-L
Office of General Counsel
660 Rosedale Road-42-L
Princeton, New Jersey 08541

Licensee will be responsible for any costs involved in the composition, reproduction, and distribution of the editions licensed herein.

Request #: 27357

LICENSING AGREEMENT (Continued)

3. Test copies or portions thereof are not to be included in research reports or other publications resulting from this use.

4. This license applies to use of the KIT material for a two-year period and will terminate on August 5, 2015, unless renewed in writing by ETS.

5. The license granted herein is nonexclusive and nontransferable to any third party without permission in writing by ETS.

6. Licensee warrants that the research materials licensed hereunder will be used under supervision of Licensee, for research purposes only, and not for counseling, or other operational purposes.

7. ETS warrants that it has the unrestricted right to grant the license specified herein. Licensee will hold ETS harmless from any liability or claims resulting from Licensee's use of these materials, including consequential damages.

8. It is agreed that any changes in the terms and conditions of this Agreement must be in writing and approved by an authorized representative of ETS.

9. This agreement shall be considered null and void if not signed, paid and returned within 30 days of the date of this letter.

<u>The signing of this license shall constitute agreement of the above terms and shall be considered a binding contract once a countersigned copy is returned to you.</u>

LICENSEE	EDUCATIONAL TESTING SERVICE
Name: SHARMILI SHANMUGARATNAM	Name: Fyiza Azeem
Title: Researcher	Title: Copyright Administrator
Date: August 21, 2013	Date: 9/11/2013

ETS®

Request #: 27357

Appendix J

NASA-TLX

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low Very High

Appendix K

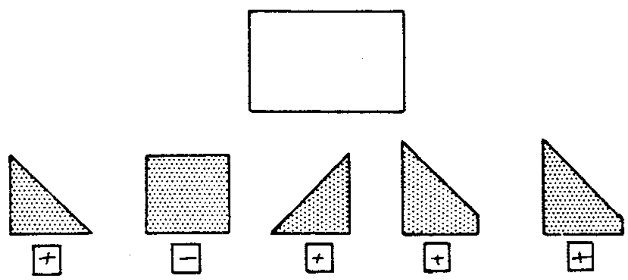
Form Board Test

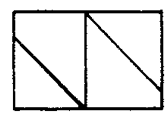
FORM BOARD TEST — VZ-1

This is a test of your ability to tell what pieces can be put together to make a certain figure. Each test page is divided into two columns. At the top of each column is a geometrical figure. Beneath each figure are several problems. Each problem consists of a row of five shaded pieces. Your task is to decide which of the five shaded pieces will make the complete figure when put together. Any number of shaded pieces, from two to five, may be used to make the complete figure. Each piece may be turned around to any position but it cannot be turned over. It may help you to sketch the way the pieces fit together. You may use any blank space for doing this. When you know which pieces make the complete figure, mark a plus (+) in the box under ones that are used and a minus (-) in the box under ones that are not used.

In Example A, below, the rectangle can be made from the first, third, fourth, and fifth pieces. A plus has been marked in the box under these places. The second piece is not needed to make the rectangle. A minus has been marked in the box under it. The rectangle drawn to the right of the problem shows one way in which the four pieces could be put together.

A.

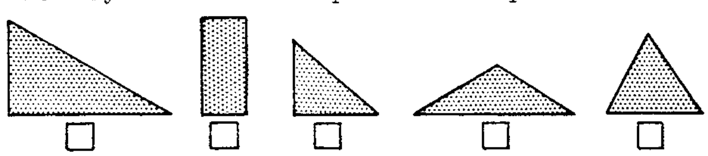




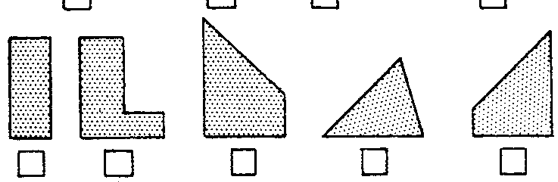
Answer

Now try to decide which pieces in Examples B and C will make the rectangle.

B.



C.



In Example B, the first, fourth, and fifth pieces are needed. You should have marked a plus under these three pieces and a minus under the other two pieces. In Example C, the second, third, and fifth pieces should be marked with a plus and the first and fourth with a minus.

Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you have some idea whether or not the piece is correct.

You will have 8 minutes for each of the two parts of this test. Each part has 2 pages. When you have finished Part 1 (pages 2 and 3), STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

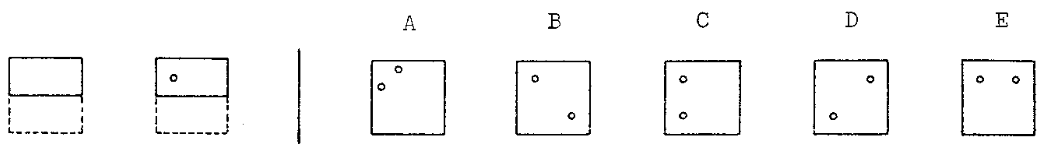
Appendix L

Paper Folding Test

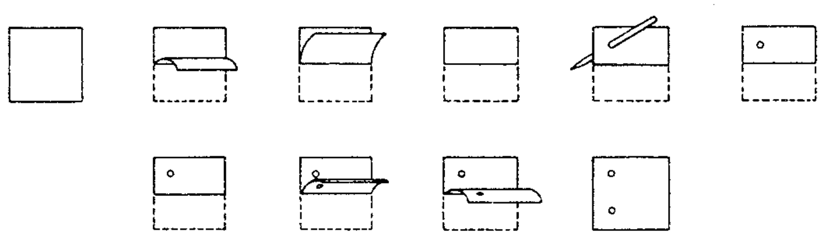
PAPER FOLDING TEST — VZ-2

In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

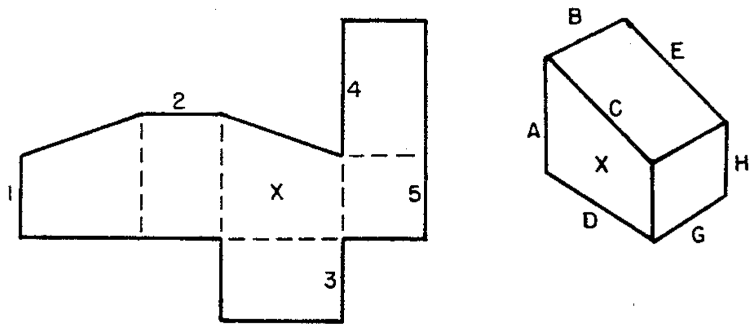
Appendix M

Surface Development Test

SURFACE DEVELOPMENT TEST — VZ-3

In this test you are to try to imagine or visualize how a piece of paper can be folded to form some kind of object. Look at the two drawings below. The drawing on the left is of a piece of paper which can be folded on the dotted lines to form the object drawn at the right. You are to imagine the folding and are to figure out which of the lettered edges on the object are the same as the numbered edges on the piece of paper at the left. Write the letters of the answers in the numbered spaces at the far right.

Now try the practice problem below. Numbers 1 and 4 are already correctly marked for you.



1:	H
2:	
3:	
4:	C
5:	

NOTE: The side of the flat piece marked with the X will always be the same as the side of the object marked with the X. Therefore, the paper must always be folded so that the X will be on the outside of the object.

In the above problem, if the side with edge 1 is folded around to form the back of the object, then edge 1 will be the same as edge H. If the side with edge 5 is folded back, then the side with edge 4 may be folded down so that edge 4 is the same as edge C. The other answers are as follows 2 is B; 3 is G; and 5 is H. Notice that two of the answers can be the same.

Your score on this test will be the number of correct letters minus a fraction of the number of incorrect letters. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have 6 minutes for each of the two parts of this test. Each part has 2 pages. When you have finished Part 1 (pages 2 and 3), STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Appendix N

Card Rotations Test

CARD ROTATIONS TEST — S-1 (Rev.)

This is a test of your ability to see differences in figures. Look at the 5 triangle-shaped cards drawn below.

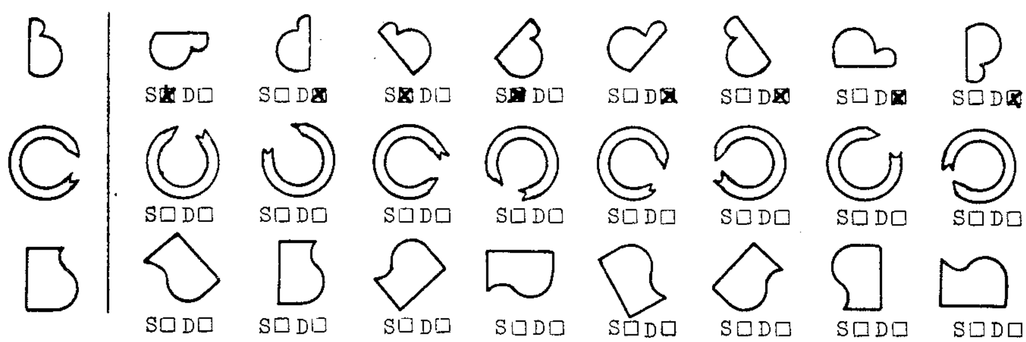
All of these drawings are of the same card, which has been slid around into different positions on the page.

Now look at the 2 cards below:

These two cards are not alike. The first cannot be made to look like the second by sliding it around on the page. It would have to be flipped over or made differently.

Each problem in this test consists of one card on the left of a vertical line and eight cards on the right. You are to decide whether each of the eight cards on the right is the same as or different from the card at the left. Mark the box beside the S if it is the same as the one at the beginning of the row. Mark the box beside the D if it is different from the one at the beginning of the row.

Practice on the following rows. The first row has been correctly marked for you.



Your score on this test will be the number of items answered correctly minus the number answered incorrectly. Therefore, it will not be to your advantage to guess, unless you have some idea whether the card is the same or different. Work as quickly as you can without sacrificing accuracy.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

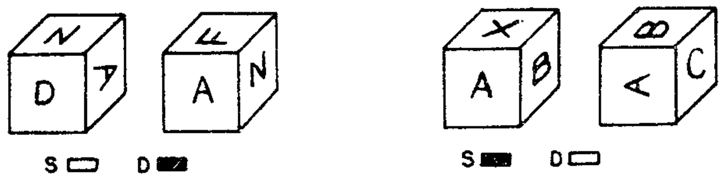
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Appendix O

Cube Comparisons Test

CUBE COMPARISONS TEST -- S-2 (Rev.)

Wooden blocks such as children play with are often cubical with a different letter, number, or symbol on each of the six faces (top, bottom, four sides). Each problem in this test consists of drawings of pairs of cubes or blocks of this kind. Remember, there is a different design, number, or letter on each face of a given cube or block. Compare the two cubes in each pair below.

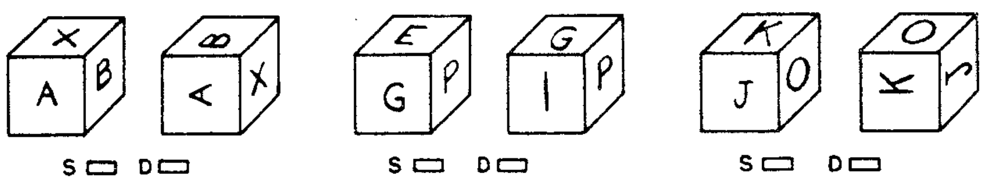


The first pair is marked D because they must be drawings of <u>different</u> cubes. If the left cube is turned so that the A is upright and facing you, the N would be to the left of the A and hidden, not to the right of the A as is shown on the right hand member of the pair. Thus, the drawings must be of different cubes.

The second pair is marked S because they could be drawings of the same cube. That is, if the A is turned on its side the X becomes hidden, the B is now on top, and the C (which was hidden) now appears. Thus the two drawings could be of the same cube.

<u>Note</u>: No letters, numbers, or symbols appear on more than one face of a given cube. Except for that, <u>any</u> letter, number or symbol can be on the hidden faces of a cube.

Work the three examples below.



The first pair immediately above should be marked D because the X cannot be at the peak of the A on the left hand drawing and at the base of the A on the right hand drawing. The second pair is "different" because P has its side next to G on the left hand cube but its top next to G on the right hand cube. The blocks in the third pair are the same, the J and K are just turned on their side, moving the O to the top.

Your score on this test will be the number marked correctly minus the number marked incorrectly. Therefore, it will <u>not</u> be to your advantage to guess unless you have some idea which choice is correct. Work as quickly as you can without sacrificing accuracy.

You will have <u>3 minutes</u> for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP.

DO NOT TURN THE PAGE UNTIL YOU ARE ASKED TO DO SO.

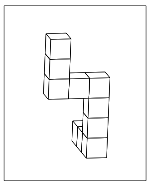
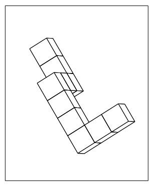
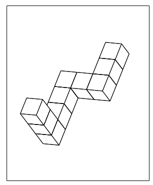
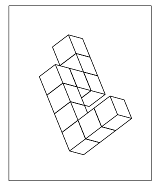
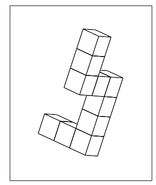
Appendix P

The Revised Vandenberg-Kuse Mental Rotations Test

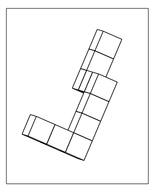
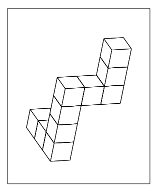
MENTAL ROTATIONS TEST (MRT-A)

This test is composed of the figures provided by Shepard and Metzler (1978), and is, essentially, an Autocad-redrawn version of the Vandenberg & Kuse MRT test.

Please look at these five figures

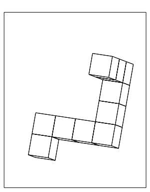
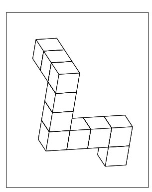
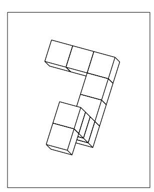
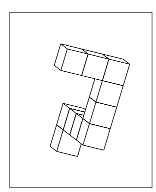
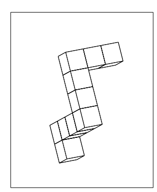
Note that these are all pictures of the same object which is shown from different angles. Try to imagine moving the object (or yourself with respect to the object), as you look from one drawing to the next.

Here are two drawings of a new figure that is different from the one shown in the first 5 drawings. Satisfy yourself that these two drawings show an object that is different and cannot be "rotated" to be identical with the object shown in the first five drawings.

Now look at this object: — Two of these four drawings show the same object. Can you find those two? Put a big X across them.

1.

If you marked the first and third drawings, you made the correct choice.

Appendix Q

Object Perspective Taking Test

Spatial Orientation Test

This is a test of your ability to imagine different perspectives or orientations in space. On each of the following pages you will see a picture of an array of objects and an “arrow circle” with a question about the direction between some of the objects. For the question on each page, you should imagine that you are standing at one object in the array (which will be named in the center of the circle) and facing another object, named at the top of the circle. Your task is to draw an arrow from the center object showing the direction to a third object from this facing orientation.

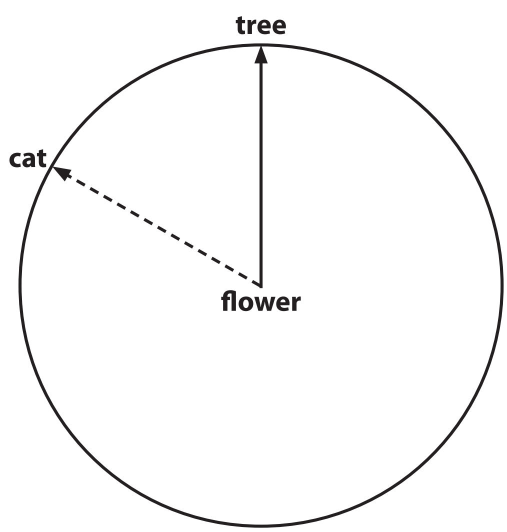
Look at the sample item on the next page. In this item you are asked to imagine that you are standing at the flower, which is named in the center of the circle, and facing the tree, which is named at the top of the circle. Your task is to draw an arrow pointing to the cat. In the sample item this arrow has been drawn for you. In the test items, your task is to draw this arrow. Can you see that if you were at the flower facing the tree, the cat would be in this direction? Please ask the experimenter now if you have any questions about what you are required to do.

There are 12 items in this test, one on each page. For each item, the array of objects is shown at the top of the page and the arrow circle is shown at the botom. Please do not pick up or turn the test booklet, and do not make any marks on the maps. Try to mark the correct directions but do not spend too much time on any one question.

You will have 5 minutes for this test.

Spatial Orientation Test **Name:** ____________________

Example:
Imagine you are standing at the **flower** and facing the **tree**.
Point to the **cat**.



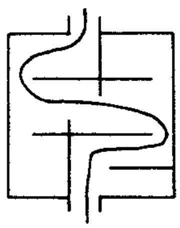
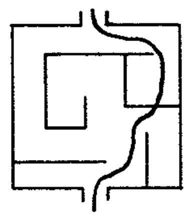
Appendix R

Mace Tracing Speed Test

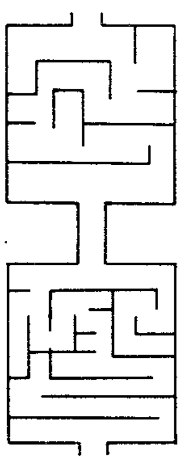
MAZE TRACING SPEED TEST — SS-1

This is a test of your ability to find a path through a maze quickly. You are to draw a pencil line through each maze without having to cross any printed lines.

Look at the two drawings below. In the left square a pencil line has been drawn to show the correct path from top to bottom. The square on the right shows an incorrect path. It is incorrect because the pencil line crosses a printed line.

CORRECT  INCORRECT 

Practice for speed on the squares below. Remember, you must make a pencil line through each square without having to cross a printed line.



Your score on this test will be the number of squares through which a line has been correctly drawn. If you should become stuck in any square, you may skip to the following one. You should try to avoid making mistakes, but you will not be penalized for lifting your pencil, for retracing a path that leads to a dead end, or for accidentally crossing lines at the sides of the path being taken. Work as quickly as you can without sacrificing accuracy. On the test, follow the squares around the page the way that they are connected, starting at the top of the left-hand column.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

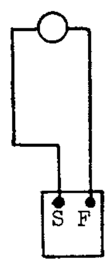
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Appendix S

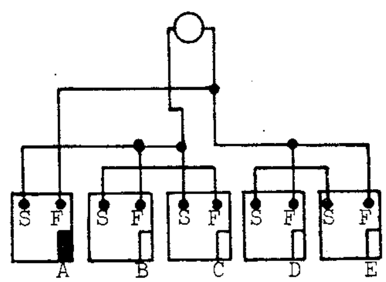
Choosing a Path Test

CHOOSING A PATH -- SS-2

This is a test of your ability to choose a correct path from among several choices. In the picture below is a box with dots marked S and F. S is the starting point and F is the finish. You are to follow the line from S, through the circle at the top of the picture and back to F.



In the problems in this test there will be five such boxes. Only one box will have a line from the S, through the circle, and back to the F in the same box. Dots on the lines show the only places where connections can be made between lines. If lines meet or cross where there is no dot, there is no connection between the lines. Now try this example. Show which box has the line through the circle by blackening the space at the lower right of that box.



The first box is the one which has the line from S, through the circle, and back to F. The space lettered A has therefore been blackened.

Each diagram in the test has only one box which has a line through the circle and back to the F. Some lines are wrong because they lead to a dead end. Some lines are wrong because they come back to the box without going through the circle. Some lines are wrong because they lead to other boxes that do not have lines going through the circle.

GO ON TO THE NEXT PAGE

Appendix T

Sample Screenshot of the Multiple Objects Tracking and Projection Tests

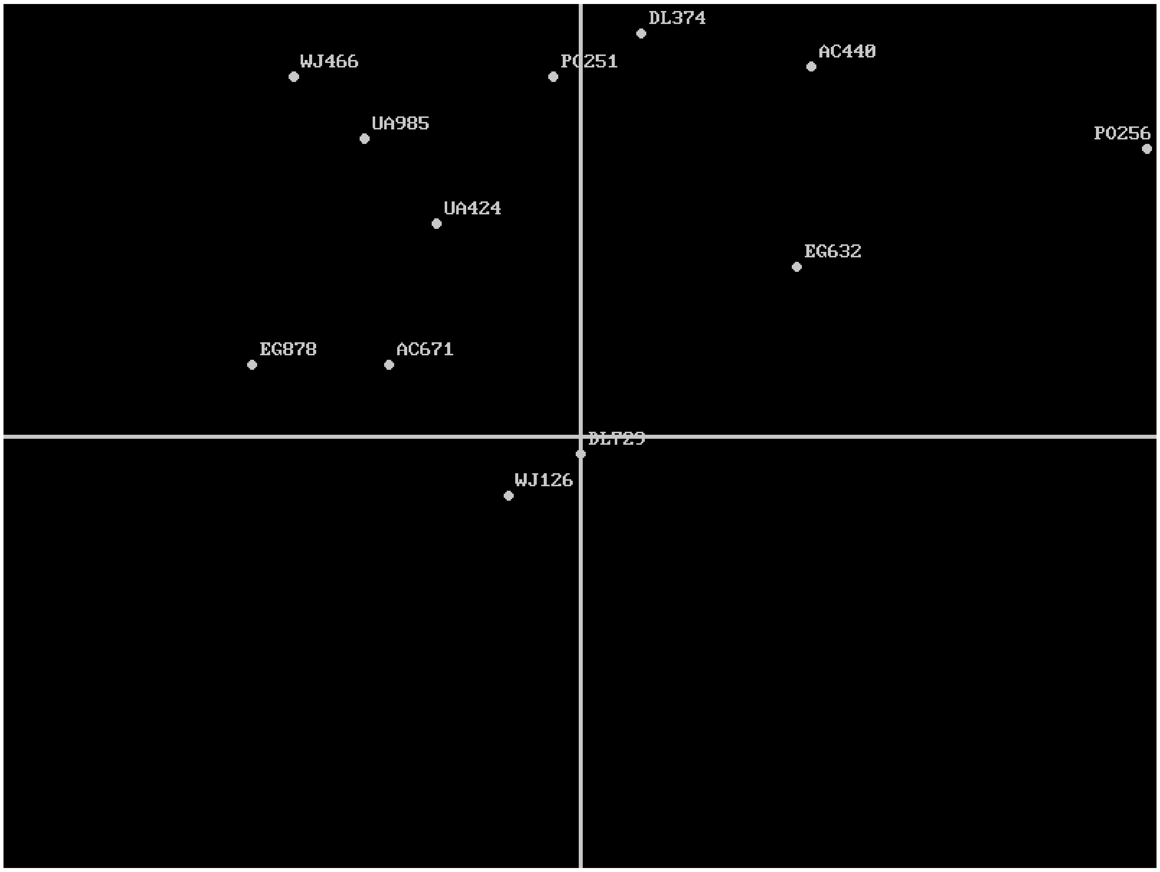


Figure A1. Sample screenshot of the MOT and Projection Tests

Appendix U

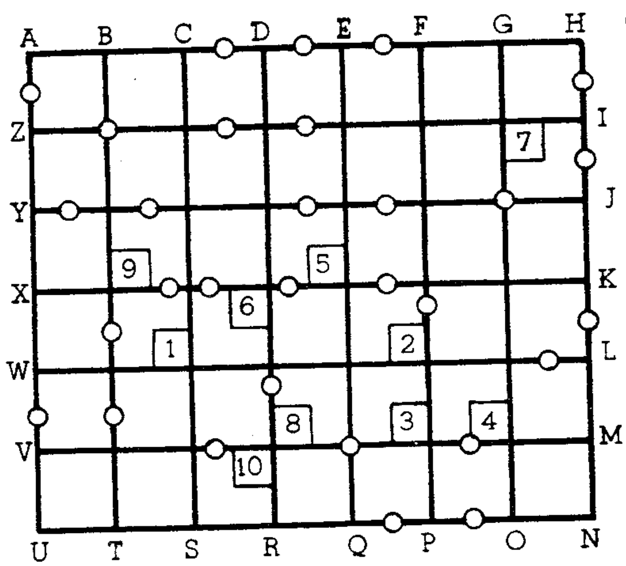
Map Planning Test

MAP PLANNING TEST — SS-3

This is a test of your ability to find the shortest route between two places as quickly as possible. The drawing below is a map of a city. The dark lines are streets. The circles are road-blocks, and you cannot pass at the places where there are circles. The numbered squares are buildings. You are to find the shortest route between two lettered points. The number on the building passed is your answer.

Rules:
1. The shortest route will always pass along the side of one and only one of the numbered buildings.
2. A building is not considered as having been passed if a route passes only a corner and not a side.
3. The same numbered building may be used on more than one route.

Look at the sample map below. Practice by finding the shortest route between the various points listed at the right of the map. The first problem has been marked correctly.



The shortest route from:	Passes building:
1. A to Z	1
2. E to S	
3. P to J	
4. V to K	
5. O to F	
6. G to M	
7. D to Q	
8. F to T	

The answers to the other practice problems are as follows: 2 passes 5; 3 passes 3; 4 passes 2; 5 passes 4; 6 passes 4; 7 passes 6; 8 passes 5.

Your score on this test will be the number of right answers. It will not be to your advantage to guess unless you have some idea which route is correct. Work as rapidly as you can without sacrificing accuracy.

You will have 3 minutes for each of the two parts of this test. Each part has one page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

Appendix V

Sample Screenshot of the Spatial Task

Figure A2. Sample screenshot of the spatial task

Appendix W

Study 2: Informed Consent

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement. The informed consent has to provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Study Title: Do you have the spatial abilities it takes to be a good Air Traffic Controller?
Study Personnel:
Sharmili Shan, sshanmu5@connect.carleton.ca
Dr. Avi Parush, Tel. 613-520-2600 ext. 6026, avi_parush@carleton.ca

If you have any ethical concerns about this study, please contact Dr. Chris Davis (acting for the Chair of Ethics, 613-520-2600 ext. 2251) or Dr. Anne Bowker (Chair, Psychology Department, 613-520-2600 ext. 8218).

Purpose and Task Requirements: The purpose of this study is to understand whether spatial ability of individuals can be best measured through general spatial ability tests, which are not similar to the task, or through task-specific spatial tests, which has high similarity to task. In order to determine this, you will be asked to complete seven tests of spatial ability (five paper-and-pencil tests, two on a computer) and a game on a computer that requires you to point, click, and drag planes to their designated destinations as indicated on the screen. The game play will be video screen captured for further analysis at a later time. The screen capture will only capture the gameplay; therefore it will not record any personal information beyond game performance. The game performance captured on the screen will show how individuals directed each aircraft to their destination, the number of path changes individuals made, the number of aircraft landed, the number of close calls (i.e., almost plane crashes), and the number of planes individuals were managing at the time of the first plane crash. During the tests and the game, you will be asked to perform a second task that will take your attention away from the test and compete with your attention in the game. The entire study will last approximately 3 hours.

Potential Risk and Discomfort: There are no physical or psychological risks in this study. If you feel anxious and/or uncomfortable in any way about your involvement in this study, please bring your concerns to the researcher's attention. You are permitted to take breaks between each of the tests.

Anonymity/Confidentiality: The data collected in this experiment are confidential. All data are coded such that your name is not associated with the data. The coded data will only be available to the researchers involved in this study. The electronic data will be stored on a password-protected computer while the paper documents will be stored in a safe room. The personal information collected in the demographics questionnaire and the video screen capture of your gameplay will be destroyed at the completion of this project (approximately in 2013).

Right to Withdraw: Your participation in this study is entirely voluntary. At any point during the study, you have the right to not answer certain questions or to withdraw from the study. You will still be provided with 3 bonus credits.

I have read the above description of "Do you have the spatial abilities it takes to be a good Air Traffic Controller?" and understand the conditions of my participation. My signature indicates that I agree to participate in the study, and that I agree to have my game performance video captured. This in no way constitutes a waiver of my rights.

Full Name (please print): ______________________________

Participant Signature: ______________________________

Date: ______________________________

Researcher Signature: ______________________________

Date: ______________________________

This study has received clearance by the Carleton University Psychology Research Ethics Board (11-148).

Demographic Questionnaire

Participant No: ______

The purpose of this questionnaire is to obtain basic information regarding the participants in this study. Using this information, I will be able to investigate the results of my study at a more in depth level. If you are uncomfortable providing any of the requested information, please leave that section blank.

Sex:
- ☐ Male
- ☐ Female

Age: ____

Degree: ____

Year of study: __________________________________

Vision:
- ☐ Normal
- ☐ Corrected-to-normal
- ☐ Other

Colour vision:
- ☐ Yes
- ☐ No

Do you have any experience with computer games?
- ☐ No
- ☐ Yes

  Please sate which type and which games:
  Strategy: ____
  1st person: ____
  Other: ____

Do you do puzzles or play other spatial board games?
- ☐ No
- ☐ Yes

  Please state which one(s): ____

Did you ever take any standardized tests?
- ☐ No
- ☐ Yes

  Please state which one(s): ____

Appendix X

Study 2: Debriefing

Study Title: Do you have the spatial abilities it takes to be a good Air Traffic Controller?

Thank you for participating in this study. Your time and effort are greatly appreciated!

The purpose of this study is to better understand whether individuals' ability to preform spatial tasks (an air traffic controller-like task in this study) can be best measured through general spatial ability tests (e.g., Vandenberg-Kuse mental rotation) or through specific, task related, tests (e.g., multiple object tracking, projection). The traditional view is that spatial ability can be measured using general tests of the ability even though there is low cognitive and physical similarity between the general tests and the tasks. However, given that task-related tests are more cognitively, and sometimes physically, similar to the task being measured, specific tests should be better predictors of task performance. Therefore, finding an answer to the current research question could impact how current spatial ability measurements are taken.

Please feel free to contact us if you would like to discuss any aspect of the research. The following people are involved in this study:

Sharmili Shan, e-mail: sshanmu5@connect.carleton.ca
Dr. Avi Parush, Tel: 613.520.2600 ext. 6026, e-mail: avi_parush@carleton.ca

If you have any ethical concerns about this study, please contact:

Dr. Monique Sénéchal (Chair, Carleton University Psychology Research Ethics Board 613-520-2600 ext. 1155)

or

Dr. Anne Bowker (Chair, Psychology Department 613-520-2600 ext. 8218)

Interested in finding out more about spatial abilities? Check out these articles:

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*, 151-176.

Parush A., & Berman, (2004). Orientation and navigation in 3D user interfaces: The impact of navigation aids and landmarks. *International Journal of Human Computer Studies*, *61(3)*, 375-395.

Tang, M. (2006). Gender differences in relationship between background experiences and three levels of spatial ability. *Unpublished doctoral dissertation, The Ohio State University*. Ohio, United States of America.
[Available at http://etd.ohiolink.edu/view.cgi?acc_num=osu1155573195]

This study has received clearance by the Carleton University Psychology Research Ethics Board (11-148).

Appendix Y

Study 2: Data Cleaning and Assumptions

Missing Data

For each of the MOT and Projection Tests, the data for the 12 trials were organized into three blocks of four trials using mean values. Missing data analysis revealed that less than 5% of the data were missing per variable, thus the data were assumed to be missing completely at random and no further protocol was taken to account for missing data.

Normality

Normality was assessed using Shapiro-Wilk test and the test revealed that most variables were not normally distributed. Outliers were examined and standard scores less than -3.29 or greater than +3.29 were attempted to be rectified using *log* and *square root* transformations. If transformations failed to remedy an offending value, then that value was brought within range, while maintaining relative rank in the dataset, such that it is no longer an outlier. These procedures resulted in some variables becoming more normally distributed.

Practice Effects

Analysis of Variance (ANOVA) was used to examine practice effects between the three blocks of the MOT and Projection Tests as well as the three trials of the ATC game and clock task. The results revealed practice effect for some variables and not others. If there were no significant differences in performances between blocks/trials, then all three blocks/trials were averaged to obtain a single value for that measure. If there was a significant difference in performance between all the blocks/trials, then regression analyses were carried out with separate values for those blocks/trials, resulting in multiple values for one measure (e.g., MOT Response Time Block 1, MOT Response Time Block 2, MOT Response Time Block 3). If the results revealed differences in performance for one block/trial and not the other two, then the values for the two blocks/trials with no difference were averaged and treated as a single measure. Sometimes difference in performance was observed only between two of the three blocks/trials; in such instances, the third block/trial was not included in regression analyses.

Sphericity

If Sphericity assumption, using Mauchly's Test, was not met for any of the ANOVA tests, then either the Huynh-Feldt or Greenhouse-Geisser corrections were used depending on whether the estimated epsilon is greater or less than .75, respectively (as recommended by Girden, 1992).

Multicollinearity and Singularity

Multicollinearity and singularity were first examined using Pearson correlations. High correlations (e.g., over .80) indicate that two variables are too highly correlated and may violate these assumptions (Tabachnick & Fidell, 2006). Multicollinearity was examined using Tolerance and VIF values. A tolerance value less than .20 (Menard, 1995, as cited in O'Brien, 2007) and a VIF value larger than 10 or average VIF value substantially greater than 1 (StataCorp, 1997, as cited in O'Brien, 2007) indicates that two or more independent variables are highly correlated. The Pearson correlation results indicated two separate variable pairs that may be of concern and the multicollinearity results confirmed this observation. For each of the two highly correlated pairs, only one of the variables was kept. This resulted in acceptable Tolerance and VIF values.

Appendix Z

Study 2: List of Variables After Data Cleaning

Table A5

List of Variables After Data Cleaning

Variable	Description
Job-non-specific	
MRT	Mental Rotation Test score
Card Rotation	Card Rotation Test score
Surface Development	Surface Development Test score
Cube Comparison	Cube Comparison Test score
Map Tracing	Map Tracing Test score
Choosing a Path	Choosing a Path Test score
Form Board	Form Board Test score
Paper Folding	Paper Folding Test score
Map Planning	Map Planning Test score
Spatial Orientation	Spatial Orientation Test score
Job-specific	
MOT CR	Mean MOT correct responses for blocks* 1, 2, and 3
MOT RT B23	Mean MOT reaction time for blocks 2 and 3 combined
Projection CR B12	Mean Projection Test correct responses for blocks 1 and 2 combined
Projection CR B3	Mean Projection Test correct responses for block 3
Projection Lead Time B1	Howe far in advance crashes were detected in block 1
Projection Lead Time B2	Howe far in advance crashes were detected in block 2
Projection Lead Time B3	How far in advance crashes were detected in block 3
Projection FA	Mean Projection false alarm rate for blocks 1, 2, and 3

*Note: Blocks refer to the early, mid, and late trials of the MOT and Projection Tests.

Block 1 = Trials 1-4, Block 2 = Trials 5-8, Block 3 = Trials 9-12

Appendix AA

Study 2: Regression Models Used

Model 1
- Step 1: Paper and Pencil Tests
- Step 2: MOT variables
- Step 3: Projection variables

Model 2
- Step 1: Paper and Pencil Tests
- Step 2: Projection variables
- Step 3: MOT variables

Model 3
- Step 1: MOT variables
- Step 2: Projection variables
- Step 3: Paper and Pencil Tests

Model 4
- Step 1: Projection variables
- Step 2: Paper and Pencil Tests
- Step 3: MOT variables

Model 5
- Step 1: MOT variables
- Step 2: Paper and Pencil Tests
- Step 3: Projection variables

Model 6
- Step 1: Projection variables
- Step 2: MOT variables
- Step 3: Paper and Pencil Tests

Model 7
- Step 1: Paper and Pencil Tests
- Step 2: MOT + Projection variables

Model 8
- Step 1: MOT + Projection variables
- Step 2: Paper and Pencil Tests

Model 9
- Step 1: Paper and Pencil Tests
- Step 2: Time-related Variables
- Step 3: Accuracy-related Variables

Model 10
- Step 1: Time-related Variables
- Step 2: Paper and Pencil Tests
- Step 3: Accuracy-related Variables

Model 11
- Step 1: Time-related Variables
- Step 2: Accuracy-related Variables
- Step 3: Paper and Pencil Tests

Model 12
- Step 1: Paper and Pencil Tests + Time-related Variables
- Step 2: Accuracy-related Variables

Model 13
- Step 1: Paper and Pencil Tests + Accuracy-related Variables
- Step 2: Time-related Variables

Model 14
- Step 1: Time-related Variables
- Step 2: Paper and Pencil Tests + Accuracy-related Variables

List of Variables

Paper and Pencil Tests

Combined:
- Mental Rotations Test
- Card Rotation Test

Group 1:
- Mental Rotations Test
- Card Rotation Test
- Map Tracing Test
- Surface Development Test

Group 2:
- Mental Rotations Test
- Card Rotation Test
- Map Planning Test
- Form Board Test

MOT Variables
- Mean Correct Response
- Reaction Time B23

Projection Variables
- Lead Time Block 1
- Lead Time Block 2
- Lead Time Block 3
- Mean False Alarm
- Correct Response B12
- Correct Response B3

Time-related Variables
- MOT Reaction Time B23
- Projection Lead Time B1
- Projection Lead Time B2
- Projection Lead Time B3

Accuracy-related Variables
- MOT Mean Correct Response
- Mean False Alarm
- Correct Response B12
- Correct Response B3

Appendix AB

Study 2: Results of Model 1 for Overlapping Variables Analysis

Table A6

Summary of Combined Hierarchical Regression Analysis for Variables Predicting Number of Planes Landed for Trials 1 and 2 of the ATC Game (N = 64)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Card Rotation	.09	.03	.37**	.09	.03	.34**	-.10	.04	.42**
MRT	.21	.18	.15	.26	.19	.19	.15	.20	.11
MOT									
Correct Response				-2.51	2.29	-.15	-2.75	2.39	-.17
Response Time				-5.10	2.62	-.25	-4.68	2.65	-.23
Projection									
Lead Time B1							1.25	.74	.32
Lead Time B2							.25	.68	.06
Lead Time B3							.27	.91	.06
False Alarm							-10.09	6.79	-.29
Correct Response B12							-1.06	.87	-.17
Correct Response B3							.78	.70	.14

Note. R^2 = .19 for Step 1 (*ps* < .01); ΔR^2 = .06 for Steps 2 (*ps* = ns), ΔR^2 = .09 for Steps 3 (*ps* = ns). ***p* < .01

Table A7

Summary of Combined Hierarchical Regression Analysis for Variables Predicting Number of Path Changes for Trial 1of the ATC Game (N = 64)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Card Rotation	.00	.01	.05	.00	.01	.01	.00	.01	.09
MRT	.00	.03	.03	.02	.03	.22	.01	.03	.03
MOT									
Correct Response				-.66	.35	-.27	-.63	.36	-.26
Response Time B23				-1.23	.40	-.42**	-1.17	.42	-.40**
Projection									
Lead Time B1							.09	.12	.16
Lead Time B2							-.06	.11	-.11
Lead Time B3							.14	.14	.20
False Alarm							-.144	1.07	-.28
Correct Response B12							-.04	.14	-.04
Correct Response B3							-.11	.11	-.14

Note. R^2 = .01 for Step 1; Δ R^2 = .16 for Step 2 (*ps* < .01); Δ R^2 = .05 for Step 3 (*ps* = ns). ***p* < .01.

Table A8

Summary of Combined Hierarchical Regression Analysis for Variables Predicting Game Duration for Trial 1 of the Game (N = 64)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
Card Rotation	.35	.23	.20	.28	.23	.16	.39	.25	.22
MRT	1.80	1.32	.18	2.39	1.33	.24	1.75	1.44	.18
MOT									
Correct Response				-26.16	16.24	-.22	-23.70	16.87	-.20
Response Time B23				-53.49	18.61	-.37**	-52.50	18.72	-.37**
Projection									
Lead Time B1							3.88	5.20	.14
Lead Time B2							2.68	4.83	.10
Lead Time B3							7.90	6.44	.23
False Alarm							-57.73	48.01	-.24
Correct Response B12							-13.57	6.16	-.31*
Correct Response B3							-.69	4.98	-.02

Note. R^2 = .10 for Step 1; Δ R^2 = .12 for Step 2 (*ps* < .05); Δ R^2 = .10 for Step 3 (*ps* = ns). ***p* < .01.

Appendix AC

Study 2: Results of Model 5 for Group 1 Variables Analysis

Table A9

Summary of Group 1 Hierarchical Regression Analysis for Variables Predicting Number of Planes Landed for Trials 1 and 2 of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	-1.05	3.64	-.07	-7.44	3.43	-.47*	-8.56	4.68	-.54
Response Time B23	-5.55	4.37	-.29	-2.62	3.91	-.14	-1.78	4.94	-.09
Card Rotations				.04	.07	.13	.06	.10	.19
MRT				.61	.25	.44*	.67	.34	.49
Map Tracing				.36	.20	.38	.25	.26	.27
Surface Development				.10	.09	.23	.15	.11	.34
Projection									
Lead Time B1							-.31	1.29	-.08
Lead Time B2							1.25	1.53	.22
Lead Time B3							.05	1.85	.01
False Alarm							5.37	12.83	.14
Correct Response B12							-.71	1.59	-.11
Correct Response B3							.70	1.35	.13

Note. R^2 = .07 for Step 1; Δ R^2 = .46 for Step 2 (*ps* < .01); Δ R^2 = .05 for Step 3 (*ps* = ns). ** *p* < .01
Table A10

Table A10

Summary of Group 1 Hierarchical Regression Analysis for Variables Predicting Number of Path Changes for Trial 1of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	-1.12	.45	-.52*	-1.44	.57	-.63*	-1.41	.72	-.62
Response Time B23	-1.45	.54	-.53*	-1.49	.65	-.55*	-1.54	.76	-.57
Card Rotations				.00	.01	.00	.00	.02	-.06
MRT				.04	.04	.22	.05	.05	.23
Map Tracing				.00	.03	.00	.03	.04	.21
Surface Development				.00	.01	.04	-.01	.02	-.20
Projection									
Lead Time B1							-.03	.20	-.05
Lead Time B2							.13	.24	.16
Lead Time B3							.19	.28	.21
False Alarm							-1.30	1.97	-.23
Correct Response B12							-.01	.24	-.02
Correct Response B3							-.38	.19	-.49

Note. R^2 = .27 for Step 1 (p < .05); ΔR^2 = .05 for Step 2 (*ps* = ns); ΔR^2 = .15 for Step 3 (*ps* = ns). *p < .05

Table A11

Summary of Group 1 Hierarchical Regression Analysis for Variables Predicting Number of Path Changes for Trial 3 of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	-.31	.53	-.13	-1.14	.63	-.48	-2.07	.58	-.87**
Response Time B2+3	-.09	.64	-.03	-.46	.71	-.16	-.87	.62	-.31
Card Rotations				-.01	.01	-.26	-.03	.01	-.68*
MRT				.04	.05	.21	.12	.04	.58*
Map Tracing				.02	.04	.15	.06	.03	.42
Surface Development				.03	.02	.41	.02	.01	.33
Projection							-.48	.16	-.84**
Lead Time B1									
Lead Time B2							.64	.19	.76**
Lead Time B3							.56	.23	.60*
False Alarm							.01	1.60	.00
Correct Response B12							.13	.20	.13
Correct Response B3							-.48	.16	-.58**

Note. R^2 = .01 for Step 1; ΔR^2 = .20 for Step 2 (*ps* = ns); ΔR^2 = .45 for Step 3 (*ps* < .05). ***p* < .01, **p* < .05

Table A12

Summary of Group 1 Hierarchical Regression Analysis for Variables Predicting Number of Alerts for Trials 1, 2, and 3 of the Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	.15	2.31	.01	-5.37	2.12	-.52*	-6.68	2.75	-.65*
Response Time B23	-.49	.28	-.40	-.04	2.42	.00	.29	2.90	.02
Card Rotations				-.01	.04	-.05	-.02	.06	-.09
MRT				.24	.15	.27	.36	.20	.41
Map Tracing				.23	.12	.37	.25	.15	.40
Surface Development				.16	.05	.58**	.17	.07	.61*
Projection									
Lead Time B1							-.68	.76	-.28
Lead Time B2							1.18	.90	.32
Lead Time B3							.34	1.09	.08
False Alarm							4.80	7.54	.19
Correct Response B12							.01	.93	.00
Correct Response B3							-.47	.73	-.13

Note. R^2 = .00 for Step 1; Δ R^2 = .51 for Step 2 (*ps* < .05); Δ R^2 = .08 for Step 3 (*ps* = ns). ***p* < .01, **p* < .05

Appendix AD

Study 2: Results of Model 3 for Group 2 Variables Analysis

Table A13

Summary of Group 2 Hierarchical Regression Analysis for Variables Predicting Number of Planes Landed for Trials 1 and 2 of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	1.57	3.21	.09	.77	4.19	.04	3.98	3.49	.23
Response Time B23	-7.26	4.12	-.32	-8.06	4.77	-.36	-5.47	4.04	-.34
Projection									
Lead Time B1				.18	1.35	.42	2.05	1.15	.49
Lead Time B2				.42	1.00	.14	-1.67	.98	-.56
Lead Time B3				.23	1.43	.06	2.60	1.23	.66
False Alarm				.47	9.77	.01	-21.85	8.84	-.71*
Correct Response B12				.05	1.30	.01	-.82	1.01	-.13
Correct Response B3				.86	1.27	.16	1.07	1.03	.20
Card Rotations							.27	.07	1.32**
MRT							-1.10	.37	-.77**
Map Planning							-.73	.28	-.79*
Form Board							.05	.03	.38

Note. R^2 = .17 for Step 1; ΔR^2 = .09 for Step 2 (*ps* = ns); ΔR^2 = .43 for Step 3 (*ps* = <.01). $**p < .01$, $*p < .05$

Table A14

Summary of Group 2 Hierarchical Regression Analysis for Variables Predicting Number of Planes Landed for Trial 3 of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	13.40	4.86	.44*	12.20	5.89	.40	16.57	4.37	.54**
Response Time B23	-13.49	6.26	-.34*	-13.98	6.70	-.35*	-5.69	5.06	-.14
Projection									
Lead Time B1				-.23	1.90	-.03	1.57	1.43	.21
Lead Time B2				1.49	1.40	.29	-1.37	1.12	-.26
Lead Time B3				-1.90	2.01	-.27	1.23	1.54	.17
False Alarm				-2.81	13.71	-.07	-28.68	11.05	-.53*
Correct Response B12				2.71	1.83	.25	2.09	1.26	.20
Correct Response B3				2.01	1.79	.21	1.08	1.29	.11
Card Rotations							.20	.80	.56
MRT							-1.81	.46	-.73**
Map Planning							-.36	.25	-.22
Form Board							.12	.03	.56**

Note. R^2 = .33 for Step 1; ΔR^2 = .16 for Step 2 (*ps* = ns); ΔR^2 = .33 for Step 3 (*ps* < .01). ***p* < .01, **p* < .05

Table A15

Summary of Group 2 Hierarchical Regression Analysis for Variables Predicting Number of Path Changes for Trial 1of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	.37	.43	.15	.35	.56	.14	1.03	.51	.41
Response Time B23	-.15	.55	-.45*	-1.16	.63	-.36	-.84	.59	-.26
Projection									
Lead Time B1				.14	.18	.23	.26	.17	.43
Lead Time B2				-.09	.13	-.21	-.34	.13	-.79*
Lead Time B3				.19	.19	.33	.52	.18	.90*
False Alarm				-1.55	1.30	-.34	-3.83	1.29	-.86
Correct Response B12				.12	.17	.14	.07	.15	.08
Correct Response B3				-.08	.17	-.10	.02	.15	.03
Card Rotations							.03	.01	1.12**
MRT							-.18	.05	-.87**
Map Planning							-.13	.04	-1.00**
Form Board							.01	.00	.31

Note. R^2 = .23 for Step 1; Δ R^2 = .10 for Step 2 (*ps* = ns); Δ R^2 = .30 for Step 3 (*ps* < .05). ***p* < .01, **p* < .05

Table A16

Summary of Group 2 Hierarchical Regression Analysis for Variables Predicting Number of Path Changes for Trial 3 of the ATC Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	1.58	.57	.44*	1.82	.60	.50**	2.16	.71	.60**
Response Time B23	-1.67	.73	-.36*	-1.59	.69	-.34*	-1.44	.83	-.31
Projection									
Lead Time B1				.40	.19	.46	.46	.23	.53
Lead Time B2				-.10	.14	-.15	-.21	.18	-.34
Lead Time B3				-.32	.21	-.39	-.17	.25	-.21
False Alarm				-1.45	1.40	-.23	-2.50	1.81	-.40
Correct Response B12				.07	.19	.05	.04	.21	.03
Correct Response B3				.28	.18	.25	.32	.21	.29
Card Rotations							.02	.01	.35
MRT							-.09	.08	-.30
Map Planning							-.06	.06	-.31
Form Board							.00	.01	.10

Note. R^2 = .35 for Step 1; Δ R^2 = .26 for Step 2 (*ps* = ns); Δ R^2 = .03 for Step 3 (*ps* = ns). *p < .05

Table A17

Summary of Group 2 Hierarchical Regression Analysis for Variables Predicting Game Duration for Trial 3 of the Game (N = 32)

	Step 1			Step 2			Step 3		
Predictor	*B*	*SE B*	*β*	*B*	*SE B*	*β*	*B*	*SE B*	*β*
MOT									
Correct Response	79.65	23.33	.52**	70.63	28.64	.46*	77.78	28.89	.50*
Response Time B23	-61.75	30.03	-.31*	-73.42	32.60	-.37*	-37.45	33.45	-.19
Projection									
Lead Time B1				-3.67	9.24	-.10	1.58	9.49	.04
Lead Time B2				6.59	6.80	.25	-3.02	7.40	-.12
Lead Time B3				-9.59	9.79	-.28	.09	10.21	.00
False Alarm				34.20	66.75	.13	-44.58	73.15	-.17
Correct Response B12				11.25	9.91	.21	9.59	8.37	.18
Correct Response B3				12.72	8.70	.27	7.66	5.56	.16
Card Rotations							.54	.55	.30
MRT							-4.90	3.02	-.39
Map Planning							-.80	2.32	-.10
Form Board							.52	.21	.47*

Note. R^2 = .39 for Step 1; ΔR^2 = .13 for Step 2 (*ps* = ns); ΔR^2 = .16 for Step 3 (*ps* = ns). $^{**}p < .01$

Appendix AE

Subjective Workload Rating (SWR)

Rate how you feel about the test, on a scale of 1 to 10, by circling the appropriate number:

Easy									Difficult
1	2	3	4	5	6	7	8	9	10

Relaxing									Stressful
1	2	3	4	5	6	7	8	9	10

Required Minimal Effort									Labour Intensive
1	2	3	4	5	6	7	8	9	10

Pleasant									Annoying
1	2	3	4	5	6	7	8	9	10

Calming									Frustrating
1	2	3	4	5	6	7	8	9	10

Simple									Complex
1	2	3	4	5	6	7	8	9	10

Appendix AF

Arithmetic Questions Used as Interruption Task

54 + 16 =	57 + 19 =	17 + 54 =
75 + 16 =	63 + 28 =	38 + 39 =
67 + 24 =	23 + 48 =	63 + 28 =
28 + 44 =	37 + 16 =	77 + 13 =
75 + 16 =	74 + 17 =	48 + 33 =
11 + 49 =	34 + 57 =	47 + 19 =
31 + 59 =	26 + 56 =	63 + 18 =
48 + 45 =	73 + 18 =	14 + 77 =
37 + 29 =	71 + 19 =	31 + 49 =
29 + 71 =	15 + 37 =	55 + 39 =
13 + 77 =	22 + 49 =	71 + 19 =
74 + 26 =	76 + 15 =	36 + 55 =
38 + 53 =	25 + 65 =	58 + 33 =
46 + 45 =	52 + 37 =	73+ 14 =
69 + 29 =	16 = 75 =	36 + 16 =
15 + 26 =	37 + 47 =	47 + 27 =
34 + 68 =	18 + 75 =	31 + 49 =
49 + 31 =	53 + 29 =	72 + 18 =
27 + 64 =	77 + 15 =	56 + 25 =
45 + 56 =	13 + 68 =	24 + 67 =
66 + 24 =	68 + 25 =	58 + 18 =
36 + 65 =	16 + 76 =	41 + 49 =
19 + 38 =	47 + 19 =	75 + 16 =
52 + 27 =	12 + 39 =	39 + 54 =
28 + 73 =	65 + 17 =	13 + 38 =
81 + 19 =	74 + 19 =	18 + 18 =
44 + 37 =	37 + 17 =	59 + 17 =
79 + 11 =	48 + 38 =	44 + 37 =
25 + 35 =	31 + 59 =	21 + 69 =
88 + 11 =	27 + 34 =	73 + 19 =
59 + 32 =	78 + 16 =	68 + 13 =
71 + 19 =	38 + 39 =	19 + 75 =
53 + 29 =	43 + 38 =	28 + 38 =
69 + 24 =	18 + 63 =	33 + 48 =
55 + 18 =	52 + 38 =	25 + 46 =
38 + 43 =	35 + 56 =	72 + 19 =
47 + 39 =	26 + 66 =	37 + 47 =
64 + 17 =	73 + 18 =	61 + 29 =
77 + 14 =	32 + 49 =	53 + 17 =
66 + 25 =	58 + 16 =	14 + 69 =

Note. The questions participants received were presented in 36pt font size to increase visibility.

Appendix AG

Study 3: Informed Consent and Debriefing

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement. The informed consent has to provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Study Title: Do you have the spatial abilities it takes to be a good Air Traffic Controller?
Study Personnel:

Sharmili Shan, sshanmu5@connect.carleton.ca
Dr. Avi Parush, Tel. 613-520-2600 ext. 6026, avi_parush@carleton.ca

If you have any ethical concerns about this study, please contact Dr. Chris Davis (acting for the Chair of Ethics, 613-520-2600 ext. 2251) or Dr. Anne Bowker (Chair, Psychology Department, 613-520-2600 ext. 8218).

Purpose and Task Requirements: The purpose of this study is to understand whether spatial ability of individuals can be best measured through general spatial ability tests, which are not similar to the task, or through task-specific spatial tests, which has high similarity to task. In order to determine this, you will be asked to complete seven tests of spatial ability (five paper-and-pencil tests, two on a computer) and a game on a computer that requires you to point, click, and drag planes to their designated destinations as indicated on the screen. The game play will be video screen captured for further analysis at a later time. The screen capture will only capture the gameplay; therefore it will not record any personal information beyond game performance. The game performance captured on the screen will show how individuals directed each aircraft to their destination, the number of path changes individuals made, the number of aircraft landed, the number of close calls (i.e., almost plane crashes), and the number of planes individuals were managing at the time of the first plane crash. During the tests and the game, you will be asked to perform a second task that will take your attention away from the test and compete with your attention in the game. The entire study will last approximately 2.5 hours.

Potential Risk and Discomfort: There are no physical or psychological risks in this study. If you feel anxious and/or uncomfortable in any way about your involvement in this study, please bring your concerns to the researcher's attention. You are permitted to take breaks between each of the tests.

Anonymity/Confidentiality: The data collected in this experiment are confidential. All data are coded such that your name is not associated with the data. The coded data will only be available to the researchers involved in this study. The electronic data will be stored on a password-protected computer while the paper documents will be stored in a safe room. The personal information collected in the demographics questionnaire and the video screen capture of your gameplay will be destroyed at the completion of this project (approximately in 2013).

Right to Withdraw: Your participation in this study is entirely voluntary. At any point during the study, you have the right to not answer certain questions or to withdraw from the study. You will still be provided with 2.5 bonus credits.

I have read the above description of "Do you have the spatial abilities it takes to be a good Air Traffic Controller?" and understand the conditions of my participation. My signature indicates that I agree to participate in the study, and that I agree to have my game performance video captured. This in no way constitutes a waiver of my rights.

Full Name (please print): ____

Participant Signature: ____

Date: ____

Researcher Signature: ____

Date: ____

This study has received clearance by the Carleton University Psychology Research Ethics Board (12-166).

Study Title: Do you have the spatial abilities it takes to be a good Air Traffic Controller?

Thank you for participating in this study. Your time and effort are greatly appreciated!

The purpose of this study is to determine how we can better predict whether or not an individual will be a good Air Traffic Controller. In order to seek answers to this research question, you were asked to perform computerized tests, paper and pencil tests, as well as to play a game that required you to perform Air Traffic Control (ATC)-related tasks. Looking at your scores on each of the tests and how you performed on the game will allow the researcher to determine if any of the tests were good at predicting your performance on the game. The researcher is also interested in determining whether the paper and pencil tests you completed, which have been known to measure general spatial ability of individuals, or the computerized tests, which are conceptually similar to the game, are better at predicting your performance on the game. The result from this portion of the study will help determine which of the tests you completed are good at predicting performance on an ATC-related task (i.e., the game in this study) and whether these tests are general measures of spatial ability (i.e., in paper and pencil format) or are more similar to the ATC-task (i.e., computerized format).

An additional interest of the researcher is to determine whether various factors in Air Traffic Controllers' working environment, such as frequent interruptions and dealing with heavy workload, affect how these individuals perform on their ATC-related tasks. Given that Air Traffic Controllers perform their job tasks under high workload and with multiple interruptions, the researcher is interested in determining whether tests that are administered in similar conditions may be better predictors of performance than tests administered under standard test-taking conditions provided by test developers. In order to assess this, this study had three conditions: In the condition you participated, you performed all the tests according to the instructions provided by the test developers. In another condition, participants will be asked to perform the tests under high workload (e.g., by requiring them to perform the computerized tests with more distractor dots on the screen). In the third condition, participants will be interrupted at various times while they perform all the computerized and paper and pencil tests. Looking at individuals' performance on the various tests in all three conditions will aid the researcher to determine whether tests that were performed with interruption and or/ higher workload are better predictors of individuals' performance on the game compared to tests administered according to test developers' instructions.

Overall, the results of this study will aid in identifying:

- which of the tests administered is best predictive of game performance.
- whether paper and pencil tests or computerized tests are better at predicting game performance.
- whether making the test taking and the spatial task situations similar (i.e., by increasing workload or by interrupting individuals during testing) will result in tests that are highly predictive of game performance.

The results from this study can aid in developing a better selection test for Air Traffic Controllers. A better test could help select individuals who have the necessary skills to perform tasks in an Air Traffic Controller occupation, thus reducing any potential job-related errors due to individuals not having adequate skills to perform spatial tasks on the job.

Please feel free to contact us if you would like to discuss any aspect of the research. The following people are involved in this study:

Sharmili Shan, e-mail: sshanmu5@connect.carleton.ca
Dr. Avi Parush, Tel: 613.520.2600 ext. 6026, e-mail: avi_parush@carleton.ca

If you have any ethical concerns about this study, please contact:
Dr. Chris Davis (acting for the Chair of Ethics, 613-520-2600 ext. 2251) or
Dr. Anne Bowker (Chair, Psychology Department 613-520-2600 ext. 8218)

Interested in finding out more about spatial abilities? Check out this article:

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*, 151-176.

Interested in learning more about the Air Traffic Controller Occupation? Visit How Stuff Work's website:
Freudenrich, C. (n.d.). How air traffic control works. Retrieved from http://science.howstuffworks.com/transport/flight/modern/air-traffic-control.htm

Want to find out how you can become an Air Traffic Controller? Have a look at the Nav Canada website:
Nav Canada (n.d.). Air Navigation Careers. Retrieved from https://takecharge.navcanada.ca/en/content/careers/ifr

This study has received clearance by the Carleton University Psychology Research Ethics Board (12-166).

Appendix AH

Study 3: Data Cleaning and Assumptions

The initial data cleaning procedure differed from Study 2 in three ways in order to be more stringent and thorough. First, the data for all twelve trials of the MOT and Projection Tests were examined individually prior to consolidating the trials into three blocks of 4 trials each. This was done to ensure that extreme outlier variables were dealt with at the trial, rather than block, level. Second, no special procedure was taken in Study 2 to deal with missing data given that less than 5% of the data were missing for each variable (Tabachnick & Fidell, 2006). Data cleaning for the Projection Test and ATC game for this study revealed that greater than 5% of the data was missing for some variables. This is due to individuals not making any selection for some trials of the Projection Test; this was not an issue in Study 2 since an average score was used for each block. Missing data for the ATC game occurred since some trials were discounted because they ended due to clicking-related difficulty experienced by participants. Missing data were accounted for using an average score of 5 imputed values generated by the SPSS software. Any imputed values that resulted in out of range values for variables were brought within range. Third, outliers in Study 2 were attempted to be rectified using transformations or by bring bringing them within range. In this study, influential outliers, rather than outliers, were examined using Studentized Deleted Residuals in order to avoid unnecessarily altering the data. No influential outliers were detected and therefore no special means were taken to deal with them.

Normality, practice effects, singularity, multicollinearity, sphericity, and homogeneity of variance were assessed using the same procedure as in Study 2. All variables violated the normality assumption. As with Study 2, this violation was not addressed due to the robustness of regression analyses. Practice effects analysis revealed differences in performance between Blocks for some of the MOT and Projection variables. Practice effect was not observed for any of the ATC game variables. Singularity was established given that no variables had a correlation of .80 or higher. Multicollinearity was not detected given Tolerance values were well above .20 and VIF values were below 5 and did not have a mean VIF value far greater than 1. Any violations in sphericity or homogeneity of variance were corrected using methods discussed in Study 2.

Appendix AI

Studies 2 and 3: Clock Task Performance Comparison

A comparison of ATC clock task performance between Studies 2 and 3 revealed that only the accuracy rate in Trial 1 and the number of missed questions in Trial 2 were significantly different between the two studies. Participants in Study 3 were more accurate for Trial 1 of the game (M = 71.88%, SD = 23.02%) than those in Study 2 (M = 63.02%, SD = 25.96%), $t(238) = -2.55$, $p < .05$ (two-tailed). The mean decrease in accuracy was 8.85 with 95% confidence interval ranging from -15.71 to -2.00, and a small to medium effect size ($\eta^2 = -0.37$). Participants in Study 3 missed significantly more questions for Trial 2 of the game (M = .21, SD = .52) than those in Study 2 (M = .07, SD = .29, $t(200.67) = -2.63$, $p < .05$. The mean decrease in number of questions missed was -0.14, with 95% confidence intervals ranging from -0.25 to -0.04 and a small effect size ($\eta^2 = -0.30$).

Appendix AJ

Study 3: Number of Questions Completed

Table A18

Number of Questions Completed for Each Paper and Pencil Test by Condition

	Condition		
Test	Control	Workload	Interruption
Spatial Visualization			
Form Board (240)	163.10	180.41	157.03
Surface Development (60)	47.00	49.89	48.62
Spatial Orientation			
Card Rotation (160)	114.98	123.54	114.24
MRT (24)	7.74	9.19	7.68
Spatial Scanning			
Map Planning (40)	25.29	28.31	24.68

Note. Maximum number of questions for each test is presented in parentheses beside test name.

Appendix AK

Study 3: Factor Analysis Results

Factor analysis using all job-specific and job-non-specific test measures initially revealed five factors. However some variables overlapped across 3 factors and one variable did not fit into a factor, which were grounds to remove these variables. A reanalysis with these variables removed revealed three factor that fell into three categories: job-non-specific spatial ability tests (Cronbach's Alpha = .64), two of the Projection Test measures (Cronbach's Alpha = .31), and correct response measures for MOT and Projection Tests (Cronbach's Alpha = .49). The communality values ranged from .49 to .78, which indicates that 100 to 200 participants are required to have reliable results (MacCallum, Widaman, Zhang, & Hong, 1999, as cited in Field, 2005). Given that the data were analyzed across all three conditions, the 176 participants should be sufficient to produce reliable results. However, Cronbach's Alpha values for all three factors were lower than the acceptance level of .70 (Nunnally, 1978), which suggests that these factors may not be reliable. The factor with all five job-non-specific spatial ability tests approached acceptance level, revealing that job-non-specific spatial ability tests could potentially be regarded as a single factor, regardless of the fact that they appear to measure different aspects of spatial ability. The analysis did not reveal spatial planning ability as a single factor given that the measures for the Projection and Map Planning Tests did not fall into a single factor.

Appendix AL

Study 4: Informed Consent and Debriefing

The purpose of an informed consent is to ensure that you understand the purpose of the study and the nature of your involvement. The informed consent has to provide sufficient information such that you have the opportunity to determine whether you wish to participate in the study.

Study Title: Do you have the abilities it takes to be a good Air Traffic Controller?

Study Personnel:
Sharmili Shan, sharmili_shanmugaratnam@carleton.ca
Dr. Avi Parush, avi_parush@carleton.ca

If you have any ethical concerns about this study, please contact Dr. Shelly Brown (Chair of Ethics, 613-520-2600 ext. 1505) or Dr. Anne Bowker (Chair, Psychology Department, 613-520-2600 ext. 8218).

Purpose and Task Requirements: The purpose of this study is to understand whether spatial tests can accurately predict your performance on a spatial task. In order to determine this, you will be asked to complete two tests of spatial ability and a game on a computer that requires you to point, click, and drag planes to their designated destinations as indicated on the screen. The game play will be video screen captured for analysis at a later time. The screen capture will only capture the gameplay; therefore it will not record any personal information beyond game performance. The game performance captured on the screen will show how individuals directed each aircraft to their destination, the number of path changes individuals made, the number of aircraft landed, the number of close calls (i.e., almost plane crashes), and the number of planes individuals were managing at the time of the first plane crash. Tower of London performance score and the number of planes landed in the ATC game are automatically generated by the respective programs so, if you wish to know these scores, please request it during the briefing period (i.e., end of study). The entire study will last approximately 1 hour.

Potential Risk and Discomfort: There are no physical or psychological risks in this study. If you feel anxious and/or uncomfortable in any way about your involvement in this study, please bring your concerns to the researcher's attention. You are permitted to take breaks between each of the tests.

Anonymity/Confidentiality: The data collected in this experiment are confidential. All data are coded such that your name is not associated with the data. The coded data will only be available to the researchers involved in this study. The electronic data will be stored on a password-protected computer while the paper documents will be stored in a safe room. The personal information collected in the demographics questionnaire and the video screen capture of your gameplay will be destroyed at the completion of this project (approximately in Summer 2014).

Right to Withdraw: Your participation in this study is entirely voluntary. At any point during the study, you have the right to not answer certain questions or to withdraw from the study. You will still be provided with 1.0 bonus credit.

I have read the above description of "Do you have the spatial abilities it takes to be a good Air Traffic Controller?" and understand the conditions of my participation. My signature indicates that I agree to participate in the study, and that I agree to have my game performance video captured. This in no way constitutes a waiver of my rights.

Full Name (please print): ____

Participant Signature: ____

Date: ____

Researcher Signature: ____

Date: ____

This study has received clearance by the Carleton University Psychology Research Ethics Board (13-136).

Study Title: Do you have the abilities it takes to be a good Air Traffic Controller?

Thank you for participating in this study. Your time and effort are greatly appreciated!

The purpose of this study is to determine how we can better predict whether or not an individual will be a good Air Traffic Controller (ATCO). My previous study revealed that planning-related tests are good at predicting individuals' ability to perform various ATCO-related tasks if they were administered under high workload environment that mimics real working environment of ATCOs. The purpose of this study is to confirm my previous finding using a classic test of planning ability (Tower of London Test) along with a test that was developed for my research (Projection Test).

Given that ATCOs perform their job tasks under high workload, I am interested in determining whether tests that are administered in similar conditions may be better predictors of performance than tests administered under standard test-taking conditions provided by test developers. In order to assess this, this study had two conditions: In the condition you participated, you took the tests according to the instructions provided by the test developers. In the other condition, participants were asked to take the same tests, but their workload was slightly increased (e.g., by having more distractor dots on the screen in the Projection Test). Looking at individuals' performance on the tests in both conditions will allow me to confirm whether tests that were performed under high workload are better predictors of individuals' performance on the game compared to tests administered according to test developers' instructions. Additionally, the results will allow me to determine if we can administer any general planning tests when selecting individuals for the ATCO position or if the tests need to be relevant to the job. The results from this study can aid in developing a better selection test for ATCOs. A better test could help select individuals who are most qualified for this job.

Please feel free to contact us if you would like to discuss any aspect of the research. The following people are involved in this study:

Sharmili Shan, sharmili_shanmugaratnam@carleton.ca
Dr. Avi Parush, avi_parush@carleton.ca

If you have any ethical concerns about this study, please contact:
Dr. Shelly Brown (Chair of Ethics, 613-520-2600 ext. 1505) or
Dr. Anne Bowker (Chair, Psychology Department 613-520-2600 ext. 8218)

Interested in finding out more about spatial abilities? Check out this article:

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence, 34*, 151-176.

Interested in learning more about the Air Traffic Controller Occupation? Visit How Stuff Work's website:

Freudenrich, C. (n.d.). How air traffic control works. Retrieved from http://science.howstuffworks.com/transport/flight/modern/air-traffic-control.htm

Want to find out how you can become an Air Traffic Controller? Have a look at the Nav Canada website:

Nav Canada (n.d.). Air Navigation Careers. Retrieved from https://takecharge.navcanada.ca/en/content/careers/ifr

This study has received clearance by the Carleton University Psychology Research Ethics Board (13-136).

Appendix AM

Study 4: Data Cleaning and Assumptions

All data, except for the ATC Duration for third trial, violated the normality assumption. Once again, due to the robustness of regression analysis, no further steps were taken to normalize the data. No influential outliers were detected (i.e., all SDR values were below 3). No practice effects were detected for any of the ATC game performance measures or for the TOL Test solution time. Practice effects were detected for two of the three Projection Test measures and were dealt with accordingly. Singularity and multicollinearity assumptions were met given that there were no correlations over .80 and Tolerance and VIF values were within acceptable range, respectively. Any violations in sphericity or homogeneity of variance were dealt with as in Studies 2 and 3.

Appendix AN

Studies 2, 3, and 4: ATC Game Performance Comparison

Table A19

ATC Performance for Studies 2, 3, and 4

	Study					
	2 (n = 64)		3 (n = 176)		4 (n = 73)	
Performance	*M*	*SD*	*M*	*SD*	*M*	*SD*
Planes Landed	17.69	6.57	19.34	6.13	20.84	7.09
Duration (Seconds)	141.01	35.39	152.15	31.15	157.31	33.11
Alerts	10.20	4.32	9.55	3.73	9.89	3.49
Path Changes	3.05	2.78	3.48	3.11	3.20	2.44

Appendix AO

Studies 2, 3, and 4: Clock Task Performance Comparison

A comparison of clock task performance between Studies 2, 3, and 4 revealed that the only potential difference in performance resulted in Missed questions for Trial 2. While participants in Studies 3 (M = .23, SD = .47) and 4 (M = .29, SD = .66) missed comparable number of questions, the number of missed questions in Study 2 (M = .07, SD = .29) approached significance, $F(2, 126) = 3.49$, $p = .046$.

Appendix AP

Study 4: Regression Models Used

Three unique model combinations were developed and, for each model, the variables entered in Step 1 and Step 2 were exchanged to obtain a total of 6 models.

Model 1:
- Step 1: Projection variables
- Step 2: TOL variables

Model 2
- Step 1: TOL variables
- Step 2: Projection variables

Model 3
- Step 1: Time-related variables
- Step 2: Non-time-related variables

Model 4
- Step 1: Non-time-related variables
- Step 2: Time-related variables

Model 5
- Step 1: Time-related variables
- Step 2: Accuracy-related variables

Model 6
- Step 1: Accuracy-related variables
- Step 2: Time-related variables

List of Variables

Projection Variables
- Lead Time Blocks 1 and 3 Combined
- Lead Time Block 2
- Mean Accuracy
- False Alarm Rate Block 2
- False Alarm Rate Block 3

TOL Variables
- Score
- Solution Time (average per move)

Time-related Variables
- Projection Lead Time Blocks 1 and 3 Combined
- Projection Lead Time Block 2
- TOL Solution Time (average per move)

Non-time-related Variables
- Mean Projection Accuracy
- Projection False Alarm Rate Block 2
- Projection False Alarm Rate Block 3
- TOL Score

Accuracy-related Variables
- Mean Projection Accuracy
- TOL score

Appendix AQ

CFAT Spatial Ability Test Instructions

PRACTICE TEST 2

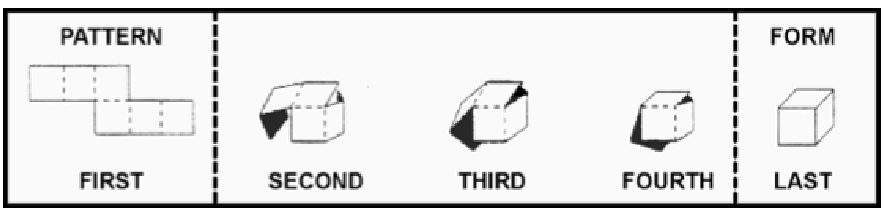
SPATIAL ABILITY (SA)

Directions

This is a test of your ability to recognize a form from its pattern, or a pattern from its form. Mark your answer on the answer sheet by blacking out the letter that corresponds to your choice.

Now look at the example questions below. There are two types of problems. In one, you are to find the form that can be made by folding the cardboard pattern and fitting it together. In the other, you are to find what the cardboard pattern would look like if the form were unfolded.

Look at the first row of pictures below:



The first row of pictures shows what is meant by folding a PATTERN to make a FORM. The broken lines in the first picture, labeled PATTERN, show where the folds are to be made. The second, third, and fourth pictures show the steps in folding the PATTERN into the FORM. The last picture in the row, labeled FORM, shows the completed form.