Naval officer selection in Canada:

An evaluation of the Maritime Officer Selection Test (MOST)

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied Psychology (Industrial/Organizational)

Sebastien J-R. Blanc

Saint Mary's University

Approved:

V. M. Catano, Ph.D.

Approved

A. L. Day, Ph.D.

Approved:

K. A. Hill, Ph.D.

October 2003

Your file Votre référence
ISBN: 0-612-85659-3
Our file Notre référence
ISBN: 0-612-85659-3

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I would like to thank the Canadian Forces for the opportunity to pursue graduate studies. Special thanks to the Director of Maritime Personnel who sponsored this research project and to the Director of Human Resources Research and Evaluation for access to their data bank.

Thanks to Dr. Vic Catano who was the best Thesis Advisor I could have wished for. His extensive knowledge was a source of inspiration and his availability was most appreciated. Thanks also to my Thesis Advisory Committee members (i.e., Dr. Arla Day, Dr. Ken Hill, and Dr. Shaun Newsome) for their review of, and comments on, this research project.

Finally, I would like to thank my wife, Amélie, for her unconditional love and absolute confidence in me. Her endless patience and unending support helped make the completion of my graduate work possible. Thank you.

Naval officer selection in Canada:
An evaluation of the Maritime Officer Selection Test (MOST)

Sebastien J-R. Blanc
Submitted October 10. 2003

Abstract

Two studies evaluating the Maritime Officer Selection Test (MOST) were conducted. Study 1 ($N = 744$) examined the psychometric properties of the individual items and their factorial structure. Study 2 ($N = 224$) assessed convergent validity and sought to determine whether using the MOST for the selection of future naval officers might discriminate against women, thereby precluding them from pursuing a career within the Canadian Navy. The results of the first study indicate that most examinees do not have enough time to complete the test, which invalidates previous evaluations of its internal consistency. Additionally, the two studies suggest that the MOST does *not* measure what it was intended to assess (i.e., memory, selective attention, and decision-making), but that it does nonetheless evaluate other ability constructs required for naval officer training performance (i.e., spatial scanning and general reasoning). With regards to adverse effect, the absence of differential item functioning and similarities in success rates across genders indicate that using the MOST does not preclude women from pursuing a career within the Canadian Navy.

Naval officer selection in Canada:

An evaluation of the Maritime Officer Selection Test (MOST)

Introduction

The design of useful and legally defensible selection procedures is generally conducted in multiple stages. In the first stage, it is typically required to conduct a job analysis to collect information about the occupation in question, and identify the knowledge, skills, abilities, and other attributes (KSAOs) required to perform critical job activities (Harvey, 1991; Whetzel & Wheaton, 1997). Next, the job analytic results are used to identify relevant predictors and criteria. When no predictors or criteria are found, the third stage consists of developing new ones using job analytic information as a framework of reference. The *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial and Organizational Psychology, 1987) provide useful guidelines in this regard.

In the final stage of the process, the prospective users (i.e., the organization) must collect multiple evidence to show that inferences made from the selection measures are both valid and reliable (Canadian Psychological Association, 1996). Where technically feasible, it is also recommended (and often necessary) to determine if the new selection procedures have an adverse effect on protected groups[1]. Adverse effect occurs when "... an employer, in good faith, adopts a policy or practice that has an unintended, negative impact on members of a protected group" (Catano, Cronshaw, Wiesner, Hacket, & Methot, 2001, p. 43). These negative effects can take several forms including adverse

[1] In Canada, protected groups include women, aboriginal peoples, persons with disabilities and members of visible minorities (Employment Equity Act, 1995 c. 44).

impact (Uniform Guidelines on Employee Selection Procedures, 1978), differential prediction (Cleary, 1968), and differential item functioning (Zumbo, 1999).

Although the Canadian Forces (CF) has made substantial efforts to satisfy most of these design requirements in the development and empirical validation of its naval officer selection procedures, a few important concerns associated with the validation of the Maritime Officer Selection Test (MOST) have been overlooked (e.g., there has been no research on its construct-related validity). The main purpose of this thesis is to address these concerns by conducting the necessary validity research.

Naval Officer Selection in Canada: An Overview

The process by which the CF selects its naval officers, specifically Maritime Surface and Sub-surface (MARS) Officers, is conducted in three phases. First, eligible applicants are screened at a Recruiting Centre where an aptitude test (i.e., the Canadian Forces Aptitude Test) is generally followed by a conditional offer of enrolment, a medical examination, a physical fitness test, and a selection interview with a military career counsellor. The main purpose of the selection interview is to assess military potential (MP). MP ratings are valid predictors ($r = .26, p < .05$) of basic officer training (BOTC) performance (Okros, Johnston, & Rodgers, 1988).

Upon completion of the initial screening phase, the personnel files of suitable applicants are forwarded to the CF Recruiting Group Headquarters where an evaluation is made as to which candidates are the most likely to succeed during the Naval Officer Assessment Board (NOAB)[II]. The NOAB is the final phase of the screening process. It incorporates two components: an assessment centre, which is "the Navy's look at the

---

[II] According to King (1989), the predictive validity of this evaluation is enhanced when the prediction is based on Military Potential (MP) ratings and aptitude test scores ($R = .58, p < .01$).

individual", and a two-day orientation program, which is "the individual's look at the Navy" (Scholtz, 2002, p. 2). Together, these two components are the direct result of more than 25 years of personnel applied research.

Background

The process by which the CF selects its naval officers has been the focus of several research projects (e.g., Bradley, Wiesner & Latham, 1993; Catano, 1989; Okros & King, 1989). One of these projects was the development and empirical validation of the NOAB, which is based upon assessment centre methodology (King, 1989). The main purpose of this section is to provide a succinct overview of its development, which will also set the stage for the subsequent evaluation of the MOST, previously known as the Passage Planning Test (Okros, 1988).

Naval Officer Interview Board (NOIB). In 1976 the Maritime Officer Production Study (MOPS) identified unusually high levels of attrition during naval officer training. To address this issue, the MOPS recommended that a Maritime Command interview board be established to screen all applicants for the MARS occupation (Okros et al., 1988). Following this recommendation, Maritime Command (MARCOM) established the Naval Officer Interview Board (NOIB), which consisted of a selection interview and an orientation program (Okros et al., 1988). The orientation program included tours of naval facilities and briefings by naval officers. Its main purpose was to provide a realistic job preview (Bradley, 1990). Despite the efforts of senior leadership, the NOIB never met its objectives, and by 1984, the CF Personnel Applied Research Unit was mandated to identify strategies to address these issues (Boswell, 1993).

Upon review of the MARS Officer production system, Rodgers (1984) confirmed that the attrition among junior naval officers was higher than for any other officer occupation (i.e., 25-35% attrition during basic officer training and 40-50% during subsequent classification training). These statistics confirmed the need to: (1) identify the abilities required for MARS Officer selection; (2) substantially revise the orientation component of the NOIB; and (3) develop MARS-specific selection tests based on job analytic results (Okros et al., 1988).

Ability requirements for MARS Officer selection. The ability requirements for MARS Officer selection were identified using an adaptation of the ability analysis procedure developed by Levine, Mallamad, and Fleishman (1978). In the first stage of the process, Rodgers (1986) used job analytic information to identify the tasks that junior MARS Officers might be called upon to perform when they are first assigned to an operation ship. The list of tasks was incorporated into a Training Importance Questionnaire, which was administered to a group of senior MARS Officers. These naval officers were asked to rate each task in terms of its training importance. These ratings were used to identify a list of the most critical tasks, which were later assessed using binary-decision flow diagrams (Mallamad, Levine, & Fleishman, 1980). The use of decision flow diagrams was expected to reduce the level of information processing and decision making on the part of the raters, which, in turn, would help them identify the ability constructs that should be considered in the development of MARS-specific selection tests. Table 1 presents the ability requirements for MARS Officer selection along with their operational definition.

Table 1

Abilities Requirements For MARS Officer Selection

Ability	Definition
Memorization	Memorization is "...the ability to remember information, such as words, numbers, pictures, and procedures" (Fleishman & Reilly, 1992, p. 15).
Oral Comprehension	This is "...the ability to understand spoken English words and sentences" (Fleishman & Reilly, 1992, p. 7).
Oral Expression	Oral Expression is "...the ability to use English words or sentences in speaking so others can understand" (Fleishman & Reilly, 1992, p. 10).
Problem Sensitivity	This is "...the ability to know when something is wrong or is likely to go wrong" (Fleishman & Reilly, 1992, p. 16).
Deductive Reasoning	Deductive Reasoning is "...the ability to apply general rules to specific problems and to come up with logical answers; for example, deciding whether or not an answer to a non-mathematical problem makes sense, or solving syllogistic reasoning problems" (Fleishman & Reilly, 1992, p. 21).
Spatial Orientation	This is "...the ability to know one's location in relation to the environment one is in or to know where an object is in relation to oneself" (Fleishman & Reilly, 1992, p. 31).

Table 1 continued

Ability	Definition
Selective Attention	Selective Attention is "...the ability to concentrate on a task over a period of time, [and] . . . without being distracted by external stimuli" (Fleishman & Reilly, 1992, p. 36).
Time Sharing	This is "...the ability to shift back and forth efficiently between two or more activities or sources of information" (Fleishman & Reilly, 1992, p. 37).
Written Comprehension	Written Comprehension involves "...reading and understanding the meaning of words, phrases, sentences, and paragraphs" (Fleishman & Reilly, 1992, p.8).
Number Facility	This is "...the ability to add, subtract, multiply, divide, and manipulate numbers quickly and accurately, [but it] ...does not involve understanding or organizing mathematical problems (Fleishman & Reilly, 1992, p.19).
Choice Reaction Time	This ability is now referred to as Response Orientation, and is defined as "...the ability to choose between two or more movements quickly and correctly when two or more different signals (lights, sounds, pictures) are given" (Fleishman & Reilly, 1992, p. 40).
Visualization	This is "...the ability to imagine how something will look when it is moved around or when its parts are moved or rearranged" (Fleishman & Reilly, 1992, p. 33)

Table 1 continued

Ability	Definition
Written Expression	According to Fleishman and Reilly (1992), it is "...the ability to use English words or sentences in writing so others can understand". (p. 11)
Originality	This is "...the ability to produce unusual or clever ideas about a given topic or situation" (Fleishman & Reilly, 1992, p.14).
Perceptual Speed	This involves "... the ability to compare letters, numbers, objects, pictures, or patterns, quickly and accurately" (Fleishman & Reilly, 1992, p. 25)
Information Ordering	According to Fleishman & Reilly (1992), it is "...the ability to correctly follow a rule or a set of rules specifying how to arrange things or actions in a certain order". (p. 25)
Arm-Hand Steadiness	It is "...the ability to keep the hand and arm steady [when using small objects]" (Fleishman & Reilly, 1992, p. 44).
Flexibility of Closure	This is the ability "...to identify or detect a known pattern (e.g., a figure, word, or object) that is hidden in other material" (Fleishman & Reilly, 1992, p. 30).

Naval Officer Selection Board (NOSB). Upon review of the ability requirements for MARS Officer selection, the Naval Officer Selection Board (NOSB) was established. The NOSB retained the orientation component of the NOIB, but the selection component was improved by incorporating other assessment instruments (i.e., conducting officer assessment, selection interview, file review, leadership task, two leaderless discussions, and in-basket exercise). Evaluation of this new selection process revealed that it was a valid predictor (see Table 2) of basic officer training performance (Okros et al., 1988). Catano (1989) arrived at a very similar conclusion when he showed that NOSB selectees would be, on average, 4.6% more productive than those selected without any board review, and that using the NOSB would increase success at BOTC from 70% to 78%, which translates into a net benefit of over $200,000.

Naval Officer Assessment Board (NOAB). In 1988-89, the Maritime Officer Selection Test (MOST) was added to the selection component of the NOSB, which was renamed the Naval Officer Assessment Board (NOAB). It was expected that the MOST would improve the predictive validity of the NOSB by measuring MARS-specific abilities (i.e., memory, selective attention, and problem-solving) that were not directly assessed by existing measures (Okros, 1988). However, the incremental validity of the MOST was never tested.

In its present form, the assessment stage of the NOAB incorporates five assessment measures including a file review, a board interview, an aptitude test, a conducting officer assessment[III], and a written assignment (Scholtz, 2002). Recent efforts to validate this new process have failed due to a number of contributing factors: (1) a change in the CF aptitude test (from the GC2 to the CFAT) that reduced the number of files available for analysis, (2) large numbers of missing data, (3) problems with criterion data, and (4) discrepancies between the MOST scores as recorded on NOAB files and those held in the CF Selection Test database (C. Mombourquette, personal communication, 2002).

[III] The conducting officer assessment is based on observing each candidate's behaviours (i.e., self-discipline, maturity, team-orientation, and motivation) in less formal settings than the board interview.

Table 2

Correlations Between Assessment Centre Merit Scores and MARS Officer Training Performance

Authors	*N*	MARS Officer Training		
		BOTC	Phase III	Phase IV
Bradley (1990)	118	_	.20*	.20*
Okros et al. (1988)	273	.34*	_	_

Note. Dashes (-) indicate that the relationship was not examined. BOTC = Basic Officer Training Course.

* $p < .05$.

Maritime Officer Selection Test (MOST)

Test development. The MOST is a *complex* cognitive-perceptual test, which purports to measure abilities required for MARS Officer training (i.e., memory, selective attention, and decision-making). It was originally modeled on the US Army Flight Planning Test (McAnulty, Cross, & Jones, 1986), which was part of an experimental battery of aviation-related ability tests. The Flight Planning Test was obtained by the CF and modified by Okros (1988) to reflect a maritime context. In its present form, the MOST contains five timed sections presented in three levels of difficulty. Each section consists of a grid route map and 12 questions about the best route between two locations on the map. Naval officer candidates must correctly answer 24 questions[IV] or more (out of 60) by memorizing and applying progressively more complex sets of navigational rules. Selective attention is assessed by including irrelevant information in about one-third of the questions (e.g., What compass headings are required in traveling from K12 to N10 *at a speed of 7.5 knots*?).

[IV] Scholtz (2003) has set this cut-off score using a modified version of Angoff's (1971) method for setting standards.

Psychometric properties. Okros (1988) has shown that the MOST is a valid predictor of MARS Phase III performance (Table 3), and that its use in selection would possibly increase the predictive validity of the NOSB. Bradley (1990) found similar results showing that the MOST is the single best predictor of both MARS Phase III and MARS Phase IV performance[V] (Table 3). More recently, Stouffer (1996) examined the internal consistency of the MOST and found a Cronbach's alpha of .85. He also looked at mean score differences between male and female Officer Cadets, and found no significant differences. With regard to item-difficulty, Okros (1988) found p-values ranging from .30 to .70 (M = .53, SD = .13) while Stouffer (1996) found p-values ranging from .26 to .87 (M = .59, SD = .15).

Table 3

Correlations Between the MOST and MARS Officer Training Performance

Authors	N	MARS Officer Training		
		BOTC	Phase III	Phase IV
Bradley (1990)	122/72	--	.21*	.30*
Okros et al. (1988)	64	--	.32*	--

Note. Dashes (-) indicate that the relationship was not examined. BOTC = Basic Officer Training Course.

* $p < .05$.

Notwithstanding the above results, several concerns remain to be addressed. First, it is unclear at this point if the MOST is a speeded or a timed-power test[VI]. If the MOST were a speeded test such that most examinees could not attempt all items (Crocker & Algina, 1986), it would be incorrect to measure its reliability in terms of internal

[V] MARS Phase III consists of learning all aspects of navigation, ship driving, and bridgemanship (e.g., receiving information from multiple sources and giving out orders). Phase IV is an iteration of Phase III, but at a much higher level.

[VI] Providing the average p-values for each section was not sufficient for making this assessment.

consistency (Nunnally & Bernstein, 1994). The reliability of a speeded test should be measured based on performance from two independent testing sessions using anyone of the following methods: (1) test-retest reliability, (2) equivalent form reliability, or (3) split-half reliability from two, separately timed half tests (Cohen, Swerdlick, & Smith, 1992).

The second issue concerns the validation process. Although performance on the MOST is related to concurrent standing during MARS Officer training (Bradley, 1990; Okros, 1988), no studies have investigated its factorial structure or its convergent and discriminant validity. Therefore, without convincing evidence of its construct-related validity, it would be erroneous to claim that anyone failing to meet the cutoff is unsuitable for a naval career.

A final concern deals with the "fairness" of using the MOST for the selection of future MARS Officers. Although the MOST was not designed to preclude members of protected groups from joining the Navy, MARCOM is concerned that using the MOST might adversely affect women (Major C. Evans, personal communication, February 2002). Several studies have found that men tend to perform better than women in spatial and mathematical tasks (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997); and if the MOST were unintentionally measuring these constructs, its use in selection would likely contravene the Employment Equity Act of 1995[VII].

The present thesis addresses each of the above concerns in order. In the first

---

[VII] The purpose of this Act is to achieve equality in the workplace so that no person shall be denied employment opportunities or benefits for reasons unrelated to ability and, in the fulfillment of that goal, to correct the conditions of disadvantage in employment experienced by women, aboriginal peoples, persons with disabilities and members of visible minorities by giving effect to the principle that employment equity means more than treating persons in the same way but also requires special measures and the accommodation of differences (Employment Equity Act, 1995 c. 44).

study, a large data set is used to inspect the psychometric properties of individual items on the MOST and to examine the factorial structure of the test. In the second study, further analyses (i.e., confirmatory factor analysis and hierarchical regression analyses) are performed to further assess the construct-related validity of the MOST. The issue of adverse effect discrimination (i.e., adverse impact and measurement bias) is addressed by comparing the performance of men with that of women and by looking for the presence of differential item functioning.

Cognitive Abilities and Gender

The topic of gender differences in cognitive abilities has been studied for more than a century (Hyde, 1990). In the early years, scientists believed that brain size was related to intelligence; and because women had smaller brains than men, they were thought to be less intelligent (Caplan & Caplan, 1999). This widely held belief did not last very long, and by 1910, the brain-size argument was dismissed (Hyde, 1990). The dismissal of this argument was due, in part, to the advent of the mental testing movement pioneered by the psychologists Alfred Binet and Lewis Terman (Hyde, 1990). These two scientists believed that there were no gender differences in general intelligence, and constructed their tests to reflect this conviction (Hyde, 1990). The next advancement in this line of research was the development of the Primary Metal Ability test (Thurstone, 1938), which laid the foundation for research on gender differences in verbal ability, mathematic ability, and spatial ability (Hyde, 1990).

In 1974, Maccoby and Jacklin (1974) published the first comprehensive literature review on gender differences in cognitive abilities. They concluded that women have greater verbal ability, that men have better visual-spatial ability, and that men perform better on mathematical ability tests. What was then considered as the definitive statement on gender differences in cognitive abilities was later tempered by Hyde (1981) who re-analyzed Maccoby and Jacklin's (1974) data and published the first meta-analysis of gender differences in cognitive abilities.

Meta-analysis can be thought of as a multi-stage procedure to combine the quantitative results of numerous studies (Hyde & Linn, 1986). In the end, it produces an effect size (*d*), which represents how far apart the means of men and women are in standard deviation units (Hyde, 1990). By convention, an effect size (*d*) of 0.10 or less is trivial, 0.20 is small, 0.50 is medium, and 0.80 is large (Hyde, 1994; MacIntyre, 1997). Using the aforementioned conventions, Hyde (1981) concluded that gender differences in verbal ability were small ($d = -0.24$) and that gender differences in spatial ($d = 0.45$) and mathematical ability ($d = 0.43$) were moderate.

A few years later, Linn & Peterson (1985) performed a more sophisticated meta-analysis of gender differences in spatial abilities, and concluded that there are three distinct types of spatial ability, each showing a different pattern of gender differences: mental rotation ($d = 0.73$), spatial perception ($d = 0.44$), and spatial visualization ($d = 0.13$). The above findings were later re-assessed (Voyer, Voyer, & Bryden, 1995), and when spatial ability was partitioned into its three constituents, the effect size for mental rotation ($d = 0.56$) was again higher than for spatial perception ($d = 0.44$) and spatial visualization ($d = 0.19$). When that partitioning method was applied to the study of gender differences in verbal abilities, Hyde and Linn (1988) found a slight female superiority in performance (average $d = -0.11$), with the exception of Speech Production where men were superior ($d = 0.33$). With regard to mathematical abilities, Hyde, Fennema, and Lamon (1990) found a slight female superiority (i.e., average $d = -0.05$), with the exception of complex-problem solving where men tended to be stronger ($d = 0.32$). Gallagher, De Lisi, Holst, McGillicuddy-De Lisi, Morely, and Cahalan (2000) found similar results and reported evidence suggesting that men are more flexible than women in applying solution strategies for solving mathematical word-problems.

In sum, the cognitive abilities of men and women are much more homogenous than what was initially expected (Caplan & Caplan, 1999), and where discrepancies remain (e.g., spatial-visualization), the gap is closing (Feingold, 1988; Stumps and Klieme, 1989). Considering the above information, some researchers question whether these differences have any practical significance (MacIntyre, 1997). One research area where gender differences are practically meaningful is personnel selection. On occasion, gender differences can result in adverse impact and/or cause measurement biases.

Adverse Impact and Measurement Bias

Adverse impact. Using cognitive ability tests in personnel selection can have an adverse impact on women (Hough, Oswald, & Ployhart, 2001; Salgado, Viswesvaran, & Ones, 2001). Adverse impact is present when the selection rate for a protected group is lower (i.e., less than four-fifths) than that for the relevant comparison group (Uniform Guidelines on Employee Selection Procedures, 1978). When those situations occur, an employer has the obligation to provide reasonable accommodation (i.e., use suitable alternative selection procedures), unless it is shown that using a litigious procedure was a *bona fide* occupational requirement (Catano et al., 2001). Table 4 provides an example of the four-fifths rule in determining adverse impact on women.

Table 4

Example of the Four-Fifth Rule Based on Achieving a Score of 24 on the MOST

Group	Total Applicant Pool (*n*)	Number of Successful Examinees	Success Rate
Women	20	5	.25
Men	100	40	.40

Note. Because .25 / .40 < .80, using 24 as a cut-off score would have had an adverse impact on women.

Measurement Bias. Measurement bias occurs when a factor inherent within a test prevents an accurate and impartial assessment of the ability being measured (Cohen et al. 1992). These inherent factors can affect the relationship between test scores and criterion, but they can also affect inter-item relationships (Zumbo, 1999). In the former situation, the presence of measurement bias is most easily detected by developing a separate regression equation for the focal and reference groups and testing the difference between their slopes and intercepts (Cleary, 1968). When a significant difference is detected, and this difference is detrimental to the focal group, then the selection procedure is biased (Norborg, 1984; Ree, Carretta, & Steindl, 2001). In the latter situation, the detection of measurement bias is best achieved by comparing the item characteristic curves (ICCs) of the reference and focal groups. ICCs plot the probability that an item will be answered correctly against ability (see Figure 1). The shape of the ICC reflects the influence of three factors (Drasgow & Hulin, 1990; Zumbo, 1999): the intercept, which represents the likelihood of finding the correct response just by guessing; the slope, which indicates how well an item discriminates among levels of ability; and the threshold, which depicts the level of item difficulty.

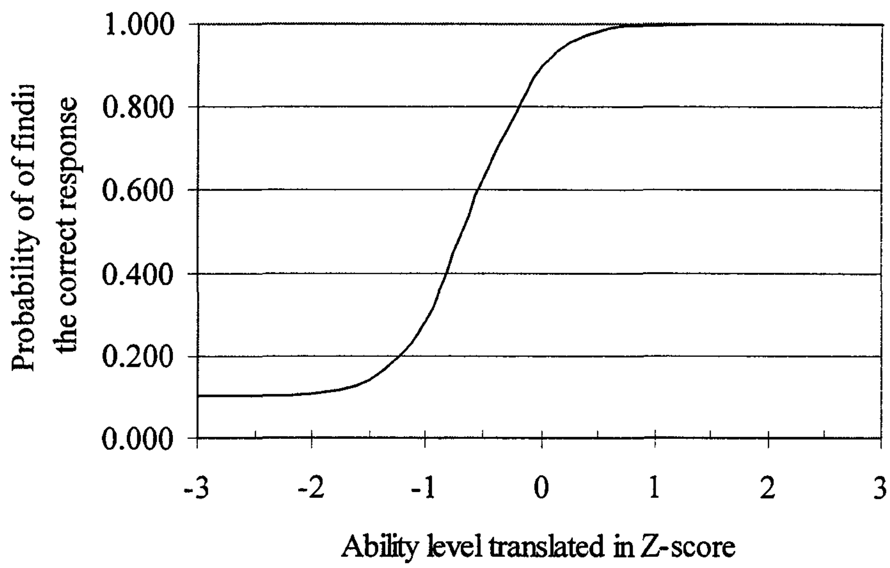

Figure 1. Example of an item characteristic curve.

When the ICCs of the focal and reference groups are significantly different from one another, the item is said to show differential item functioning (DIF[VIII]; Nunnally & Bernstein, 1994). DIF is caused by a nuisance source of variation affecting the item under consideration, placing the focal group at a disadvantage (Swanson, Clauser, Case, Nungester, & Featherman, 2002; Whitmore & Schumacker, 2001; Zumbo, 1999). There are two categories of DIF: uniform DIF and non-uniform DIF. When DIF is uniform, there is no interaction between ability level and group membership (Swaminathan & Rogers, 1990). As a result, the ICCs of the focal and reference groups are parallel. Conversely, when DIF is non-uniform the interaction between ability level and group membership causes ICCs to intersect (Swaminathan & Rogers, 1990).

One of the most powerful and flexible methods for detecting DIF is through the use of logistic regression (Clauser, Nungester, Mazor, & Ripkey, 1996; Mazor, Kanjee, &

[VIII] The accepted definition of DIF is that an item shows DIF if examinees of the same ability but belonging to different groups do not have the same probability of success on an item (Mazor, Kanjee, & Clauser, 1995; Swaminathan & Rogers, 1990).

Clauser, 1995; Swaminathan & Rodgers, 1990). This procedure is based on the statistical modeling of the probability of finding the correct response to an item by group membership and a conditioning variable (usually the scale or subscale score; Zumbo, 1999). The logistic regression equation for DIF detection can be written as:

$$\ln\left[\frac{p_i}{(1-p_i)}\right] = b_0 + b_1 TOT + b_2 GROUP + b_3 (TOT * GROUP),$$

where p_i is the proportion of individuals that endorse an item in the direction of the latent variable, TOT is the total test score, and GROUP is the grouping variable (dummy coded 0 = women and 1 = men).

The main advantage of using the logistic regression method is its power to detect both uniform and non-uniform DIF. Simulation studies have shown that logistic regression is less sensitive to sample size than IRT-based procedures (Drasgow & Hulin, 1990; Zumbo, 1999) and is more powerful than most other DIF detection methods (i.e., Mantel-Haenszel, ANOVA, and SIB procedures) in "flagging" non-uniform DIF (Jodoin & Gierl, 2001; Swaminathan & Rogers, 1990; Whitmore & Schumacker, 1999). One drawback of using logistic regression is the inflated risk of making Type I errors (Jodoin & Gierl, 2001). However, purifying the conditioning variable by removing multidimensional items can alleviate the risk of making Type I errors (Holland & Thayer, 1988; Mazor, Hambleton, & Clauser, 1998; Navas-Ara & Gomez-Benito, 2002; Zumbo, 1999). Another method for reducing the risk of making Type I errors is to use a conservative measure of effect size ($R^2\Delta$) in conjunction with a significant 2-*df* chi-squared test. Although Zumbo and Thomas (1996) suggest using $R^2\Delta \geq .13$ as a minimum standard, Jodoin and Gierl (2001) have obtained adequate results using $R^2\Delta \geq .035$.

Study 1: Item Analysis and Factorial Structure

The initial objective of this study is to identify poorly functioning items, and determine if the MOST is a speeded or timed-power test. Its second objective is to assess the factorial structure of the MOST to see if the MOST measures what it purports to measure. Because the MOST was designed to measure memory, selective attention, and decision-making (Okros, 1988), it is expected that a factor analysis of the MOST will yield a three-factor solution corresponding to the three constructs that it purports to measure.

Method

Sample

The sample was composed of 744 English-speaking officer candidates who were recruited by the Canadian Forces between 1987 and 2002. Information about their age, gender, and level of education, was not available.

Measure

Maritime Officer Selection Test (MOST). The MOST is a complex-cognitive perceptual test designed to assess abilities required to plan a simulated ship passage (i.e., memory, selective attention, and decision making; Okros, 1988). It contains five timed sections presented in three levels of difficulty. Each level begins with a set of directions followed by a series of route selection rules (e.g., If two or more routes have the same length, the "best" route has the fewest turns). For the first two levels, the route selection rules are followed by a set of practice problems (e.g., Which landmark(s) would you pass in traveling from A1 to E4?). The practice problems are also timed. After receiving feedback on correct responses to the practice problems, examinees are given time to read and memorize the route selection rules pertaining to the level and section they are at. Each section consists of a grid route map (Figure 2) and 12 four-response multiple-choice questions about the best route between two locations on the map (e.g., What speed is required in traveling from B2 to D7?).

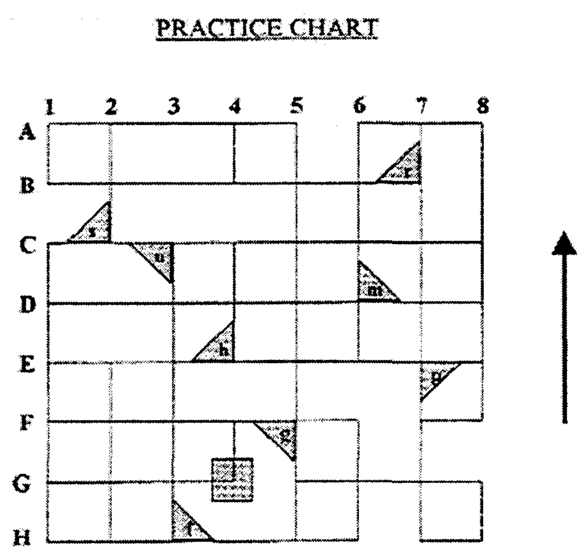

Figure 2. Example of a grid route map. *Notes.* From *Maritime Officer Selection Test,* by A. C. Okros, 1988, Ottawa, ON, Canada: Director Maritime Personnel. Copyright 2000 by Her Majesty the Queen in Right of Canada. Reprinted with permission.

Level 1 contains three types of items: (1) items pertaining to the landmark(s) passed in traveling between two locations, (2) items pertaining to the number of turns needed to get to the destination, and (3) items pertaining to the compass heading(s) required in sailing between two grid coordinates. Level II items add two variables (i.e., distance and speed), and formulae are provided for determining speed (represented in nautical miles per hour or knots) and time required in traveling between two locations. Speed is presented either numerically or symbolically using a speed indicator. Both the knot values on the indicator and the formulae must be memorized, as they are not provided in the text. Level II incorporates two types of items: (1) items dealing with the time required in traveling between two locations at a given speed, and (2) items dealing with the speed required in sailing between two grid coordinates in a given time. Level III adds additional route selection restrictions based on tide levels. Tide level information is provided by either words or tidal graphs. Again, examinees must memorize the ranges on the tidal graphs, as they are not labelled during the test.

Procedures

Analyses are based on test scores for 744 officer candidates who attended the NOAB between 1987 and 2002. These test scores were obtained from a database maintained by the Director of Human Resources Research and Evaluation, Department of National Defence.

Data Analysis

Assumptions. The dataset was screened for violations of critical assumptions (i.e., linearity, absence of outliers among cases, absence of multicolinearity and singularity, and factorability of *R*; Tabachnick & Fidell, 2001), but no violations were found.

Descriptive statistics and correlations. Descriptive statistics were used to examine the general pattern of scores across sections. Pearson product-moment correlations were used to determine the size of intersection and section-total relationships.

Item analysis. The primary purpose of the item analysis was to examine the psychometric properties of individual test items such that poor performing items could be identified. This objective was achieved by looking at multiple indices including item-difficulty, item-discrimination, and corrected item-total correlation.

The item-difficulty index (*p*) represents the proportion of examinees who answered an item correctly (Cohen, Swerdlick, & Smith, 1992). That index can range from .00 to 1.00. A large *p*-value indicates that an item was easy while a small *p*-value indicates that an item was difficult. Item true score variance is maximized when the *p*-value of a 4-choice item lies between .62 and .74 (Crocker & Algina, 1986).

Discrimination indices such as the item-discrimination index and corrected item-total correlation indicate how well an item discriminates between high- and low-scoring examinees (Cohen et al., 1992). The item-discrimination index (*d*) measures the difference between the *p*-values of high- and low-scoring examinees while the corrected item-total correlation represents the strength of the relationship between each item and the total test score. Both indices can range from .00 to 1.00; and by convention, values below .20 suggest that an item should be completely revised or discarded (Crocker & Algina, 1986).

Factor analysis. A principal axis factoring analysis was used to identify the underlying constructs that caused the test items to form coherent clusters. This factor extraction technique conforms to the factor analytic model and has the advantage of being widely used and understood (Tabachnick & Fidell, 2001).

Results

Descriptive Statistics

As shown in Table 5, the pattern of scores across levels was very similar. The mean scores gradually increased from section 1 to 4, but total test scores decreased after that point. The inter-section correlations were all significant, but often moderate in size (Cohen, date). The highest inter-section correlations were between section 2 and 3 (r = .50, $p < .01$) and between section 3 and 4 (r = .51, $p < .01$). Section-total correlations were generally much stronger and homogenous.

Table 5

Descriptive Statistics for the MOST and Intersection Correlations

Level	Section	*M*	*SD*	Time (Min)	1	2	3	4	5	Total
I	1	6.17	1.99	6	-	.29	.21	.31	.34	.59
II	2	6.23	2.19	8		-	.50	.43	.41	.74
II	3	6.55	2.02	8			-	.51	.44	.74
II	4	7.28	2.36	8				-	.46	.77
III	5	5.54	2.07	10					-	.73
-	Total	31.77	7.63	-						-

Note. N = 744. All correlations are significant at the .01 level (two-tailed).

Item Analysis

As shown in Table 6, 32 of 60 items were omitted by at least 10% of the examinees. As a result of this, the p-value of omitted items was low and their discrimination indices were spuriously elevated. The remaining items had a much lower rate of omission, but their levels of discrimination were often lower than .20. This observation applied to all items but two (i.e., items 14 and 38). With regards to item-difficulty, 43 items had a p-value situated outside of the optimum range (i.e., $.62 \leq p \leq .74$), which means that only one item (i.e., item 14) functioned effectively across all indices.

In addition to the above, there were multiple problems with item alternatives. As shown below (see Table 6), 37 of 60 items had at least one distractor that was so obviously incorrect that it was selected by less than 5% of the examinees. There were also five items (items 7, 23, 31, 42, and 55) that had an abnormally high percentage of examinees choosing the same incorrect answer. These five items should be reviewed to insure the accuracy of their keyed response.

Table 6

Item Analysis Results for the 60 Items on the MOST

Item	Item Responses (%)				Omit	Diff.		Item-total
	A	B	C	D	(%)	*p*	d^a	correlation
1	7.3	0.7	1.1	91.0*	0.0	.91	.03	.02
2	2.7	9.1	42.7*	44.4*	1.1	.44	.17	.06
3	1.9	6.7	17.6	73.5*	0.3	.74	.05	.01
4	1.9	2.2	16.5	78.9*	0.5	.79	.15	.07
5	4.0	35.9	7.4	48.8*	3.9	.49	.23	.14
6	1.9	7.8	9.9	76.3*	4.0	.76	.21	.16
7	4.7	24.7	33.2	25.1*	12.2	.25	.11	.04
8	1.2	12.8	3.1	67.3*	15.6	.67	.47	.36
9	27.2	4.2	7.4	33.3*	28.0	.33	.37	.31
10	5.2	2.6	21.1	33.5*	37.6	.34	.30	.22
11	3.4	7.1	18.5	23.7*	47.3	.24	.31	.27
12	5.0	10.8	7.7	21.0*	55.6	.21	.54	.30

Table 6 continued

Item	Item Responses (%)				Omit	Diff.		Item-total
	A	B	C	D	(%)	*p*	d^a	correlation
13	1.2	0.9	9.5	87.8*	0.5	.88	.10	.13
14	10.2	5.8	9.0	73.0*	2.0	.73	.26	.21
15	15.6	6.6	5.2	72.2*	0.4	.72	.06	.03
16	2.4	7.9	13.3	70.7*	5.6	.71	.21	.15
17	4.3	3.8	21.9	69.5*	0.5	.70	.14	.06
18	18.8	10.5	12.9	46.4*	11.4	.46	.32	.27
19	1.7	9.4	32.1	42.9*	13.8	.43	.30	.22
20	5.0	18.5	11.0	34.5*	30.9	.35	.50	.41
21	2.4	6.7	7.7	51.6*	31.6	.52	.58	.43
22	8.5	10.2	5.9	31.9*	43.5	.32	.51	.41
23	6.3	15.2	12.1	12.5*	53.9	.13	.25	.27
24	3.8	8.9	3.4	30.4*	53.6	.30	.40	.32

Table 6 continued

Item	Item Responses (%)				Omit	Diff.		Item-total
	A	B	C	D	(%)	*p*	d^a	correlation
25	.40	2.0	12.0	83.2*	2.4	.83	.16	.16
26	3.9	13.6	8.7	70.4*	3.4	.70	.23	.19
27	17.5	0.4	4.8	77.3*	0	.77	.16	.09
28	1.3	6.5	3.9	84.9*	3.4	.85	.19	.18
29	31.7	3.1	1.6	62.0*	1.6	.62	.08	.02
30	15.9	10.3	23.1	37.6*	13.0	.38	.21	.13
31	59.7	4.6	1.2	27.8*	6.7	.28	.09	.06
32	4.4	3.2	18.5	60.5*	13.3	.61	.53	.39
33	6.5	15.6	9.9	37.6*	30.4	.38	.54	.41
34	4.0	8.9	10.8	47.8*	28.5	.48	.51	.39
35	6.3	10.9	14.4	22.6*	45.8	.23	.27	.26
36	2.0	3.9	9.5	43.0*	41.5	.43	.38	.31

Table 6 continued

Item	Item Responses (%)				Omit (%)	p	d^a	Item-total correlation
	A	B	C	D				
37	7.3	8.7	13.3	68.8*	1.9	.69	.22	.12
38	10.6	14.0	17.7	52.3*	5.4	.52	.30	.21
39	5.6	9.7	14.0	66.4*	4.3	.66	.27	.19
40	1.5	1.3	12.0	84.9*	0.3	.85	.08	.05
41	3.4	2.6	2.4	89.9*	1.7	.90	.14	.19
42	6.9	12.9	43.4	30.8*	6.0	.31	.12	.08
43	9.4	8.3	30.9	40.7*	10.6	.41	.38	.28
44	5.5	2.6	6.2	72.0*	13.7	.72	.49	.42
45	4.8	4.7	15.1	50.3*	25.1	.50	.67	.47
46	1.2	9.0	4.0	65.5*	20.3	.66	.59	.45
47	2.0	4.4	2.7	63.4*	27.4	.63	.61	.46
48	9.1	13.3	5.5	43.0*	29.0	.43	.44	.33

Table 6 continued

Item	Item Responses (%)				Omit	Diff.		Item-total
	A	B	C	D	(%)	*p*	*d*[a]	correlation
49	0.0	32.8	6.7	60.5*	0.0	.61	.21	.15
50	31.6	5.2	20.4	41.4*	1.3	.41	.02	-.05
51	22.6	8.2	7.5	55.2*	6.5	.55	.26	.18
52	2.7	8.6	24.2	61.8*	2.7	.62	.28	.16
53	30.2	1.3	5.0	62.0*	1.5	.62	.24	.18
54	1.9	11.8	18.5	57.0*	10.8	.57	.42	.18
55	10.6	8.6	41.5	25.2*	10.1	.25	.13	.09
56	0.7	4.3	7.9	73.3*	13.8	.73	.37	.32
57	11.0	21.0	15.2	23.8*	29.0	.24	.28	.23
58	18.0	4.8	6.5	36.8*	33.9	.37	.52	.38
59	9.1	2.6	10.1	31.9*	46.3	.32	.55	.44
60	4.8	6.0	19.8	21.6*	47.7	.22	.23	.20

Note. $N = 244$. * The position of the keyed responses was changed to protect the integrity of the test. [a]The index of discrimination (d) $= p_u - p_l$, where p_u is the proportion in the upper group who answered the item correctly and p_l is the proportion in the lower group who answered the item correctly (Crocker & Algina, 1986). The groups are composed of the top 30 percent and the bottom 30 percent of the examinee group.

Factor Analysis

A factor analysis of the Maritime officer Selection Test (MOST) was expected to yield a three-factor solution corresponding to the three constructs that it purports to measure (i.e., selective attention, decision-making, and memory). This hypothesis was tested by means of principal axis factoring with equamax rotation. This extraction technique is robust to violations of multivariate normality (Fabrigar et al., 1999), and the equamax rotation was the technique yielding the best simple structure. Because 22 factors had eigenvalues ≥ 1, the number of factors was determined by a scree test (Tabachnick & Fidell, 2001). Using this criterion, three factors emerged, which accounted for 13.69% of the variance. The results of this analysis are presented in Table 7.

Factor interpretation. Using a cut of .32 for inclusion of an item in the interpretation of a factor (Crocker & Algina, 1986), 29 of 60 items loaded on at least one factor. The first factor accounted for 5.96% of the variance, and was comprised of items that were omitted by a large number of examinees. The second factor accounted for 4.38% of the variance, and seamed to measure the spatial scanning[IX] construct described by Ekstrom, French, Harman, and Dermen (1976). The third factor seamed to measure general reasoning[X], and accounted for a meager 3.34% of the variance. Failure of numerous variables to load on a factor demonstrates the heterogeneity of the test items (Tabachnick & Fidell, 2001) and the small size of the communalities suggests that the test items were poorly defined by this factor solution.

[IX] Spatial scanning was defined as "speed in exploring visually a wide or complicated spatial field" (Ekstrom et al., 1976, p. 155).
[X] General reasoning was defined as "the ability to select and organize relevant information for the solution of a problem" (Ekstrom et al., 1976, p. 133).

Table 7

Rotated Factor Loadings and Communalities ($N = 744$)

Item	Communalities	Factor Loadings		
		1	2	3
1	.01	-.08	.01	.03
2	.01	-.01	.02	.11
3	.00	-.03	-.00	.05
4	.00	-.00	.02	.05
5	.04	.16	-.00	.11
6	.04	.18	-.04	.08
7	.01	-.01	.07	.05
8	.16	.29	.20	.18
9	.11	.25	.13	.16
10	.11	.31	.08	.00
11	.14	**.36**	.09	.01
12	.21	**.44**	-.02	.12

Table 7 continued

Items	Communalities	Factor Loadings		
		1	2	3
13	.02	.08	.07	.07
14	.09	.04	.12	.27
15	.03	.01	-.10	.13
16	.02	.05	.11	.08
17	.02	-.04	-.00	.12
18	.16	.15	.04	**.36**
19	.07	.14	.23	.03
20	.26	**.32**	.22	**.33**
21	.35	**.56**	.14	.12
22	.29	**.50**	.13	.17
23	.12	**.32**	.10	.10.
24	.34	**.57**	.04	-.08

Table 7 continued

Items	Communalities	Factor Loadings		
		1	2	3
25	.05	.09	.03	.21
26	.18	.00	.05	**.43**
27	.01	.05	.04	.04
28	.10	-.00	.19	.25
29	.02	-.06	-.02	.14
30	.07	.05	.00	.26
31	.04	-.09	.19	.00
32	.25	.22	**.45**	.07
33	.25	.29	.30	.28
34	.26	**.35**	**.37**	.00
35	.11	.25	.22	.01
36	.33	**.50**	.20	-.19

Table 7 continued

Items	Communalities	Factor Loadings		
		1	2	3
37	.09	-.02	.02	.29
38	.14	.16	-.00	**.34**
39	.04	.04	.18	.10
40	.00	.00	.02	.02
41	.10	-.04	.19	.24
42	.05	-.11	.18	.08
43	.16	.07	.20	**.34**
44	.33	.15	**.52**	.18
45	.32	.29	**.40**	.27
46	.45	**.39**	**.54**	-.11
47	.49	**.42**	**.54**	-.13
48	.26	.30	**.40**	-.10

Table 7 continued

Items	Communalities	Factor Loadings		
		1	2	3
49	.04	.03	.09	.17
50	.01	-.06	.01	-.05
51	.15	-.02	.07	**.38**
52	.06	.00	.10	.23
53	.08	.08	-.02	.27
54	.16	.09	.20	**.34**
55	.02	-.02	.13	.02
56	.26	.13	**.49**	.05
57	.08	.26	.12	.03
58	.25	**.41**	.30	-.03
59	.28	**.47**	.19	.13
60	.14	**.37**	.05	-.08

Note. $N = 744$. Factor loadings ≥ .32 are in bold (Tabachnick & Fidell, 2001).

Discussion

The initial objective of this study was to identify poorly functioning items, and determine if the MOST is a speeded or timed-power test. The second objective was to examine its factorial structure to see whether it measures what it purports to measure (i.e., memory, selective attention, and decision making). As discussed below, the results of this research depart substantially from what Okros (1988) and Stouffer (1996) have reported.

Descriptive Statistics and Intersection Correlations

The patterns of scores across sections and intersection correlations were consistent with past research (Okros, 1988). However, there was no evidence to suggest that Level II items are any more difficult than Level I items. Mean scores increased across sections (Table 5), which demonstrates that Level II items were in fact easier than Level I items. Only Level III items were more difficult than Level I and II items.

Item Analysis

The item analysis results have revealed several problems with item alternatives (e.g., 37 of 60 items had at least one distractor that was so obviously incorrect that it was selected by less than 5% of examinees). These problems should be addressed promptly. However, it may be premature to discard items or make substantial revisions to the test on the basis of these analyses. The high rate of omission (due to test-takers running out of time to complete all items on each scale) has had a dramatic impact on the psychometric properties of individual test items (i.e., the *p*-values of omitted items were low and their discrimination indices were spuriously elevated). Should it be necessary to make substantial revisions to the MOST, it is recommended to extend the time limits and to discard dysfunctional items on the basis of a new item analysis. Ideally, 90 percent of examinees should have enough time to complete the test (Nunnally & Bernstein, 1994).

Factor Analysis

Thirdly, the factor analytic results suggest that the MOST is *not* measuring what it purports to measure. Eighty six percent of its variance was accounted for by "noise"; and of the three factors that came out, only two were relevant to the MARS occupation (i.e., spatial scanning and general reasoning). In fact, the "spatial scanning" factor may be considered a second-order visualization factor (Ekstrom, French, & Harman, 1979) and the "general reasoning" factor may be interpreted as an index of number facility.

One possible explanation for the "messy" factor analytic results is that patterns of correlations among items on the MOST were highly affected by the time limits and the ordering of items within the test. As Nunnally and Bernstein (1994, p. 351) phrased it, "items [on a speeded test tend to] correlate more highly with items near their own ordinal position on the test than they do with items further removed in the ordering because of the similarities of their distribution". Considering the above information, the best method to learn more about the factorial composition of the MOST is to reassess its factorial structure after its time limits have been extended. However, this procedure might conflict with the concurrent assessment of its other psychometric properties, and as an alternative, it might be more convenient to look at its convergent-related validity.

Reliability

Finally, all analyses converge in suggesting that the MOST has more in common with speeded tests than timed-power tests. As a result, previous findings attesting to its reliability are no longer applicable (Ree et al., 2002; Stouffer, 1996) and this limitation prevents the calculation of its standard error of measurement[XI]. Future research will have to demonstrate that test scores are reliable if the MOST is to be used again for making personnel decisions (Canadian Psychological Association, 1996). Because there is only one version of the MOST, there are only two viable options for re-assessing its reliability: (a) test-retest reliability or (b) split-half reliability from two, separately timed half tests (Cohen et al., 1992). The former method would provide information on the amount of measurement error associated with changes in examinees (e.g., fatigue, nervousness, physical discomfort, practice effects, etc.) while the latter would provide information on error variance due to content sampling (Crocker & Algina, 1986).

[XI]The standard error of measurement (SEM) may be defined as "the standard deviation of scores that a person would receive if we could obtain an infinite number of independent test scores from this individual" (Saks, 2000, p. 208). The SEM is used to establish cut-off scores on selection tests (Scholtz, 2003) and its computation is needed to determine whether a person's test score is significantly above or below the minimum passing mark (Saks, 2000).

Study 2: Construct Validity and Fairness-Related Issues

The first objective of this study is to examine the convergent and discriminant validity of the MOST using relevant measures from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). Considering the factor analytic results of the first study and the requirement to memorize progressively more complex sets of navigational rules prior to writing each section of the test), it is expected that correlations between the MOST and measures of relevant constructs (i.e., spatial scanning, general reasoning, and memory) will exceed its correlations with other variables (e.g. demographic variables).

The second objective of this research is to re-examine the factorial structure of the MOST using confirmatory factor analyses. It is expected that a three-factor model derived from the factor analytic results of the first study will provide a better fit to the data than two models derived a content analysis. The third objective is to determine if using the MOST in selection is having an adverse impact on women. Because the MOST appears to be measuring constructs that tend to favour men (i.e., spatial scanning and general reasoning), it is expected that: (1) the performance of men on the MOST will significantly exceed that of women, and (2) that using a cut-off score of 24 on the MOST will have an adverse impact on women.

The final objective of this research is to determine whether men and women of equal ability have the same probability of finding the correct response to each item on the MOST. This research objective was set in collaboration with MARCOM whose short-term objective is to remove all barriers precluding women from joining the Navy.

Method

Sample

Focal group. The focal group was composed of 120 women from Saint Mary's University. Most of them were White (90%), and 86.7% indicated that English was their primary language. The age of the focal group ranged from 18 to 50 (M = 21.71, SD = 5.46), and 96.7% of its constituents were undergraduate students. A breakdown of participants by academic background is presented in Table 8.

Reference group. The reference group was composed of 104 men from Saint Mary's University. The majority was White (86.5%), and 92.3% reported that English was their primary language. The age of the reference group ranged from 17 to 35 (M = 21.10, SD = 3.25), and 95.2% were undergraduate students. A breakdown of participants by academic background is presented in Table 8.

Table 8

Breakdown of Participants by Academic Background

Faculty	Focal Group - Women (n = 120)		Reference Group - Men (n= 104)	
	Frequency	%	Frequency	%
Arts	61	50.8	48	46.1
Science	19	15.8	16	15.4
Commerce	23	19.2	26	25.0
Graduate Studies	4	3.3	5	4.8
Undeclared / Missing	13	10.9	9	8.7
Total	120	100	104	100

Note. N = 224

Measures

Background Information Questionnaire. The background information questionnaire was composed of 10 items designed to measure relevant demographic variables (e.g., gender, age, primary language, and race). A copy of this questionnaire is available in Appendix A.

Mathematics Aptitude Test. The Mathematics Aptitude Test (RG-2) is part of the Kit of Factor-Referenced Cognitive Tests, which was developed to provide scientists with a means of identifying certain aptitude factors in factor-analytic studies (Ekstrom et al., 1979). This test measures general reasoning, which is "the ability to select and organize relevant information for the solution of a problem" (Ekstrom et al., 1976, p. 133). It requires 20 minutes to administer, and is composed of 30 items assessing one's ability to solve arithmetic and algebraic word problems. Its published reliability is .81 (Ekstrom et al., 1976).

Auditory Number Span Test. The Auditory Number Span Test (MS-1) is part of the Kit of Factor-Referenced Cognitive Test (Ekstrom et al., 1976). It measures memory span, which is "the ability to recall a number of distinct elements for immediate recall" (Ekstrom et al., 1979, p. 19). This test takes about 10 minutes to administer, and consists of 24 items assessing one's ability to remember progressively more complex series of numbers called out by an examiner. Its published reliability is .74 (Ekstrom et al., 1976).

Map Planning Test. The Map Planning Test (SS-3) is also part of the Kit of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). It measures spatial scanning, which is defined as "speed in exploring visually a wide or complicated spatial field" (Ekstrom et al., 1976, p. 155). This test contains 40 items presented in two parts. Each part takes 3 minutes to administer and contains 20 items assessing one's ability to find rapidly the shortest route between two points on a grid route map. Each grid route map depicts a set of lettered grid coordinates as well as several numbered buildings and roadblocks. The subjects must find the shortest route between two lettered points without encountering any roadblocks. The published reliability of this test is .79 (Ekstrom, et al., 1976).

Procedure

Participants were tested in a classroom setting using the standard directions presented in Appendix B. At the beginning of each testing session, participants were asked to read and sign an informed consent form (Appendix C). Those who agreed to participate were told that they were taking part in a research project designed to validate an aptitude test for the selection of naval officers. Because the MOST is classified, participants were asked not to discuss the nature of the aptitude tests with anyone else. Next, they were informed that everyone would get $10 for participation and perhaps more (i.e., an additional $50) for displaying the abilities required for naval officer training (i.e., for placing in the top 20% of the examinee group).

Upon completion of the Background Information Questionnaire, participants were instructed to read a two-page description of the MARS Officer occupation (Appendix D). Next, they were asked to complete a series of cognitive ability tests. The presentation order of the ability tests varied across testing sessions. In one testing condition, the MOST was followed by the three marker tests from Kit of Factor-Referenced Cognitive Tests. In the other testing condition, the three marker tests were followed by the MOST. Regardless of the testing condition, the three marker tests were always administered in the same order, that is, the general reasoning test was always followed by the memory and spatial scanning tests. Each testing session lasted approximately 2 hr and 30 min.

Data Analysis

Data screening process. Study 2 involved 228 participants. One participant was removed from subsequent analyses because the individual failed to complete the second part of the experiment. Upon removal of that case, the sample was split into two groups (i.e., Focal and Reference groups), and each variable was screened for out-of-range values and missing values. None were found, but mean scores on the Mathematics Aptitude Test (RG-2) were lower than what was previously reported in the literature (Ekstrom et al., 1976). The few univariate outliers were recoded into one unit larger/smaller than the next most extreme score in the distribution. Pairwise linearity was assessed, and one variable (i.e., level of education) failed to meet the linearity assumption. Because a square-root transformation made very little difference, that variable was left unchanged. Next, each group was screened for multivariate outliers. None were found within the Reference group, but three were found and removed from the Focal group. These multivariate outliers were caused by large differences in test-taking motivation between the MOST and the Kit of Factor-Referenced Cognitive tests.

Assumptions. The compliance with or violation of relevant assumptions was assessed in accordance with the procedures outlined in Tabachnick and Fidell (2001). For all analyses, the data met the most critical assumptions (i.e., multivariate normality, absence of outliers, homogeneity of variance-covariance matrices, linearity, homogeneity of regression, absence of multicolinearity and singularity, homoscedasticity of residuals, and independence of error; Tabachnick & Fidell, 2001).

Descriptive statistic and intercorrelations. Descriptive statistics were used to examine the pattern of scores across gender. Correlations were performed to examine the relationships between study variables and identify sources of unwanted effects (i.e., covariates). The Cronbach's alpha was calculated for the MS-1 test, and coefficients of equivalence were calculated for the RG-2 and SS-3 tests. However, for the reasons discussed in the first study, the reliability of the MOST could not be assessed.

Convergent and discriminant validity. Correlations were used to examine the convergent and discriminant validity of the MOST. The joint effects of spatial scanning, general reasoning, and memory were assessed by means of hierarchical regression. As recommended by Tabachnick and Fidell (2001), the effect of covariates was statistically removed from the analysis before the MOST was regressed onto the variables of primary interest (RG-2, MS-1, and SS-3).

Confirmatory factor analysis (CFA). The hypothesized factorial structure was assessed through *LISREL 8.53* (Jöreskog & Sörbom, 2002). The goal of the analysis was to determine if the hypothesized factorial structure would fit the data. A factorial structure fits when the covariance matrix implied by the hypothesized factor loadings is consistent with the observed covariance matrix. The adequacy of the fit is typically assessed using absolute fit indices (e.g., chi-squared [χ^2] test statistic and RMSEA), comparative fit indices (e.g., NFI and CFI), and parsimonious fit indices (e.g., PNFI and PGFI). A factorial structure fits when its χ^2 test statistic is not significant; and by convention, when its RMSEA-value is smaller than .10 and its other fit indices (i.e., NFI, CFI, PNFI, and PGFI) are greater than .90 (Kelloway, 1998; Tabachnick & Fidell, 2001).

Mean score differences and adverse impact. A multivariate analysis of covariance (MANCOVA) was used to determine the presence of reliable mean differences among groups after adjusting a linear combination of all dependent variables for differences on two covariates (Tabachnick & Fidell, 2001). The "four-fifth" rule was used to determine whether using the MOST in selection has an adverse impact on women.

Differential item functioning (DIF) analysis. The DIF analyses were performed using Zumbo' s (1999) syntax for SPSS (Appendix E). For an item to be classified as displaying DIF, two conditions have to be met: the *p*-value for the two-degrees of freedom (*df*) χ^2 test in logistic regression has to be $\leq .01$ (Zumbo, 1999) and the Zumbo-Thomas effect size measure (R^2) has to be $\geq .13$ (Zumbo & Thomas, 1997).

Results

Descriptive Statistics and Intercorrelations

Table 9 and 10 present the means and standard deviations for all study variables. As shown in Table 9, the men performed better than women on all cognitive ability tests. The largest difference was on the RG-2 test where the effect size (*d*) was equal to 0.48 standard deviation units[XII]. The effect sizes for the other tests were as follows: MOST (0.31), SS-3 (0.21), and MS-1 (0.16).

In addition to presenting the means and standard deviations for all study variables, Table 9 and 10 present the correlations between all study variables. An inspection of these correlational patterns revealed the presence of two covariates. These covariates

---

[XII] $d = (M_m - M_f) / \surd\, [(SD^2_m + SD^2_f) / 2]$

(i.e., Presentation Order and Science)[XIII] correlate with the variables of primary interest and their effect should be accounted for when assessing whether mean score differences were larger than expected by chance.

Table 9

Group Means, Standard Deviations, and Intercorrelations

Variable	Reference Group (Women)		Focal Group (Men)		1.	2.	3.	4.
	M	SD	*M*	*SD*				
1. **MOST**	29.59	7.98	27.30	6.92	-	.50**	.48**	.26**
2. **RG-2**	10.29	5.31	7.86	4.76	.40**	-	.34**	.36**
3. SS-3	25.89	7.87	24.07	9.46	.33**	.29**	-	.19*
4. MS-1	8.03	3.17	7.51	3.19	.29**	.30**	.20*	-

Note. RG-2 = General Reasoning test, SS-3 = Spatial Scanning test, and MS-1 = Auditory Number Span test. Variables in bold indicate the presence of reliable mean score differences between the Reference and Focal groups (Table 15). Correlations for the Reference group ($n = 104$) are above the diagonal and those for the Focal group ($n = 120$) are below the diagonal. Dashes indicate that the value was 1.00.

* $p < .05$. ** $p < .01$.

[XIII] Presentation Order (PO) variable was dummy coded 1 when the MOST was presented first and 0 when it was presented last. Science is a "scientific background" indicator, which was derived from adding-up the number of science-related courses that each participant had completed since graduation from High School.

Table 10

Pooled Means, Standard Deviations, and Intercorrelations (N = 224)

Variable	*M*	*SD*	1.	2.	3.	4.	5.	6.	7.	8.
1. PO	-	-	1.00	.03	-.12	-.09	-.18**	-.19**	-.13**	-.22**
2. Gender	-	-		1.00	.07	-.11	-.15*	-.24**	-.10	-.08
3. Age	21.42	4.57			1.00	.18**	-.07	.14*	-.11	.04
4. Science	2.74	3.32				1.00	.20**	.24**	.15*	.05
5. MOST	28.36	7.50					-	.47**	.40**	.28**
6. RG-2	8.99	5.16						(.67)	.32**	.34**
7. SS-3	24.92	8.79							(.74)	.20**
8. MS-1	7.75	3.18								(.74)

Note. PO = Presentation Order, RG-2 = General Reasoning test, SS-3 = Spatial Scanning test, and MS-1 = Auditory Number Span test. Dashes indicate that a value was not estimated. Values in parentheses are reliability coefficients. Gender was dummy coded 1 for male and 2 for female.

* $p < .05$. ** $p < .01$.

Convergent Validity

As expected (see Table 9 and 10), the correlations between the MOST and the three variables of primary interest (i.e., RG-1, SS-3, and MS-1) were much higher than its correlations with any other variables ($r_{56} = .47$, $r_{57} = .40$, and $r_{58} = .28$; *ps* < .01). The unique effect of these three variables was further assessed by means of hierarchical regression where the MOST was used as the dependant variable. Following the recommendations of Tabachnick and Fidell (2001), the control variables were entered in the regression equation before the variables of primary interest. Table 11 presents the results of a hierarchical regression analysis in which the cognitive ability tests were the last variables entered. Together, the variables of primary interest accounted for 23% of the variance in aptitude test performance, but the MS-1 test did not contribute significantly to the regression ($\beta = .10$, ns.)

Table 11

Summary of Hierarchical Regression Analysis for Variables Correlating with the MOST

Variable	β	R	R^2	ΔR^2	$F\Delta$	Sig. $F\Delta$
Step 1		.29	.08	.08	6.85	.00
PO	-.16**					
Gender	-.13*					
Science	.18**					
Step 2		.56	.31	.23	24.10	.00
RG-2	.32**					
SS-3	.26**					
MS-1	.10					

Note. $N = 224$. PO = Presentation Order, RG-2 = General Reasoning test, SS-3 = Spatial Scanning test, and MS-1 = Auditory Number Span test.

* $p < .05$. ** $p < .01$.

Confirmatory Factor Analysis (CFA)

Hypothesized model. The hypothesized model, *Model 1*, was tested based on its covariance matrix using maximum likelihood estimation as implemented in *LISREL 8.53* (Jöreskog & Sörbom, 2002). This model was comprised of three oblique factors. Factor 1 included largely omitted items from Study 1 that loaded on the Speediness factor (i.e., items 11, 12, 21, 22, 24, 36, and 58-60). Factor 2 included all items from Study 1 that loaded on the Spatial Scanning factor as well as all items beginning with either "How many turns are required...", "Which landmarks would you pass...", or "What compass headings are required..." (i.e., items 1-10, 13, 15-17, 19, 23, 25, 27-29, 31, 32, 34, 39-42, 44-50, 52, 53, 55, and 56). The third factor, General Reasoning, was comprised of all remaining items (i.e., items 14, 18, 20, 26, 30, 33, 35, 37, 38, 43, 51, 54, and 57).

Alternative models. Two alternative models were examined. *Model 2* was composed of three oblique factors. Factor 1 included items beginning with "How many minutes are required...", or "What speed is required..." (i.e., items 14, 18, 20, 23, 26, 30, 33, 35, 37, 38, 43, 45, 54, 57, and 59) . Factor 2 was comprised of items beginning with either "How many turns are required...", "Which landmarks would you pass...", or "What compass headings are required..." (i.e., items 1-13, 15, 17, 21, 25, 27, 31, 36, 40, 41, 46, 48-50, 53, and 56). Factor 3 was composed of items designed to measure selective attention (i.e., items 16, 19, 22, 24, 28, 29, 32, 34, 39, 42, 44, 47, 51, 52, 55, 58, and 59). *Model 3* was composed of two oblique factors. Factor 1 was composed of all items beginning with "How many minutes are required...", or "What speed is required..." (i.e., 14, 18, 20, 23, 26, 30, 33, 35, 37, 38, 43, 45, 51, 54, 57, and 59), whereas Factor 2 was composed of all remaining items beginning with either "How many turns are required...", "Which landmarks would you pass...", or "What compass headings are required...". Items designed to measure selective attention were imbedded within factor 1 or 2, using the wording of each item as a criterion for inclusion in a factor. For instance, items measuring selective attention and beginning with "How many minutes are required..." were included in factor 2 while items beginning with "How many minutes are required..." were imbedded in factor 1.

Assessment of fit. With the exception of the RMSEA index, all fit indices are outside the bounds that indicate a good fit to the data (GFI, NFI, CFI, PNFI, PGFI < .90). However, they all converge in suggesting the marginal superiority of the hypothesized model (Table 12). Comparison with the other models shows that Model 1 (χ^2_{1707} = 2260.58, $p < .000$) provides a better fit to the data than does Model 2 (χ^2_{1707} = 2299.54, $p < .000$) and Model 3 (χ^2_{1709} = 2276.54, $p < .000$). Inspection of the parsimonious fit indices (i.e., PNFI and PGFI) suggests that a two-factor model is no better than a three-factor model. The comparative fit indices (i.e., NFI and CFI) suggest that the hypothesized model is better fitting than the null model (Kelloway, 1998).

Table 12

Goodness-of-Fit Indices for Model Comparisons

Model	χ^2	*df*	GFI	RMSEA	NFI	CFI	PNFI	PGFI
1. 3-factor oblique	2260.58	1707	.76	.03	.46	.77	.44	.71
2. 3-factor oblique	2299.24	1707	.75	.04	.45	.75	.43	.70
3. 2-factor oblique	2268.53	1709	.75	.04	.46	.77	.44	.70

Note. N = 224. GFI = Goodness of Fit Index, RMSEA = Root Mean Square Error of Approximation, NFI = Normed Fix Index, CFI = Comparative Fit Index, PNFI = Parsimony Normed Fit Index, and PGFI = Parsimony Goodness of Fit Index.

Table 13 presents the standardized parameter estimates for the hypothesized model. Although most parameters were significant, they typically explained trivial amounts of item variance (R^2 range from .00 to .42). As shown in Table 14, the three factors were significantly correlated (r_{12} = .73, r_{13} = .20, and r_{23} = .37).

Table 13

Standardized Parameter Estimates for the Hypothesized Model

Item	Speediness	Spatial Scanning	General Reasoning	R^2
1		.23		.05
2		.10		.01
3		.16		.02
4		.18		.03
5		.07		.00
6		.26		.07
7		.07		.01
8		.39		.15
9		.09		.01
10		.31		.10
11	.23			.05
12	.36			.13

Table 13 continued

Item	Speediness	Spatial Scanning	General Reasoning	R^2
13		.22		.05
14			.32	.10
15		-.08		.01
16		.24		.06
17		.27		.07
18			.48	.23
19		.30		.09
20			.29	.09
21	.64			.42
22	.46			.21
23		.24		.06
24	.51			.26

Table 13 continued

Item	Speediness	Spatial Scanning	General Reasoning	R^2
25		.18		.03
26			.49	.24
27		.05		.00
28		.37		.14
29		.00		.00
30			.29	.09
31		.03		.00
32		.38		.14
33			.28	.08
34		.35		.12
35			.21	.04
36	.51			.26

Table 13 continued

Item	Speediness	Spatial Scanning	General Reasoning	R^2
37			.26	.07
38			.34	.12
39		.33		.11
40		.22		.05
41		.35		.12
42		.26		.07
43			.42	.17
44		.51		.26
45		.28		.08
46		.47		.22
47		.54		.29
48		.28		.08

Table 13 continued

Item	Speediness	Spatial Scanning	General Reasoning	R^2
49		.13		.02
50		.15		.02
51			.49	.24
52		.14		.02
53		.34		.12
54			.50	.25
55		.16		.02
56		.44		.20
57			.10	.01
58	.39			.15
59	.21			.04
60	.28			.08

Note. $N = 224$. Standardized parameter estimates >.14 are significant at the .01 level.

Table 14

Interfactor Correlations

	1.	2.	3.
1. Speediness	1.00		
2. Spatial Scanning	.73**	1.00	
3. General Reasoning	.20*	.37**	1.00

Note. $N = 224$. * $p < .05$. ** $p < .01$.

Mean Score Differences and Adverse Impact

Mean score differences. A multivariate analysis of covariance (MANCOVA) examined whether mean score differences between the Reference and Focal groups (Table 15) were larger than expected by chance after adjusting for the effect of two covariates (i.e., PO and Science). Results of evaluation of assumptions were satisfactory (Tabachnick & Fidell, 2001). The two covariates provided significant adjustment to five of nine dependant variables (i.e., MOST, RG-2, SS-3, and MS-1) with β values ranging from -.16 to .23 ($p < .05$).

The results of the MANCOVA (Wilks' Lambda = .95, $p < .05$) indicated the presence of a significant multivariate effect and univariate F-ratios indicated the presence of significant group differences on two measures (i.e., on the MOST and RG-2 test). As shown in Table 15, Reference group members performed better on the MOST (adjusted Mean score = 29.40, SE = .71) than members of the Focal group (adjusted mean score = 27.46, SE = .66). Reference group members did also better on the RG-2 test (adjusted Mean score = 10.15, SE = .48) than their counterparts (adjusted mean score = 7.99, SE = .44. There were no significant group differences on any other assessment measures.

Table 15

Univariate Analysis of Variance for the Reference and Focal Groups

Source	SS	*df*	MS	*F*	η^2
MOST	208.38	1	208.38	3.99*	.02
RG-2	256.41	1	256.41	10.94**	.05
SS-3	129.90	1	129.90	1.74	.01
MS-1	12.08	1	12.08	1.24	.01

Note. $N = 224$. PO = Presentation Order, RG-2 = General Reasoning test, SS-3 = Spatial Scanning test, and MS-1 = Auditory Number Span test.

$^{*}p < .05$. $^{**}p < .01$

Adverse Impact. The results of the adverse impact analysis are presented in Table 16. In order to demonstrate that using the MOST in selection is not having an adverse impact on women, the success rate of the Focal group had to be equal to or greater than 80% of the success rate of the Reference group. As shown below, this condition was met; which means that using the MOST in selection is not having an adverse impact on women.

Table 16

Adverse Impact Analysis Based on Achieving a Score of 24 on the MOST

Group	*n*	Number of Successful Examinees	Success Rate
Focal (Women)	120	82	.68
Reference (Men)	104	75	.72

Note. Because .68 / .72 = .94, using 24 as a passing mark has no adverse impact on women.

Differential Item Functioning (DIF) Analysis

A DIF analysis was performed for each of the 60 MOST items. For an item to be classified as displaying DIF, the *p*-value for the two-*df* χ^2 tests in logistic regression had to be ≤ 0.01 (Zumbo, 1999). In addition, the Zumbo-Thomas R^2 had to be $\geq .13$ (Zumbo & Thomas, 1997). Using the above guidelines, none of the MOST items were classified as displaying DIF (see Table 17).

Table 17

Summary of DIF Analysis for all MOST Items

Item	R^2 at Each Step in Sequential Hierarchical Regression			χ^2 2-df / (*p*)	Zumbo-Thomas R^2	DIF?
	R^2_1	R^2_2	R^2_3			
1	.16	.16	.20	2.27 (.32)	.04	No
2	.07	.12	.13	4.62 (.10)	.01	No
3	.07	.07	.16	6.11 (.05)	.09	No
4	.14	.15	.15	0.63 (.73)	.00	No
5	.01	.01	.06	3.37 (.19)	.05	No
6	.20	.20	.20	0.53 (.77)	.00	No
7	.07	.06	.06	1.31 (.52)	.00	No
8	.27	.29	.34	6.01 (.05)	.05	No
9	.08	.14	.16	6.04 (.05)	.02	No
10	.26	.30	.32	3.65 (.16)	.02	No
11	.11	.13	.24	10.47 (.01)	.11	No
12	.24	.24	.24	0.36 (.84)	.00	No

Table 17 continued

Item	R^2 at Each Step in Sequential Hierarchical Regression			χ^2 2-df / (p)	Zumbo-Thomas R^2	DIF?
	R^2_1	R^2_2	R^2_3			
13	.15	.20	.21	4.59 (.10)	.01	No
14	.23	.23	.24	.78 (.68)	.01	No
15	.00	.01	.02	1.22 (.54)	.01	No
16	.20	.20	.22	1.33 (.51)	.02	No
17	.21	.26	.26	4.22 (.12)	.00	No
18	.23	.25	.25	1.42 (.49)	.00	No
19	.07	.12	.13	4.62 (.10)	.01	No
20	.13	.16	.17	3.85 (.15)	.01	No
21	.39	.43	.43	3.65 (.16)	.00	No
22	.20	.31	.31	9.69 (.01)	.00	No
23	.19	.19	.21	2.13 (.35)	.02	No
24	.25	.25	.27	2.54 (.28)	.02	No

Table 17 continued

Item	R^2 at Each Step in Sequential Hierarchical Regression			χ^2 2-df / (p)	Zumbo-Thomas R^2	DIF?
	R^2_1	R^2_2	R^2_3			
25	.11	.11	.12	0.76 (.68)	.01	No
26	.33	.33	.33	0.00 (.99)	.00	No
27	.13	.13	.13	0.02 (.99)	.00	No
28	.31	.33	.36	3.08 (.21)	.03	No
29	.03	.10	.12	5.59 (.06)	.02	No
30	.12	.20	.23	6.11 (.05)	.03	No
31	.04	.04	.04	0.11 (.95)	.00	No
32	.28	.32	.36	7.35 (.03)	.04	No
33	.27	.30	.33	5.56 (.06)	.03	No
34	.28	.28	.28	0.19 (.91)	.00	No
35	.18	.19	.20	2.00 (.37)	.01	No
36	.30	.30	.32	1.80 (.41)	.02	No

Table 17 continued

Item	R^2 at Each Step in Sequential Hierarchical Regression			χ^2 2-df / (p)	Zumbo-Thomas R^2	DIF?
	R^2_1	R^2_2	R^2_3			
37	.18	.19	.20	1.52 (.47)	.01	No
38	.07	.07	.08	0.66 (.72)	.01	No
39	.25	.28	.29	3.06 (.22)	.01	No
40	.24	.24	.25	0.69 (.71)	.01	No
41	.18	.20	.20	1.79 (.41)	.00	No
42	.31	.32	.32	0.91 (.63)	.00	No
43	.21	.28	.30	6.38 (.04)	.02	No
44	.13	.13	.15	4.34 (.11)	.02	No
45	.32	.33	.34	1.38 (.50)	.01	No
46	.33	.34	.35	1.94 (.38)	.01	No
47	.41	.41	.47	5.46 (.07)	.06	No
48	.25	.25	.25	0.11 (.95)	.00	No

Table 17 continued

Item	R^2 at Each Step in Sequential Hierarchical Regression			χ^2 2-df / (p)	Zumbo-Thomas R^2	DIF?
	R^2_1	R^2_2	R^2_3			
49	.12	.13	.15	2.50 (.29)	.02	No
50	.13	.18	.18	4.20 (.12)	.00	No
51	.16	.17	.20	2.66 (.27)	.03	No
52	.06	.06	.06	0.11 (.95)	.00	No
53	.26	.26	.27	0.85 (.65)	.01	No
54	.34	.35	.36	1.44 (.49)	.01	No
55	.12	.12	.12	0.23 (.89)	.00	No
56	.30	.32	.32	1.41 (.49)	.00	No
57	.02	.03	.03	1.13 (.57)	.00	No
58	.37	.39	.39	1.73 (.42)	.00	No
59	.21	.23	.23	1.66 (.44)	.00	No
60	.10	.19	.19	7.34 (.03)	.00	No

Note. $N = 224$

Discussion

The first objective of this study was to assess the convergent and discriminant validity of the MOST using marker tests from the Kit of Factor-Referenced Cognitive tests (Ekstrom et al., 1976). The next objective was to re-examine its factorial structure using maximum likelihood estimation as implemented in *LISREL 8.53* (Jöreskog & Sörbom, 2002). The third objective was to determine if using the MOST in selection has an adverse impact on women and the final goal was to examine whether men and women of equal ability have the same probability of finding the correct response to each item on the MOST.

Convergent and Discriminant Validity

As expected, correlations between the MOST and measures of general reasoning, spatial scanning, and memory, were much larger than its correlations with other variables. However, when the MOST was regressed onto measures of general reasoning, spatial scanning, and memory, the effect of memory did not contribute to the variation in aptitude test performance. General reasoning and spatial scanning were the only two significant variables; and together, these two variables explained only 23% of the variation in aptitude test performance. One possible explanation for this result is that most examinees had no difficulty to memorize the progressively more complex sets of navigational rules. Another possibility is that the variance accounted for by memory may already have been predicted by the other two factors (i.e., general reasoning and spatial scanning). A third possibility is that passed a certain point, having a good memory does not improve your score on the MOST. Future studies should investigate these possibilities and examine whether general intelligence (*g*) accounts for more variance in aptitude test performance than lower-order factors (i.e., spatial scanning and general reasoning). Past research has shown that *g* typically accounts for about 30% to 65% of variance in aptitude test performance and that specific abilities provide little incremental validity beyond *g* (Ree, Carretta, and Steindl, 2002; Salgado et al., 2002).

Confirmatory Factor Analysis (CFA)

As sown previously (see Table 12), none of the assessed models have provided a good fit to the data. Although Model 1 was slightly better fitting than other models, the difference between their corresponding fit indices was marginal. These consistent lacks of fit reaffirmed the relevance of a previous statement, which suggested that most of the variability in aptitude test performance is accounted for by "noise" and random error of measurement. Future studies should determine the amount of measurement error that may be attributed to unwanted sources of variation and confirm whether general intelligence accounts any portion of the unexplained variance.

Adverse Impact & Measurement Bias

Consistent with past research (e.g., Hyde et al., 1990; Linn & Peterson, 1985; Voyer et al., 1995) the men performed better than women on all cognitive ability tests administered in this study. However, these differences were small ($d \leq .48$), and although using the MOST tends to favour men ($d = 0.31$), there was no evidence that using this test in selection has an adverse impact on women. When the passing mark was set to 24 (out of 60), the success rate of men (.72) and women (.68) were nearly equal.

With regards to measurement bias, there was no evidence of DIF. The necessary next step is to reassess the criterion-related validity of the MOST and look for evidence of *differential prediction* (Canadian Psychological Association, 1996). This other type of measurement bias is found when the criterion score predicted from the use of a common regression line (relating criterion performance to the score on the predictor measure) is either to high or too low for different groups of examinees (Cleary, 1968; Maxwell & Arvey, 1993; Young & Kobrin, 2001).

General Discussion

The main objective of this last section is to summarize the findings of this thesis, integrate the results of its two studies with those of earlier research, discuss possible limitations to the findings, identify further research requirements, and make recommendations to improve the selection of future MARS Officers. These five topics are discussed under the following headings: (a) summary of research findings, (b) limitations, (c) implications for future research, and (d) recommendations.

Summary of Research Findings

Item analysis. The item analysis performed in the first study shows that most examinees do *not* have enough time to answer all items on the MOST. This finding invalidates Okros's (1988) evaluation of the psychometric properties of individual items on the MOST as well as Stouffer's (1996) assessment of its internal consistency. As discussed previously, the difficulty level of omitted items is spuriously elevated and large rates of omission yield inflated estimates of internal consistency. Therefore, all we may conclude at this point is that: (a) 37 of 60 items have a distractor that is so obviously incorrect that it is selected by less than 5% of examinees; and that (b) five items have an abnormally high percentage of examinees choosing the same incorrect answer. This latter finding suggests that these five items (items 7, 23, 31, 42, and 55) might have been miskeyed. The correct responses should be reviewed to insure their accuracy.

Construct-related validity. In contrast with what has been previously reported (see Okros, 1988), it seems that the MOST is not measuring what it purports to measure (i.e., memory, selective attention, and problem-solving). Instead, all analyses converge in suggesting that it measures three oblique factors including speediness (which may be defined as speed in finding the correct answer to each problem), spatial scanning (which may be defined as "speed in exploring visually a wide or complicated field", Ekstrom et al., 1976, p. 155), and general reasoning (which may be defined as "the ability to select and organize relevant information for the solution of a problem", Ekstrom et al., 1976, p. 133). Although the MOST correlates with a measure of memory, hierarchical regression analyses suggest that individual differences in mnemonic ability do not contribute to variation in aptitude test performance. As shown, when the MOST is first regressed onto measures of spatial scanning and general reasoning, the effect of memory is not significant. As discussed previously, there are at least three possible explanations for this result: (1) it may be that most examinees have no difficulty to memorize the progressively more complex sets of navigational rules; (2) that passed a certain point, having a good memory has no direct impact on aptitude test performance; or that the variance accounted for by memory may already have been predicted by general reasoning and spatial scanning.

Adverse effect discrimination. As explained previously, adverse effect discrimination occurs when employers use selection procedures that have an unintended negative effect on protected groups (e.g., using the Canadian Forces Aptitude Test for the selection of military personnel might prevent some aboriginal people from enrolling into several military occupations; Vanderpool, 2003). Adverse effect discrimination can be detected by measuring group differences in selection/success rates (i.e., adverse impact), detecting differential item functioning (DIF), and evaluating group differences in predicted score on a criterion (e.g., training performance). Although the former two categories of adverse effect can be assessed without having access to criterion data, the latter type cannot be assessed without that information. As a result, only the former two categories could be evaluated at this time.

As shown, when a score of 24 is used as a passing mark on the MOST, the success rate of women is 94% that of the men. As a result, there is no indication that using the MOST for the selection of future MARS Officers has an adverse impact on women. Furthermore, the absence of DIF suggests that none of the MOST items are biased against women. Together, these two results suggest that using the MOST for the selection of male and female applicants may be legally defensible (Employment Equity Act, 1995 c. 44).

Limitations

In addition to the typical limitations associated with correlational designs (i.e., no manipulation of variables, no control of the research environment, no random selection of research participants, design does not eliminate possibility of alternate explanations for results, and no possibility to draw cause and effect conclusions from the results), several factors may have had an impact on the findings. Firstly, the sample for the second study was too small for the type of analyses performed. As a result, logistic regressions lacked statistical power to identify DIF and the results of the confirmatory factor analysis may be unstable (i.e., they may not replicate with a larger sample). Secondly, errors of measurement (low levels of test reliability) may have obscured the true relationship between constructs (Crocker & Algina, 1986). For instance, a stronger relationship may have been found between the MOST and memory if more reliable measures had been used.

Implications for Future Research

Future research should first improve the quality of item alternatives and confirm the accuracy of the keyed response to items 7, 23, 31, 42, and 55. Next, it will be necessary to reassess the reliability of the MOST to determine the amount of measurement error associated with content sampling and changes in examinees (e.g., nervousness, physical discomfort, and practice effects). If the amount of measurement error is acceptable (i.e., reliability $\geq .80$), the next step will consist of adjusting the minimum passing mark (based on a new evaluation of the *SEM*) and assessing whether using the new cut-off score has an adverse impact on protected groups.

Recommendations

Firstly, the MOST should *not* be used for decision making until its reliability has been established. In the interim, the use of relevant measures from the Kit of Factor-Referenced Cognitive Tests (i.e., Mathematics Aptitude Test, Spatial Scanning Test, and Auditory Number Span Test) is recommended. Although these tests were not designed for use in selection (Ekstrom et al., 1976), this thesis indicates that they are suitable for MARS Officer selection. In fact, they all measure MARS-specific abilities and their published reliabilities are acceptable.

Secondly, MARCOM should reassess the predictive validity of the NOAB measures; and if necessary, look for suitable alternative selection procedures. Table 18 presents a sample of alternative selection procedures that may help increase the criterion-related validity of the NOAB. As shown, work sample tests are usually the best predictors of occupational performance ($r = .54$) followed by general intelligence tests ($r = .51$), and integrity tests ($r = .41$). Personality tests are also good predictors of occupational performance ($r = .31$), and MARCOM should examine the possibility of adding personality tests to its current selection procedures. Recent research by the Royal Navy has shown that high levels of extraversion and low levels of agreeableness are associated with higher leadership ratings at the Admiralty Interview Board (Perry, 1999).

Table 18

Alternative Selection Procedures

Selection Procedure	Mean Validity	Gender Differences	Perceptions of Procedural Fairness
Work sample tests	.54[a]	Small[c]	Above Average[d]
General intelligence (g) tests	.51[a]	None[b]	—
Ability tests	—	Some[b]	Average[d]
Biodata	.35[a]	None[e]	Average[d]
Academic performance	.32[e]	—	—
Personality tests	.31[a]	Small[b]	Below Average[d]
Reference checks	.26[a]	—	Average[d]

Note. Dashes (–) indicate that no information was found. Perceptions of fairness may vary across cultures (Steiner & Gilliland, 1996). *Sources*: [a]Schmidt and Hunter (1998); [b]Hough et al., 2001; [c]Salgado et al., 2001); [d]Steiner and Gilliland (1996); [e]Reilly and Chao (1982).

Finally, MARCOM should undertake research to identify other possible reasons for the high levels of attrition among junior MARS Officers (Rodgers, 1986). The NOAB was established to improve the quality of MARS applicants, but there has been no research to show that the lack of MARS-specific abilities is an issue. It may be equally useful to examine junior officers' reactions towards occupational training because organizational research has shown that personal evaluations of organizational justice predict both turnover intentions and organizational commitment (Cohen-Charash & Spector, 2001; Cropanzano & Greenberg, 1997).

References

Angoff, W. H. (1971). Norms, scales, and equivalent scores. In R. L. Thorndike (ed.). Educational Measurement (2nd Ed.). Washington, DC: American Council on Education

Bradley, J. P. (1990). *A validation study on the Naval Officer Assessment Board's ability to predict MARS Officer training success* (Working Paper 90-7). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Bradley, J. P., Wiesner, W. H., & Latham, G. P. (1993). *Comparison of the situational interview and patterned behaviour description interview formats at the Naval Officer Assessment Board* (Technical Note 23/93). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Boswell, R. A. (1993). Review of naval officer selection methods in TCCP countries (Technical Note 21/93). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Canadian Psychological Association (1996). *Guidelines for educational and psychological testing* (2nd Ed.), Ottawa, Ontario, Canada: Author.

Caplan, P. J., & Caplan, J. B. (1999). *Thinking critically about research on sex and gender*. New York: Addison-Wesley Educational Publishers.

Catano, V. M. (1989). *Naval officer selection board: Methods for assessing utility* (Working Paper 89-9). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Catano, V. M., Cronshaw, S. F., Wiesner, W. H., Hackett, R. D., & Methot, L. L. (2001). *Recruitment and selection in Canada*. Toronto, Ontario, Canada: Nelson Thomson Learning

Clauser, B. E., Nungester, R. J., Mazor, K., & Ripkey, D. (1996). A comparison of alternative matching strategies for DIF detection in tests that are multidimensional. *Journal of Educational Measurement, 33*(2), 202-214.

Cleary, T. A. (1968). Test bias: Prediction of grades of Negro and white students in integrated college. *Journal of Educational Measurement, 5*, 115-124.

Cohen, R. J., Swerdlik, M. E., & Smith, D. K. (1992). *Psychological testing and assessment: An introduction to tests and measurement*. Mountain View, CA: Mayfield Publishing Company.

Cohen-Charash, Y., & Spector, P. E. (2001). The role of justice in organizations: A meta-analysis. *Organizational Behavior and Human Decision Processes, 86*(2), 278-321.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Orlando, FL: Harcourt Brace Jovanovich.

Cropanzano, R., & Greenberg, J. (1997). Progress in organizational justice: Tunneling through the maze. In C.L. Cooper & I. T. Robertson (Eds.). *International Review of Industrial and Organizational Psychology*, vol 12 (pp.317- 372). London: John Wiley & Sons.

Department of Justice Canada. (1995, c.44). *Employment Equity Act*. Retrieved on October 27, 2002, from http://laws.justice.gc.ca/en/E-5.401/index.html.

Drasgow, F., & Hulin, C. L. (1990). Item response theory. In M. D. Dunnette, & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology*, (pp. 577-636). Palo Alto, CA: Consulting Psychologists Press.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1979). Cognitive factors: Their identification and replication. *Multivariate Behavioral Research Monographs, 79*(2).

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

Elkins, T. J., & Phillips, J. S. (2000). Job context, selection decision outcome, and the perceived fairness of selection tests: Biodata as an illustrative case. *Journal of Applied Psychology, 85*(3), 479-484.

Eyres, S. A. T., & Rodgers, M. N. (1988). *Validation of the Combat Arms Officer Selection Board* (Working Paper 88-2). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272-299.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*(2), 95-103.

Fleishman, E. A., & Reilly, M. E. (1992). *Handbook of human abilities*. Palo Alto, CA: Consulting Psychologist Press, Inc.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-De Lisi, A. V., Morely, M., & Cahalan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology, 75*(3), 165-190.

Gilliland, S. W., & Chan, D. (2001). Cognitive ability. In N. Anderson, D. S. Ones, H. K. Sinangil, &C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (pp. 143-165). Thousand Oaks, CA: Sage Publications.

Harvey, R. J. (1991). Job analysis. In *M.D.* Dunnette, & L. M. Hough (Eds.), *Handbook of Industrial and Organizational Psychology* (pp. 71-163). Palo Alto, CA: Consulting Psychologists Press.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer, & H. L. Braun (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*(1/2), 152-194.

Hyde, J. S. (1981). How large are cognitive gender differences? *American Psychologist, 36*(8), 892-901.

Hyde, J. S. (1990). Meta-analysis and the psychology of gender differences. *Signs, 16*(1), 55-73.

Hyde, J. S. (1994). Can meta-analysis make feminist transformations in psychology? *Psychology of Women Quarterly, 18*(4), 451-462.

Hyde, J. S., & Linn, M. C. (1986). *The psychology of gender: Advances through meta-analysis*. Baltimore, MD: The Johns Hopkins University Press.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53-69.

Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139-155.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. Applied Measurement in Education, 14(4), 329-349.

Jöreskog, K.G., & Sörbom, D. (2002). *Lisrel 8.53 user's guide*. Chicago, IL: Scientific Software International Inc.

Kelloway, E.K. (1998). *Using LISREL for structural equations modeling: A researcher's guide.* Thousand Oaks, CA: Sage Publications.

King, W. D. (1989). *Assessment centre pre-screening: An empirical double cross-validation approach.* Unpublished master's thesis, University of Waterloo, Waterloo, Ontario, Canada.

Levine, N. G., Mallamad, M. S., & Fleishman, A. E. (1978). *Decision aids in estimating personnel requirements.* Washington, DC: Advanced Research Resources Organization.

Linn, M. C., & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67-101). Baltimore, MD: The Johns Hopkins University Press.

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

MacIntyre, T. (1997). Gender differences in cognition: A minefield of research issues. *The Irish Journal of Psychology, 18*(4), 386-396.

Mallamad, S. M., Levine, J. M., & Fleishman, E. A. (1980). Identifying ability requirements by decision flow diagrams. *Human Factors, 22*(1), 57-68.

Mallamad, S. M., Levine, J. M., & Fleishman, E. A. (1980). Identifying ability requirements by decision flow diagrams. *Human Factors, 22*(1), 57-68.

Mazor, K. M., Kanjee, A., & Clauser, B. E. (1995). Using logistic regression and the Mantel-Haenszel with multiple ability estimates to detect differential item functioning. *Journal of Educational Measurement, 32*(2), 131-144.

McAnulty, D. M., Cross, K. D., & Jones, D. H. (1986). *The development of an experimental battery of aviation-related ability tests.* (Draft Technical Report). Fort Rucker, Alabama: United States Army Research Institute Aviation Research and Development Activity.

Navas-Ara, M. J., & Gomez-Benito, J. (2002). Effects of ability scale purification on the identification of DIF. *European Journal of Psychological Assessment, 18*(1), 9-15.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory. New York: McGraw-Hill.

Okros, A. C. (1988). *Analysis of Maritime Surface officer classification results* (Technical Note 10/88). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Okros, A. C., & King, W. D. (1989). *Pre-screening of Naval Officer Assessment Board candidates* (Technical Note 7/89). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Okros, A. C., Johnston, V. W., & Rogers, M. N. (1988). *An evaluation of the effectiveness of the Naval Officer Selection Board as a predictor of success on the Basic Officer Training Course* (Working Paper 88-1). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Perry, A. (1999). Intelligence and personality at assessment centers: Examining the relationship between intelligence and personality measures to performance at the Royal Navy's Admiralty interview Board. University of London, London, United Kingdom.

Ree, M. J., Carretta, T. R., & Steindl, J. R. (2001). Cognitive ability. In N. Anderson, D. S. Ones, H. K. Sinangil, &C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (pp. 166-199). Thousand Oaks, CA: Sage Publications.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35*, 1-62.

Rodgers, M. N. (1984). *An overview of the Naval Officer production system* (Technical Note 12/84). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Rodgers, M. N. (1985). *An evaluation of the effectiveness of Naval Officer Interview Board serial 8305 to 8404* (Technical Note 18/85). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Rodgers, M. N. (1986). *The development of Maritime Surface and Sub-surface Officer training performance measures* (Technical Note 11/86). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Rodgers, M. N. (1986). *The identification of ability requirements for MARS Officer Selection* (Technical Note 23/86). Willowdale, Ontario: Canadian Forces Personnel Applied Research Unit.

Saks, A. M. (2000). *Research, measurement, and evaluation of human resources.* Toronto: Nelson.

Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D. S. Ones, H. K. Sinangil, &C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology* (pp. 166-199). Thousand Oaks, CA: Sage Publications.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.

Scholtz, D. (2002). *Naval officer selection board* (Annex A to MS: 4500-6-3 [DMTE]). Ottawa, Ontario, Canada: Director of Maritime Command.

Scholtz, D. (2003). *Establishing defensible cut-off scores for the Naval Officer Assessment Board* (Research Report). Ottawa, Ontario, Canada: Director of Maritime Command.

Society For Industrial and Organizational Psychology, Inc. (1987). *Principles for the validation and use of personnel selection procedures* (3rd ed.). College Park, MD: Author.

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81*(2), 134-141.

Stouffer, J. M. (1996). *Psychometric evaluation of the Canadian Forces Aptitude Test (CFAT) and the Passage Planning (PPT) Test* (Technical Note 1/96). Willowdale, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

Stumpf, H., & Klieme, E. (1989). Sex-related differences in spatial ability: More evidence for convergence. *Perceptual and Motor Skills, 69(*3), 915-921.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27(4), 361-370.

Swanson, D. B., Clauser, B. E., Case, S. M., Nungester, R. J., & Featherman, C. (2002). Analysis of differential item functioning (DIF) using hierarchical logistic regression models. *Journal of Educational and Behavioral Statistics, 27*(1), 53-75.

Tabachnick, B. G., & Fidelll, L. S. (2001). *Using multivariate statistics* (4th Ed.).Boston: Allyn and Bacon.

Uniform Guidelines on Employee Selection Procedures. (1978). *Federal Register* 43: 38290-315.

Vanderpool, M. A. (2003). *Determining if the Canadian Forces Aptitude Test is Biased Against Canadian Aboriginal Peoples.* Unpublished master's thesis, Saint- Mary's University, Halifax, Nova Scotia, Canada.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial ability: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Whetzel, D. L., & Wheaton, G. R. (Eds.). (1997). *Applied measurement methods in industrial psychology.* Palo Alto, CA: Davis-Black Publishing.

Whitmore, M. L., & Schumacker, R. E. (1999). A comparison of logistic regression and analysis of variance differential item functioning detection methods. *Educational and Psychological Measurement, 59*(6), 910-927.

Zumbo, B. D. (1999). A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores. Ottawa, ON: Directorate of Human Resources Research and Evaluation, Department of National Defence

Appendix A

Background Information Questionnaire

Student Identification Number (last five digits only): ___ ___ ___ ___ ___

Gender: ❑ Male ❑ Female Age: ________ First Language: ________________

Looking at these categories, which one best describes your ethnic background?

❑ White ❑ Black ❑ First Nation ❑ Asian ❑ Hispanic ❑ Arabic
❑ Other (specify): ____________________

What is your present level of education?

❑ 1st year university ❑ 2nd year university ❑ 3rd year university
❑ 4th year university ❑ Graduate student

What academic program are you enrolled in (check only one box)?

Faculty of Arts

❑ Anthropology ❑ English ❑ Geography ❑ History
❑ Irish Studies ❑ Philosophy ❑ Linguistic ❑ Women Studies
❑ Political Science ❑ Psychology ❑ Religious Studies ❑ Sociology
❑ Atlantic Canada Studies ❑ Asian Studies ❑ Criminology
❑ International Development Studies ❑ Modern Languages and Classics

Faculty of Science

❑ Biology ❑ Engineering ❑ Geology
❑ Psychology ❑ Astronomy & Physics ❑ Environmental Studies
❑ Math & Computing Science ❑ Chemistry

Faculty of Commerce

❑ Accounting ❑ Commercial Law ❑ Communications ❑Economics
❑ Management ❑ Marketing ❑ Finance & Management Science
❑ Global Business Management ❑ Computing Science/Business Admin.

Faculty of Graduate Studies

❑ Ph.D. in Management ❑ M.A. in Atlantic Canada Studies ❑ M.A. in Criminology
❑ M.A. in History ❑ M.A. in Philosophy ❑ MBA
❑ M.A. in Women's Studies ❑ EMBA ❑ MSc. in Astronomy
❑ MSc. in Applied Psychology ❑ MSc. In Applied Science
❑ M.A. in International Development Studies

❑ Other (Specify): ____________________

How many Math courses have you completed since you graduated from High School? ___
How many Physics courses have you completed since you graduated from high School? ___
How many Chemistry courses have you completed since you graduated from High School? ___
How many Computer Science courses have you completed since you graduated from High School? ___

Appendix B

VERBATIM INSTRUCTIONS

FOR EXPERIMENT NAVY

CANADIAN FORCES

Chief of the Maritime Staff

WINTER 2003

001

When everyone is seated and you are ready to begin testing, say:

GOOD MORNING/AFTERNOON, I AM ______________________ AND I WILL BE CONDUCTING THIS TESTING SESSION. PLEASE MAKE YOURSELF COMFORTABLE. IF YOU WISH TO REMOVE YOUR JACKET OR SWEATER DO SO NOW AND HANG IT ON THE BACK OF YOUR SEAT. ANYTHING ELSE YOU BROUGHT WITH YOU INTO THIS ROOM SHOULD NOT BE ON YOUR DESK, BUT SHOULD BE PLACED ON THE FLOOR UNDER YOUR SEAT/DESK. DO NOT OPEN THE TEST BOOKLETS UNTIL YOU ARE TOLD TO DO SO.

Wait until all noise ceases, then say:

THIS MARKS THE BEGINNING OF EXPERIMENT NAVY. I WANT YOU TO PAY CLOSE ATTENTION TO EVERYTHING I SAY. IF YOU DO NOT UNDERSTAND ANY POINT, PLEASE RAISE YOUR HAND.

THIS STUDY INVOLVES THE COMPLETION OF TWO COGNITIVE ABILITY TEST BATTERIES AND FOUR QUESTIONNAIRES. THE TOTAL TIME REQUIRED FOR THE COMPLETION OF THESE TEST BATTERIES AND QUESTIONNAIRES CAN RANGE FROM TWO AND A HALF HOURS TO 3 HOURS. IF YOU CANNOT STAY HERE FOR THE COMPLETE DURATION OF THE EXPERIEMENT, PLEASE RAISE YOUR HAND AND ANOTHER APPOINTMENT WILL BE MADE FOR YOU.

AS MENTIONED BEFORE, YOU WILL BE PAID $10 FOR APPROXIMATELY 2.5 HOURS OF TESTING. YOU MIGHT ALSO RECEIVE AN ADDITIONAL $50 IF YOU DISPLAY THE ABILITIES REQUIRED TO BECOME A NAVAL OFFICER IN THE CANADIAN FORCES.

FOR THE PURPOSE OF THIS STUDY, YOUR SUITABILITY TO BECOME A NAVAL OFFICER WILL BE DETRMINED BASED ON YOUR CUMULATIVE PERFORMANCE ON TWO COGNITIVE ABILITY TEST BATTERIES. IF YOUR TOTAL SCORE ON THESE TEST BATTERIES EXCEED THAT OF YOUR PEERS,

THAT IS IF YOU PLACE IN THE TOP 20%, YOU WILL RECEIVE THE $50 PERFORMANCE BONUS. IF YOUR TOTAL SCORE FALLS BELOW THE 20TH PERCENTILE, YOU WILL <u>NOT</u> RECEIVE THE $50 PERFORMANCE BONUS. THE PERORMANCE BONUSES WILL BE DISTRIBUTED IN ROOM _____ FROM MARCH 10 TO MARCH 14. THE LIST OF ELIGIBLE PARTICIPANTS WILL BE POSTED OUTSIDE ROOM _______. IF THE LAST FIVE DIGITS OF YOU STUDENT IDENTIFICATION NUMBER ARE ON THAT LIST IT MEANS THAT YOU ARE ELIGIBLE FOR A PERFORMANCE BONUS. IF THE LAST FIVE DIGITS OF YOUR STUDENT IDENTIFICATION NUMBER ARE NOT ON THE LIST IT MEANS THAT YOU ARE NOT ELIGIBLE FOR A BONUS. DO YOU HAVE ANY QUESTIONS?

Answer any questions, and then say:

IF YOU HAVE NO OTHER QUESTIONS, PLEASE TAKE A FEW MINUTES TO READ AND SIGN THE INFORMED CONSENT FORM. SHOULD YOU WISH TO KEEP A COPY OF THIS FORM FOR YOU OWN RECORDS, YOU CAN HAVE ONE AT THE END OF THE EXPERIMENT. PLEASE READ AND SIGN THE CONSENT FORM NOW.

Wait until everyone has signed the Consent Form, and then say:

ONCE YOU HAVE SIGNED THE CONSENT FORM YOU CAN GO TO THE NEXT PAGE, AND COMPLETE THE BACKGROUND INFORMATION QUESTIONNAIRE. <u>PLEASE DO NOT FORGET TO WRITE DOWN THE LAST FIVE DIGITS OF YOUR STUDENT IDENTIFICATION NUMBER</u>. THIS INFORMATION WILL BE NEEDED FOR THE ADDITIONAL PAYMENT OF PERFORMANCE BONUSES. ONCE YOU HAVE COMPLETED THE BACKGROUND INFORMATION QUESTIONNAIRE, PUT YOUR PENCIL DOWN AND WAIT FOR FURTHER INSTRUCTIONS.

Wait until everyone has completed the questionnaire, and then say:

NOW GO TO THE NEXT PAGE, AND PAY CLOSE ATTENTION TO EVERYTHING I SAY.
(pause)
BECAUSE YOUR ELIGIBILITY FOR A $50 PERFORMANCE BONUS IS CONTINGENT UPON YOUR DISPLAYED ABILITY TO BECOME A NAVAL OFFICER IT IS IMPORTANT THAT YOU CLEARLY UNDERSTAND WHAT IT MEANS TO BE A NAVAL OFFICER IN THE CANADIAN FORCES.

TO HELP YOU FORM A MENTAL PICTURE OF THAT UNIQUE CAREER PATH, I HAVE ENCLOSED A TWO-PAGE DESCRIPTION OF THAT OCCUPATION. PLEASE READ PARAGRAPH 1, 2, AND 5 TO YOURSELF WHILE I READ THEM ALOUD. (Read only the first five paragraphs)

DO YOU HAVE ANY QUESTIONS?

Answer any questions, and then say:

GO TO THE NEXT PAGE, AND PAY ATTENTION TO WHAT I SAY.

FOR THE FIRST PART OF THE TESTING PROCESS YOU ARE TO TAKE A SERIES OF 5 SUB-TESTS COMPRISING THE MARITIME OFFICER SELECTION TEST. THESE TESTS WILL BE USED TO HELP DETERMINE YOUR SUITABILITY TO BECOME A NAVAL OFFICER IN THE CANADIAN FORCES. IT WILL TAKE ABOUT ONE AND A HALF HOURS TO COMPLETE ALL FIVE TESTS.

IT IS IMPORTANT THAT YOU BE IN THE BEST CONDITION FOR WRITING THESE TESTS TODAY. IF YOU DO NOT FEEL WELL, IF YOU ARE VERY TIRED, OR IF YOU ARE FEELING THE EFFECTS OF MEDICATION, DRUGS OR ALCOHOL, YOU SHOULD WRITE THESE TESTS AT ANOTHER TIME.

IF ANY OF THESE CONDITIONS APPLY TO YOU, RAISE YOUR HAND AND

ANOTHER APPOINTMENT WILL BE MADE FOR YOU. ARE THERE ANY QUESTIONS?

Answer any questions, and the say:

FOR SECURITY REASONS, IT IS IMPORTANT THAT YOU DO NOT REMOVE ANY OF THE TEST MATERIALS, OR EVEN YOUR SCRAP PAPER FROM THE ROOM. SIMILARLY, UNDER CANADIAN FORCES REGULTIONS, YOU ARE PROHIBITED FROM DISCUSSING THE ITEMS ON THIS TEST WITH ANYONE OUTSIDE THIS ROOM. DO YOU UNDERSTAND?

Pause, ensure everyone complies, answer any questions, and continue.

IN ORDER TO BE FAIR TO EVERYONE, IT IS IMPORTANT THAT CERTAIN BASIC RULES BE OBSERVED STRICTLY. YOU MUST WORK SILENTLY. YOU MAY NOT ASK ANY QUESTIONS ONCE THE TESTS HAVE STARTED. YOU MAY NOT TALK TO OTHERS TAKING THE TEST OR LOOK AT ANYONE ELSE'S WORK. USE OF CALCULATORS IS NOT ALLOWED DURING TESTING. IF YOU HAVE A CALCULATOR OF ANY TYPE, PLEASE PLACE IT IN YOUR POCKET OR ON THE FLOOR UNDER YOUR SEAT/DESK. TURN OFF ANY WATCH CHIMES, ALARMS, CELLULAR PHONES, BEEPERS AND PAGERS.

If required, wait for all activity to cease, then say:

TURN TO PAGE 1 IN YOUR TEST BOOKLET. READ THE DIRECTIONS TO YOURSELF WHILE I READ THEM ALOUD.

DO NOT WRITE OR MAKE MARKS OF ANY KIND IN THIS TEST BOOKLET. DO YOUR ROUGH WORK ON THE SCRAP PAPER PROVIDED. YOUR ANSWERS TO THE TEST ARE TO BE RECORDED ON A SEPARATE ANSWER SHEET.

PLEASE IGNORE PARAGRAPH NUMBER 2.

TURN TO PAGE 3 IN YOUR TEST BOOKLET.

FOR EACH QUESTION BE SURE TO PICK OUT THE BEST ONE OF THE POSSIBLE ANSWERS LISTED. WHEN YOU HAVE DECIDED WHICH ONE OF THE CHOICES GIVEN IS THE BEST ANSWER TO A QUESTION, BLACKEN THE CIRCILE ON YOUR SEPARATE ANSWER SHEET IDENTIFIED BY THE SAME NUMBER AS THE QUESTION YOU ARE ANSWERING. BLACKEN THE CIRCLE NEATLY WITH A HEAVY BLACK MARK AS SHOWN IN THE SAMPLE PROBLEM 1 BELOW.

THE CORRECT ANSWER TO THE SAMPLE PROBLEM IS 3. NOTE HOW CIRCLE 3 OPPOSITE QUESTIONS NUMBER 1 HAS BEEN BLACKENED ABOVE. YOUR MARKS SHOULD LOOK JUST LIKE THIS AND BE PLACED IN THE CIRCLE IDENTIFIED BY THE SAME NUMBER AND LETTER AS THE CORRECT ANSWER TO EACH ITEM. IF YOU WANT TO CHANGE AN ANSWER, ERASE YOUR FIRST MAK COMPLETELY. QUESTIONS ANSWERED TWICE WILL BE COUNTED AS WRONG.

DO NOT SPEND TOO MUCH TME ON ANY ONE ITEM. WORK AS RAPIDLY AS POSSIBLE WITHOUT SACRIFICING ACCURACY. TRY TO ANSWER EVERY QUESTION. IF YOU CANNOT DETERMINE THE ANSWER TO AN ITEM BUT CAN ELIMINATE SOME OF THE CHOICES, MAKE YOUR BEST GUESS, SINCE YOU ARE NOT PENALIZED FOR WRONG ANSWERS. ALWAYS MAKE SURE THAT THE NUMBER ON THE ANSWER SHEET IS THE SAME AS THE NUMBER OF THE QUESTION IN THE TEST BOOKLET. DO NOT MARK IN THE TEST BOOKLET.

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO.

Pause and then say:

EACH OF THESE TESTS IS TIMED. AT THE END OF EACH TEST I WILL SAY, "STOP". YOU ARE TO STOP IMMEDIATELY AND PLACE YOUR PENCIL ON THE DESK. IF YOU FINISH A TEST BEFORE TIME IS CALLED, YOU MAY GO

BACK AND CHECK YOUR WORK <u>ON THAT TEST ONLY</u>. DO NOT GO ON TO THE NEXT TEST UNTIL YOU ARE TOLD TO DO SO, AND DO NOT TURN BACK TO ANY OF THE PREVIOUS TESTS, OR YOUR TEST WILL BE INVALID AND YOUR PROCESSING WILL STOP. DO YOU UNDERSTAND?

Pause and continue:

YOU ARE RESPONSIBLE FOR MONITORING YOUR OWN TIME.

IF YOU NEED ANOTHER PENCIL OR MORE SCRAP PAPER DURING THE TEST, RAISE YOUR HAND WITH THE ITEM IN IT. ARE THERE ANY QUESTIONS?

Pause and continue:

THE TEST YOU ARE ABOUT TO TAKE IS DESIGNED TO TEST YOUR ABILITY TO LEARN AND APPLY A SET OF RULES TO SOLVE PROBLEMS. THE TEST CONSISTS OF THREE LEVELS. FOR EACH LEVEL, THERE ARE A SET OF DIRECTIONS WHICH YOU WILL HAVE TO READ AND LEARN. YOU ARE REQUIRED TO HOLD THESE DIRECTIONS IN MEMORY. YOU ARE NOT PERMITTED TO WRITE DOWN THE DIRECTIONS ON YOUR SCRAP PAPER, OR MAKE NOTES ABOUT THE DIRECTIONS. IS THAT UNDERSTOOD?

Pause and continue:

WITHIN EACH LEVEL OF THE TEST, YOU WILL BE GIVEN A SPECIFIC PERIOD OF TIME TO READ AND LEARN THE DIRECTIONS. FOR THE FIRST TWO LEVELS, THIS IS FOLLOWED BY A SET OF PRACTICE PROBLEMS TO TEST YOUR UNDERSTANDING OF THE DIRECTIONS. THE PRACTICE PROBLEMS ARE ALSO TIMED. FINALLY, YOU WILL COMPLETE THE TEST ITEMS. ARE THERE ANY QUESTIONS?

Pause and continue:

PLEASE TURN TO PAGE 4 IN YOUR TEST BOOKLET. YOU WILL HAVE 6 MINUTES TO READ AND LEARN THE DIRECTIONS ON PAGE 4 AND THE TOP OF PAGE 5 <u>AND</u> COMPLETE THE PRACTICE ITEMS ON PAGE 5. FOR THE

PRACTICE ITEMS, USE THE PRACTICE CHART ON PAGE 4. THERE ARE FOUR ANSWER CHOICES FOR EACH PRACTICE ITEM. CHOOSE THE BEST ANSWER THEN MARK THE SAME LETTER ON THE ANSWER SHEET. PLEASE FILL IN YOUR RESPONSES DIRECTLY ON YOUR ANSWER SHEET IN SPACE 1, 2, AND 3. AGAIN YOU HAVE 6 MINUTES TO COMPLETE THESE TWO PAGES. BEGIN NOW.

Time 6 minutes, then say:

STOP. TURN TO PAGE 7, WHERE YOU WILL FIND THE ANSWERS TO THE PRACTICE ITEMS. THE ANSWERS ARE ILLUSTRATED ON THE CHART BELOW. PLEASE CHECK YOUR ANSWERS.
(pause)
ARE THERE ANY QUESTIONS?

Pause, then say:

TURN TO PAGE 9. THE RULES FROM LEVEL 1 ARE SUMMARIZED FOR YOU AT THE TOP OF THE PAGE. BEFORE PROCEEDING WITH THIS PORTION OF THE TEST, YOU WILL HAVE 30 SECONDS TO REVIEW THE BASIC RULES OF PASSAGE PLANNING. BEGIN NOW.

Time 30 seconds, then say:

YOU HAVE 6 MINUTES TO COMPLETE TWELVE ITEMS. TURN TO PAGE 10. BEGIN NOW.

Time 6 minutes, then say:

STOP. PUT YOUR PENCILS DOWN.

YOU HAVE NOW FINISHED LEVEL 1 OF THE TEST. LEVEL 2 HAS THE SAME FORMAT. YOU WILL READ AND LEARN AN ADDITIONAL SET OF DIRECTIONS. PLEASE NOTE, THE DIRECTIONS FROM LEVEL 1 STILL APPLY. THE LEVEL 2 DIRECTIONS ARE TO BE USED IN ADDITION TO LEVEL 1 DIRECTIONS. ONCE AGAIN, THERE ARE PRACTICE ITEMS BEFORE THE TEST ITEMS.

TURN TO PAGE 12. YOU HAVE 8 MINUTES TO READ THE DIRECTIONS ON PAGE 12 AND THE TOP OF PAGE 13 AND COMPLETE THE PRACTICE ITEMS ON PAGE 13. USE THE PRACTICE CHART TO ANSWER THE PRACTICE ITEMS. DECIDE ON THE BEST ANSWER, THEN MARK THE SAME LETTER ON THE ANSWER SHEET. AGAIN, YOU HAVE 8 MINUTES TO COMPLETE THESE TWO PAGES. BEGIN NOW.

Time 8 minutes. Then say:

STOP. TURN TO PAGE 15, WHERE YOU WILL FIND THE ANSWERS TO THE PRACTICE ITEMS. THE ANSWERS ARE ILLUSTRATED ON THE CHART BELOW. PLEASE CHECK YOUR ANSWERS. (pause) ARE THERE ANY QUESTIONS?

Pause, then say:

TURN TO PAGE 17. THE RULES FROM LEVEL 2 ARE SUMMARIZED FOR YOU. BEFORE PROCEEDING WITH THIS PORTION OF THE TEST, YOU WILL HAVE 30 SECONDS TO REVIEW THE LEVEL 2 RULES. BEGIN NOW.

Time 30 seconds, then say:

STOP. PLEASE GIVE ME YOUR COMPLETE ATTENTION NOW, AS I EXPLAIN THE PROCEDURES FOR THE NEXT PORTION OF THE TEST. YOU WILL HAVE 8 MINUTES TO ANSWER 12 TEST ITEMS. WORK AS RAPIDLY AS POSSIBLE WITHOUT SACRIFICING ACCURACY. IF YOU CANNOT DETERMINE THE ANSWER TO AN ITEM BUT CAN ELIMINATE SOME OF THE CHOICES, MAKE YOUR BEST GUESS. IF YOU CHANGE AN ANSWER, ERASE YOUR FIRST ANSWER COMPLETELY. MAKE SURE THE ANSWER YOU SELECT CORRESPONDS TO THE CORRECT QUESTION NUMBER ON THE ANSWER SHEET. IF YOU FINISH THE CHART ITEMS, CHECK YOUR ANSWERS FOR THIS SET OF ITEMS ONLY. DO NOT GO ON TO

THE NEXT CHART. DO NOT TURN BACK TO THE INSTRUCTIONS OR TO THE LEVEL 1 ITEMS. ARE THERE ANY QUESTIONS?

Pause, then say:

YOU HAVE 8 MINUTES TO COMPLETE TWELVE ITEMS. TURN TO PAGE 18. BEGIN NOW.

Time 8 minutes, then say:

STOP. TURN TO PAGE 21. BEFORE PROCEEDING WITH THE NEXT PORTION OF THE TEST, YOU WILL HAVE 30 SECONDS TO REVIEW THE LEVEL 2 RULES. BEGIN NOW.

Time 30 seconds, then say:

STOP. PLEASE GIVE ME YOUR COMPLETE ATTENTION NOW, AS I EXPLAIN THE NEXT PORTION OF THE TEST. AGAIN, YOU WILL HAVE 8 MINUTES TO ANSWER 12 TEST ITEMS. WORK AS RAPIDLY AS POSSIBLE WITHOUT SACRIFICING ACCURACY. IF YOU FINISH THE CHART ITEMS, CHECK YOUR ANSWERS FOR THIS SET OF TEST QUESTIONS ONLY. DO NOT GO ON TO THE NEXT CHART. DO NOT TURN BACK TO THE INSTRUCTIONS OR TO ANY PREVIOUS TEST ITEMS. ARE THERE ANY QUESTIONS?

Pause, then say:

YOU HAVE 8 MINUTES TO COMPLETE TWELVE ITEMS. TURN TO PAGE 22. BEGIN NOW.

Time 8 minutes, then say:

STOP TURN TO PAGE 25. BEFORE PROCEEDING WITH THE NEXT PORTION OF THE TEST, YOU WILL HAVE 30 SECONDS TO REVIEW THE LEVEL 2 RULES. BEGIN NOW.

Time 30 seconds, then say:

STOP. PLEASE GIVE ME YOUR COMPLETE ATTENTION, AS I EXPLAIN THE NEXT PORTION OF THE TEST.

AGAIN, YOU WILL HAVE 8 MINUTES TO ANSWER 12 TEST ITEMS. WORK AS RAPIDLY AS POSSIBLE WITHOUT SACRIFICING ACCURACY. IF YOU CANNOT DETERMINE THE ANSWER TO AN ITEM BUT CAN ELIMINATE SOME OF THE CHOICES, MAKE YOUR BEST GUESS. IF YOU CHANGE AN ANSWER, ERASE YOUR FIRST ANSWER COMPLETELY. MAKE SURE THE ANSWER YOU SELECT CORRESPONDS TO THE CORRECT QUESTION NUMBER ON THE ANSWER SHEET. IF YOU FINISH THE CHART ITEMS, CHECK YOUR ANSWERS FOR THIS SET OF QUESTIONS ONLY. DO NOT GO ON TO THE NEXT CHART. DO NOT TURN BACK TO THE INSTRUCTIONS OR TO ANY PREVIOUS TEST ITEMS. ARE THERE ANY QUESTIONS?

Pause, then say:

YOU HAVE 8 MINUTES TO COMPLETE TWELVE ITEMS. TURN TO PAGE 26. BEGIN NOW.

Time 8 minutes, then say:

STOP. TURN TO PAGE 29. YOU WILL HAVE 2 MINUTES AND 30 SECONDS TO READ AND LEARN THE LEVEL 3 DIRECTIONS, PRESENTED ON PAGE 29. THERE ARE NO PRACTICE ITEMS FOR THIS LEVEL. BEGIN READING NOW.

Time 2 minutes and 30 seconds, then say:

STOP. TURN TO PAGE 31. BEFORE PROCEEDING WITH THE FINAL PORTION OF THE TEST, YOU WILL HAVE 30 SECONDS TO REVIEW THE LEVEL 3 RULES, SUMMARIZED FOR YOU ON PAGE 31. BEGIN NOW.

Time 30 seconds, then say:

STOP. PLEASE PAY CLOSE ATTENTION TO MY LAST SET OF INSTRUCTIONS. YOU WILL HAVE 10 MINUTES TO ANSWER THE TWELVE LEVEL 3 ITEMS.

WORK AS RAPIDLY AS POSSIBLE WITHOUT SACRIFICING ACCURACY. MAKE SURE THE ANSWER YOU SELECT CORRESPONDS TO THE CORRECT QUESTION NUMBER ON THE ANSWER SHEET. IF YOU FINISH THE CHART ITEMS, CHECK YOUR ANSWERS FOR THE LEVEL THREE ITEMS ONLY. DO NOT TURN BACK TO THE INSTRUCTIONS, OR ANY PREVIOUS ITEMS. ARE THERE ANY QUESTIONS?

YOU HAVE 10 MINUTES TO COMPLETE TWELVE ITEMS. TURN TO PAGE 32. BEGIN NOW.

Time 10 minutes, then say:

STOP. PUT DOWN YOUR PENCILS AND CLOSE YOUR TEST BOOKLET. YOU HAVE NOW COMPLETED THE FIRST PORTION OF THE TESTING PROCESS. YOU ARE REMINDED THAT YOU ARE NOT PERMITTED TO DISCUSS THE CONTENT OF THIS TEST OUTSIDE OF THIS ROOM. ARE THERE ANY QUESTIONS?

Answer any questions, and then say:

BEFORE WE CONTINUE WITH THE SECOND PART OF THE TESTING PROCESS, I AM INTERESTED IN YOUR ATTITUDES TOWARD THE FIRST PART OF THE TESTING PROCESS.

PLEASE GO TO THE NEXT PAGE AFTER THE ANSWER SHEET AND COMPLETE THE 39-ITEM QUESTIONNAIRE LABELED "SPJS". WHILE YOU COMPLETE THIS QUESTIONNAIRE, PLEASE KEEP IN MIND THAT I AM USING YOUR SCORE ON THE FIRST CONGITIVE ABILITY TEST BATTERY AS AN INDICATOR OF YOUR SUITABILITY TO BECOME A NAVAL OFFICER.

ONCE YOU HAVE COMPLETED THIS QUESTIONNAIRE, PUT YOUR PENCIL DOWN AND WAIT FOR FURTHER INSTRUCTIONS. BEGIN NOW.

Wait until everyone has completed the questionnaire, and then say:

GO TO THE NEXT PAGE. YOU WILL NOW COMPLETE TWO MORE QUESTIONNAIRES LABELED "VIEMS"AND "MCSDS". PLEASE RESPOND HONESTLY. ONCE YOU HAVE COMPLETED THESE QUESTIONNAIRES PUT DOWN YOUR PENCILS, GO TO THE PAGE LABELED "EXPERIMENT NAVY CANADIAN FORCES", AND WAIT FOR FURTHER INSTRUCTIONS.

Wait until everyone has completed the questionnaire, and then say:

FOR THE NEXT PART OF THE TESTING PROCESS YOU ARE TO TAKE A SERIES OF FOUR SUB-TESTS FROM THE KIT OF COGNITIVE FACTOR-REFERENCED TESTS. THESE TESTS WILL BE USED TO HELP DETERMINE YOUR SUITABILITY TO BECOME A NAVAL OFFICER IN THE CANADIAN FORCES. IT WILL TAKE ABOUT ONE HOUR TO COMPLETE ALL FOUR TESTS.

GO TO PAGE 1 IN THE SECOND TEST BOOKLET. READ THE DIRECTIONS TO YOURSELF WHILE I READ THEM ALOUD.

Pause, then say:

IN THIS TEST YOU WILL BE ASKED TO SOLVE SOME PROBLEMS IN MATHEMATICS. THERE ARE FIVE ANSWER CHOICES FOR EACH PROBLEM. FOR EACH QUESTION BE SURE TO PICK THE BEST ONE OF THE POSSIBLE ANSWERS LISTED. WHEN YOU HAVE DECIDED WHICH ONE OF THE CHOICES GIVEN IS THE BEST ANSWER TO A PROBLEM, RECORD YOUR ANSWER ON YOUR SEPARATE ANSWER SHEET. <u>PLEASE WRITE DOWN THE LAST FIVE DIGITS OF YOUR STUDENT IDENTIFICATION NUMBER ON TOP OF YOUR ANSWER SHEET.</u>

EXAMPLE: HOW MANY CANDY MINTS CAN YOU BUY FOR 50 CENTS AT THE RATE OF 2 FOR 5 CENTS? (1) 10; (2) 20; (3) 25); (4) 100; 0R (5) 125.

THE CORRECT ANSWER TO THIS PROBLEM IS 20. THEREFORE, YOU SHOULD HAVE BLACKENED NUMBER 2 BESIDE THE WORD "EXAMPLE" ON YOUR

ANSWER SHEET.

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 10 MINUTES FOR EACH OF THE TWO PARTS OF THIS TEST. EACH PART HAS 3 PAGES WITH 15 ITEMS. WHEN YOU HAVE FINISHED PART 1, STOP. PLEASE DO NOT GO ON TO PART 2 UNTIL YOU ARE ASKED TO DO SO. ARE THERE ANY QUESTIONS?

Answer any questions, and then say:

GO TO THE NEXT PAGE AND BEGIN.

Time 10 minutes, then say:

STOP. YOU HAVE 10 MINUTES TO COMPLETE THE SECOND PART OF THIS TEST. IF YOU FINISH BEFORE THE TIME IS UP, CHECK YOUR ANSWERS FOR THE PART 2 ITEMS ONLY. DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO.

NOW GO TO THE NEXT PAGE IN YOUR TEST BOOKLET AND WORK ON PART 2.

Time 10 minutes, then say:

STOP. PUT DOWN YOUR PENCILS, AND GO TO THE CALENDAR TEST IN YOUR OTHER TEST BOOKLET (show them). READ THE DIRECTIONS TO YOURSELF WHILE I READ THEM ALOUD.

THIS IS A TEST OF YOUR ACCURACY IN FOLLOWING DIRECTIONS. EACH DIRECTION WILL ASK YOU TO FIND A DATE ON A CALENDAR, WHICH IS PRINTED ON THE LAST PAGE OF THIS TEST, THAT IS, THREE PAGES FROM

THIS ONE.

IN THIS CALENDAR YOU ARE TO REMEMBER THAT: (1) A CIRCLED NUMBER IS A HOLIDAY; (2) SATURDAYS AND SUNDAYS ARE WEEKEND DAYS; (3) ALL DAYS EXCEPT HOLIDAYS AND WEEKENDS ARE WORK DAYS; (4) THE FIRST DAY OF SPRING IS MARCH 21; (5) THE FIRST DAY OF SUMMER IS JUNE 21; (6) THE FIRST DAY OF FALL IS SEPTEMBER 21; AND (7) THE FIRST DAY OF WINTER IS DECEMBER 21.

LOOK AT THE SAMPLE ITEMS BELOW. PUT AN "X" ON THE LETTER IN FRONT OF THE CORRECT ANSWER. 1-WHAT IS THE THIRD TUESDAY OF THE MONTH? THE CORRECT ANSWER IS "A". 2-WHAT IS THE THIRD WORKING DAY AFTER THE HOLLIDAY? THE CORRECT ANSWER IS "D". FINALLY, WHAT IS THE SEVENTH WORKING DAY AFTER THE THIRD MONDAY OF THE MONTH? THE CORRECT ANSWER IS "D".

YOUR SCORE WILL BE THE NUMBER OF DATES YOU MARK CORRECTLY <u>MINUS</u> A FRACTION OF THOSE MARKED INCORRECTLY. THEREFORE, IT WILL <u>NOT</u> BE TO YOUR ADVANTAGE TO GUESS UNLESS YOU HAVE SOME IDEA ABOUT WHICH DATE IS CORRECT.

THIS TEST HAS TWO PARTS. EACH PART HAS 10 DATES FOR YOU TO SELECT. YOU WILL HAVE <u>7 MINUTES</u> FOR EACH PART. WHEN YOU HAVE FINISHED PART 1, STOP. PLEASE DO NOT GO TO PART 2 UNTIL ASKED TO DO SO.

TEAR OFF THE LAST PAGE OF THIS TEST NOW SO YOU WILL BE ABLE TO REFER TO THAT CALENDAR EASILY AS YOU TAKE THE REST OF THIS TEST. DO YOU HAVE ANY QUESTIONS?

Answer any questions, and then say:

GO TO THE NEXT PAGE AND BEGIN.

Time 7 minutes, then say:

STOP. PUT DOWN YOUR PENCILS. YOU HAVE 7 MINUTES TO COMPLETE THE SECOND PART OF THIS TEST. IF YOU FINISH BEFORE THE TIME IS UP, CHECK YOUR ANSWERS FOR THE PART 2 ITEMS ONLY. DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO.

NOW GO TO THE NEXT PAGE AND BEGIN.

Time 7 minutes, then say:

STOP PUT DOWN YOUR PENCILS AND GO TO THE NEXT PAGE IN YOUR TEST BOOKLET. READ THE DIRECTIONS TO YOURSELF WHILE I READ THEM ALOUD.

THIS IS A TEST OF YOUR ABILITY TO REMEMBER SERIES OF NUMBERS. THE EXAMINER WILL CALL OUT THE NUMBERS. AFTER HE/SHE FINISHES, YOU ARE TO WRITE DOWN THE NUMBERS IN THE EXACT ORDER IN WHICH THEY WERE CALLED OUT. PLEASE DO NOT WRITE ANY NUMBERS UNTIL THE EXAMINER HAS FINISHED THE WHOLE SERIES.

SOME OF THE SERIES WILL BE TOO LONG FOR YOU TO REMEMBER ALL OF THE NUMBERS. IF YOU DO NOT REMEMBER SOME OF THEM, LEAVE A BLANCK SPACE FOR THEM AND WRITE DOWN ALL THE NUMBERS YOU DO REMEMBER. TRY TO REMEMBER ALL THE NUMBERS IF POSSIBLE, AND BE SURE TO WRITE THEM DOWN IN THE EXACT ORDER IN WHICH THEY WERE CALLED OUT.

FOR EXAMPLE, THE EXAMINER MIGHT CALL OUT, "SERIES ONE. 7-2-4 BEGIN." WHEN HE OR SHE SAYS "BEGIN", WRITE THE NUMBERS ON THE ANSWER PAGE IN THIS MANNER: 7-2-4.

IT IS VERY IMPORTANT THAT YOU DO NOT WRITE NUMBERS WHILE A SERIES IS BEING CALLED OUT, SINCE THIS IS A TEST OF YOUR MEMORY FOR NUMBERS. DO YOU HAVE ANY QUESTIONS?

Answer any questions, and then say:

GO TO THE NEXT PAGE, AND LISTEN CAREFULLY.

Call out the numbers, one number per second. Once you have finished calling out the numbers, then say:

PUT DOWN YOUR PENCILS. GO TO THE NEXT PAGE IN YOUR TEST BOOKLET. READ THE DIRECTIONS TO YOURSELF WHILE I READ THEM ALOUD.

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 QUARES 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 <u>ONE AND ONLY ONE</u> 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.

Pause for 90 seconds, then say:

THE ASWER 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; AND 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 YOU HAVE ANY QUESTIONS?

Answer any questions, then say:

GO TO THE NEXT PAGE AND BEGIN.

Time 3 minutes, then say:

STOP. PUT DOWN YOUR PENCILS. YOU HAVE 3 MINUTES TO COMPLETE THE SECOND PART OF THIS TEST. IF YOU FINISH BEFORE THE TIME IS UP, CHECK YOUR ANSWERS FOR THE PART 2 ITEMS ONLY. DO NOT GO BACK TO PART 1 AND DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO.

NOW GO TO THE NEXT PAGE AND BEGIN.

Time 3 minutes, then say:

STOP. PUT DOWN YOUR PENCILS AND GO TO THE NEXT PAGE. I AM NOW INTERESTED IN YOUR ATTITUDES TOWARD THE SECOND PART OF THE TESTING PROCESS.

PLEASE GO TO THE NEXT PAGE AND COMPLETE THE 39-ITEM QUESTIONNAIRE LABELED "SPJS". WHILE YOU COMPLETE THIS

QUESTIONNAIRE, PLEASE KEEP IN MIND THAT I AM USING YOU TOTAL SCORE ON THE SECOND CONGITIVE ABILITY TEST BATTERY AS AN INDICATOR OF YOUR SUITABILITY TO BECOME A NAVAL OFFICER.

ONCE YOU HAVE COMPLETED THIS QUESTIONNAIRE, PUT YOUR PENCIL DOWN AND WAIT FOR FURTHER INSTRUCTIONS. BEGIN NOW.

Wait until everyone has completed the questionnaire, and then say:

GO TO THE NEXT PAGE. YOU WILL NOW COMPLETE ONE MORE QUESTIONNAIRES LABELED "VIEMS". PLEASE RESPOND HONESTLY. ONCE YOU HAVE COMPLETED ALL QUESTIONNAIRES BRING ME BACK THE TESTING MATERIALS, AND SIGN THE REGISTER TO CONFIRM THAT I GAVE YOU $10. YOU CAN ALSO PICK UP A COPY OF THE CONSENT FORM. THE BACK OF THE CONSENT FORM CONTAINS ADDITIONAL INFORMATION CONCERNING THIS EXPERIMENT. THANK YOU VERY MUCH FOR YOUR TIME.

Appendix C

Informed Consent Form

Experiment Navy
Sébastien Blanc
Department of Psychology
Saint Mary's University
Halifax, NS, B3H 3C3
Tel: 902-420-5946 Fax: 902-496-8287
e-mail: Lorraine.Huston@STMARYS.CA

I am graduate student in the Department of Psychology at Saint Mary's University. As part of my Masters Thesis, I am conducting research under the supervision of Dr. Catano. I am inviting you to participate in my study. The purpose of the study is to examine the psychometric characteristics of the Maritime Officer Selection Test (MOST) as well as your reactions toward the testing process.

This study involves the completion of five cognitive ability tests and five questionnaires (three of these five questionnaires are administered twice). The total time required for the completion of these tests and questionnaires is approximately 2.5 hours.

There are no risks involved in this study. **Your participation is completely voluntary. You may withdraw from this study at any time without penalty**. You will receive $10 for 2.5 hours of testing, or portion thereof, for the time spent in the study. The top 20% will receive an additional $50 as a reward for their excellent performance.

All information obtained in this study will be kept strictly **confidential and anonymous.** The tests and questionnaires will be numerically coded using the last five digits of your Student Identification Number. This identification number will be needed for the additional payment of above average test performance. Please do not put any identifying information on any of the forms. To further protect individual identities, this consent form will be sealed in an envelope and stored separately. Furthermore, the results of this study will be presented as a group and no individual participants will be identified.

If you have any questions, please contact the principal researcher, Sébastien Blanc, at 902-431-6051 (sebastienblanc@hotmail.com).

This research has been reviewed and approved by the Saint Mary's University Research Ethics Board. If you have any questions or concerns about the study, you may contact Dr. John MacKinnon at ethics@stmarys.ca, Chair, Research Ethics Board.
By signing this consent form, you are indicating that you fully understand the above information and agree to participate in this study. In addition, you agree not to discuss the nature of the Maritime Officer Selection Test (MOST) with anyone else.

Participant's Signature: ______________________ Date: ____________________

Please keep one copy of this form for your own records.

Appendix D

A Career as an Officer in Maritime Surface and Sub-surface (MARS 71)[1]

What They Do

The primary function of the Maritime Surface and Sub-surface military occupation is to provide officers to man the seagoing combatant units of MARCOM. The primary task of officers within the military occupation is command, co-ordination, and control of Military Maritime Operations. To do this, you must be able to lead and make decisions, often under adverse conditions of physical discomfort and mental stress. Furthermore, you will be required to gain knowledge and expertise in a wide spectrum of activities relating to the exercise of sea power, including maritime strategy, tactics and procedures in the operation of ships, submarines and aircraft, maritime sensors, combat information, and weapons systems. You may also be called upon to provide an input into the design, the procurement, and the evaluation of ships or systems. In addition to your primary tasks as an officer to the MARS military occupation, you will be required to perform staff, training and administrative duties which require this background.

Qualification Requirements

The minimum education required is an equivalent to a high school leaving diploma with strong emphasis in selection being placed on good standing in English, Mathematics and Science. You will be expected to meet Canadian Forces medical standards and go through a selection process, which includes tests and interviews and an acceptance board. The entry plans that provide access to the MARS military occupation are described below. However, for complete information, you should contact your Military Career Counsellor at 1-800-856-8488.

Plans for Entry

Regular Officer Training Plan (ROTP) - This plan involves first completing a university education under government sponsorship prior to beginning full-time employment as a Maritime Surface and Sub-surface Officer. For this plan you should have attained, or be in the process of attaining, a high school leaving diploma with university-oriented credits.

Direct Entry Officer (DEO) - to apply for direct entry as an officer, you must have a university degree.

Training

Basic Officer Training (PHASE 1) - Upon selection into the Maritime Surface and Sub-surface occupation, you will be enrolled in the Canadian Forces and then proceed to the Canadian Forces Officer Candidate School at CFB Saint-Jean, Quebec, for a seven-week initial Basic Officer Training Course. At Saint-Jean, you will be introduced to life in the Canadian Forces. You will learn military regulations and customs as well as leadership techniques, and acquire the fundamental military skills of drill, dress, deportment, weapon handling and first aid. You will also

---

[1] Adapted from National Defence. (n.d.). A Career as an Officer in Maritime Surface and Sub-surface (MARS 71). Retrieved December 10, 2002, from http://www.recruiting.dnd.ca/html/navy/careers/career_profiles/mari_surf.html.

participate in physical training and sports programmes. The Basic Officer Training Course is given in either English or French and successful completion is a prerequisite for further training. For those enrolled under the DEO entry plan, commissioning in the rank of Second-Lieutenant (Acting Sub-Lieutenant) follows. You will then attend Basic Officer Training Part II and second language training, which could be from two to seven months, depending on the occupation selected and the entry plan.

PHASE II - The training that you will undergo is intense, varied, and spans some 27 months after which you will be competent to stand your own watch on the bridge of an operational ship at sea. You will spend 12 months at the Naval Officer Training Centre in Esquimalt, British Columbia, after which you will be awarded your Certificate of competency Level I. The remaining time will be spent either in classrooms, or in an Operational Destroyer. Your training will consist of learning all aspects of navigation, bridgemanship, communications, relative motion, ship safety, emergency procedures, and rules of seamanship. From there, you will join the fleet and continue on with the Naval Operations Course, which is oriented towards operations and tactics. It will include topics of study, such as tactics, communications, and helicopter operations and procedures, which will allow you, as a MARS Officer, to gain an appreciation for a warship combat role. While on this course, you will also received detailed instructions in military law, general service knowledge, personnel administration and financial administration. Upon successful completion of the Naval Operations Course and a period of approximately 12 months in an operational ship, you will be awarded your Bridge Watchkeeping Certificate and the Certificate of Competency Level II.

Working Environment

The working environment of a MARS Officer can vary in a short span of time from leisure activities to prolonged periods of physically and mentally demanding duties. Regardless of the activity at any given time, you are, as a Naval officer, continually at the mercy of the sea. Seasickness can be a problem, however, it is usually temporary, and chronic seasickness (e.g., continuous sickness, even under moderate seagoing conditions) is quite rare. You very soon gain your "sea legs", and your newly chosen environment becomes second nature to you. You must be ready to work long hours while at sea, to live in small quarters aboard ships, and to be separated from your family during your sea tours. As a junior MARS Officer, you may expect to remain on your first ship for another 20 months during which time you will receive further specialized training as a Destroyer Navigation Officer, Above Water Weapons Director, Under Water Weapons Director, Destroyer Antisubmarine Warfare Officer, Air Control Officer and/or Ships Diving Officer. As your process, you may receive the Combat Control Officer Course, and eventually hold the position of Department Head in a destroyer. If you are selected to serve in submarines, you will be sent on a Basic Submarine Training course. This course is carried out after completing your General Service Bridge Watchkeeping Certificate and Certificate of Competency Level II. It is, therefore, a very intense programme which, when you are qualified, will allow you to wear the Submarine Badge, which is a specialist skill badge also referred to as "Dolphins".

Related Civilian Occupations

Some of the related civilian occupations encompass positions and responsibilities in various seagoing vessels from fishing vessels and merchant ships, to Coast Guard vessels and passenger liners. These positions (for example, Mate, Master, Captain) depend on the level of MARS qualifications held, on whether the vessel is of the inland type or ocean going, and also on the vessel's specific or design role. Many civilian companies view Canadian Naval Training with respect. The intensity, quality and completeness of training ensure ideal employees who can easily adapt to any seagoing vessel.

Appendix E

SPSS Syntax for DIF with Logistic Regression[1]

```
* SPSS SYNTAX written by: .
* Bruno D. Zumbo, PhD  .
* Professor of Psychology and Mathematics,  .
* University of Northern British Columbia  .
* e-mail:

* Instructions .
* Change the filename, currently 'binary.sav' to your file name.
* Change 'item', 'total', and 'grp', to the corresponding variables in your file.
* Run this entire syntax command file.

compute item= item1.
compute total= scale.
compute grp= group.

*   Aggregation.
*   Working with the Centered data.

* Hierarchical regressions approach with the following order of steps:.
*     1. total.
*     2. total + group.
*     3. total + group + interac.
* This also, of course, allows one to compute the relative Pratt Indices.


* Saves the standardized versions of group and total with the.
* eventual goal of centering before computing the cross-product term.

DESCRIPTIVES
 VARIABLES=group total  /SAVE
 /FORMAT=LABELS NOINDEX
 /STATISTICS=MEAN STDDEV MIN MAX
 /SORT=MEAN (A).

* Allows for both uniform and non-uniform DIF.
* Provides the 2df Chi-square test for DIF.
```

---

[1] From *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistical Regression Modelling as a Unitary Framework for Binary and Likert Type (Ordinal) Item Scores*, by B. D. Zumbo, 1999, Ottawa, ON: Director Human Resources Research and Evaluation. Copyright 1999 by Her Majesty the Queen in Right of Canada. Reprinted with permission.

```
LOGISTIC REGRESSION item
 /METHOD=ENTER ztotal /method=enter zgroup ztotal*zgroup
 /SAVE PRED(pre1).
execute.

* The following command is required to deal with the repeaters in.
* the data. The WLS regression will be conducted on the aggregate.
* file entitled "AGGR.SAV".

AGGREGATE
 /OUTFILE='aggr.sav'
 /BREAK=zgroup ztotal
 /item = SUM(item) /pre1 = MEAN(pre1)
 /Ni=N.

GET
 FILE='aggr.sav'.
EXECUTE .

compute interact=zgroup*ztotal.
execute.

COMPUTE v1 = Ni*pre1 *(1 - pre1) .
EXECUTE .

COMPUTE z1 = LN(pre1/(1-pre1))+ (item-Ni*pre1)/Ni/pre1/(1-pre1) .
EXECUTE .

FORMATS v1, z1 (F8.4).
execute.

* Overall logistic regression.
* Both Uniform and Non-uniform DIF.
REGRESSION
 /MISSING LISTWISE
 /REGWGT=v1
 /descriptives=corr
 /STATISTICS COEFF OUTS R ANOVA COLLIN TOL CHA
 /NOORIGIN
 /DEPENDENT z1
 /METHOD=ENTER ztotal / method=enter zgroup / method= enter interact .
execute.
```

Sébastien Blanc

Tel: (
>Fro
>To:
>Subject: RE: Authorization Request
>Date: Fri, 13 Jun 2003 14:42:39 -0400
>
>Capt Blanc,
>
>Approved.
>
>
>DC Scholtz
>Maj
>D Mar Pers 2-5

>
>-----Original Message-----
>From: Sébastien Blanc ı]
>Sent: Friday, 13 June, 2003 13:20
:
>Subject: Authorization Request
>
>
>
>Maj Scholtz,
>
>Please confirm if you authorize me to copy one practice chart from the MOST
>(i.e., the grid route map on p. 4) and provide a sample of a route selection
>rule (i.e., If two or more routes are the same length...). This information
>would be presented in the Method section of my thesis and would make my
>general description of the MOST much easier to understand.
>
>Regards,
>
>Capt. Sébastien Blanc, CHRP
>M.Sc. Graduate Student Saint-Mary's University
>Halifa
>Tel:

Sébastien Blanc

>Date: Fri, 18 Jul 2003 11:39:43 -0400
>
>Sebastion: you are authorized to include Zumbo's table's and syntax in your
>thesis.
>
>This email authorizes Sebastien Blanc to reprint Figures 1, 2 and 3 and spss
>syntax (from A Handbook on the Theory and Methods of Differential Item
>Functioning (DIF): Logistical Regression Modelling as a Unitary Framework
>for Binary and Likert Type (Ordinal) Item Scores, by B. D. Zumbo, 1999,
>Ottawa, ON: Director Human Resources Research and Evaluation) for use in his
>Master's Thesis.
>
>
>
>R.A. Boswell LCol

>
>Intellegere

Saint Mary's University

Halifax, Nova Scotia
Canada B3H 3C3

Research Ethics Board
tel
fax
e-ma
web

Saint Mary's University

Certificate of Ethical Acceptability of Research Involving Human Subjects

This is to certify that the Research Ethics Board has examined the research proposal or other type of study submitted by:

Principal Investigator:	Sebastien Blanc
Name of Research Project:	Naval Officer Selection in Canada: An Evaluation of the Maritime Officer Selection Test (MOST)
REB File Number:	2002-104

and concludes that in all respects the proposed project meets appropriate standards of ethical acceptability and is in accordance with the Tri-Council Policy Statement on the Conduct of Research Involving Humans. Please note that approval is only effective for one year from the date approved. (If your research project takes longer than one year to complete, submit form #3 to the REB at the end of the year and request an extension.)

Date: Nov. 27, 2002

Signature of REB Chair:

Dr. John E. MacKinnon