

Examining Coaching and Retest Effects on Aptitude Tests

by

Lauren Florko

A Thesis Submitted to  
Saint Mary's University, Halifax, Nova Scotia  
In Partial Fulfillment of the Requirements for  
The Degree of Master of Science in Applied Psychology  
(Industrial/Organizational)

August 31, 2010, Halifax, Nova Scotia

© Lauren Florko, 2010

Approved: \_\_\_\_\_  
Arla Day, Ph.D.  
Supervisor

Approved: \_\_\_\_\_  
Victor Catano, Ph.D.  
Committee Member

Approved: \_\_\_\_\_  
James Cameron, Ph.D.  
Committee Member

Approved: \_\_\_\_\_  
Damian O'Keefe, Ph.D.  
External Examiner

Date: August 31, 2010



Library and Archives  
Canada

Published Heritage  
Branch

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque et  
Archives Canada

Direction du  
Patrimoine de l'édition

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
ISBN: 978-0-494-68827-4  
*Our file* *Notre référence*  
ISBN: 978-0-494-68827-4

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

## Table of Contents

List of Appendices .....	iii
List of Tables .....	iv
List of Figures .....	v
Acknowledgments.....	vi
Abstract .....	1
Examining Coaching and Retest Effects on Aptitude Tests .....	2
Cognitive Tests .....	3
Aptitude Tests .....	3
Coaching Effects.....	4
Retesting Effects .....	6
Summary and Hypotheses .....	10
Method .....	13
Sample .....	13
Procedure .....	14
Measures .....	16
Results.....	20
Preliminary Analyses.....	20
Analyses.....	22
Discussion .....	33
Practical Implications .....	37
Limitations and Future Research.....	38
Conclusion .....	43
References.....	44

List of Appendices

Appendix A: Saint Mary’s University Research Ethics Board Approval.....	50
Appendix B: Recruitment Tools .....	51
Appendix C: Socioeconomic Status and Demographic Questions .....	54
Appendix D: Mini Royal Canadian Mounted Police Aptitude Test .....	55
Appendix E: Test-Taking Anxiety Scale, Test-Taking Stress Scale, and Test-Taking Familiarity Scale .....	65
Appendix F: Test-Taking Motivation Scale .....	66
Appendix G: Coaching Materials .....	67
Appendix H: Mini International Personality Item Pool.....	81
Appendix I: Coaching Effects Questions.....	84

## List of Tables

Table 1: Means, standard deviations, and intercorrelations among the study variables.....	23
Table 2: Means, standard deviations, and reliability coefficients among the study variables by condition group.....	24
Table 3: Intercorrelations among the study variables split by condition group.....	25
Table 4: Mean score differences over consecutive tests split by condition group.....	26
Table 5: Source table for the repeated measures ANOVA on the Wonderlic .....	27
Table 6: Source table for the repeated measures ANOVA on the RPAT .....	28
Table 7: Results of regression analysis of test preparation on the second administration of aptitude tests.....	30
Table 8: Results of hierarchical regression analysis of the relationships between retesting effects and aptitude test scores.....	32
Table 9: Differences among participants who dropped out or completed the study.....	42

List of Figures

Figure 1: Different types of changes in aptitude test scores upon retesting .....7

Figure 2: Measures used within each study session.....15

### Acknowledgments

I would like to thank my supervisor, Dr. Arla Day, for continually challenging and pushing me to advance my KSAOs. I would also like to thank my committee member, Dr. Vic Catano, for providing me with support and hope. I would like to thank Dr. Jim Cameron for assisting me with any ethical concerns and any theoretical discrepancies. Additionally, I would like to thank my external examiner, Dr. Damian O’Keefe, for challenging me with ‘hard’ questions and taking an interest in my thesis.

I would like to thank my parents (all three of them) to keeping me motivated and see the light at the end of the tunnel. They kept saying, “Get a job!” and I knew it meant, “Keep up the good work!” Without their support I would be eating instant noodles every night and couldn’t afford to have any work-life balance. I want to thank my brother for complaining about me getting more degrees as he felt it “made him look bad”. Needless to say I’m starting the Ph.D. program. I want to thank my friends and classmates, particularly Meghan Donohoe and Matt MacPhee, who helped me work out my concerns and, more importantly, keep my sanity.

## Examining Coaching and Retest Effects on Aptitude Tests

By Lauren Florko

### Abstract

Cognitive tests are commonly used in employment settings because they are strong predictors of job performance. Although cognitive tests are beneficial in selection settings, they have potential drawbacks. Test scores may artificially increase from retesting and from studying test materials and aids beforehand (coaching effects). Studies have examined these effects, but have yet to investigate why these effects exist. The purpose of this study was to conduct an experiment to examine coaching and retesting effects on two aptitude tests. Thirty-nine participants completed two testing sessions composed of these aptitude tests and questions about general test-taking attitudes. The combined total effect of coaching and retesting led to a significant increase in aptitude test scores. However, test-taking anxiety, test-taking stress, test-taking familiarity, and test-taking motivation did not uniquely contribute to these effects. Future research should examine other possible explanations of why these effects exist.

Date August 31, 2010

### Examining Coaching and Retest Effects on Aptitude Tests

Tests of cognitive ability, or general mental ability, are frequently used by organizations to select the best applicant for the job (Catano, Wiesner, Hackett, & Methot, 2005; Hunter & Hunter, 1984; Schmidt & Hunter, 1998). These tests also are used in police selection to not only predict successful job performance, but also successful job training (Hunter & Hunter, 1984). Cognitive tests can be used independently, but are more likely used in combination with other selection methods (e.g., interviews, reference checks). Their popularity is primarily due to the fact that they can strongly predict future job performance (Cohen, Swedlik, & Phillips, 1996; Goldstein & Hersen, 1990; Hunter & Hunter, 1984). Moreover, cognitive tests allow employers a standardized selection method and a means to compare applicants objectively (Catano et al., 2005).

Applicants may try to increase their cognitive test scores by obtaining test instruction and/or by preparing for the test (Kulik, Bangert-Drowns, & Kulik, 1984). Unfortunately, these actions may lead to inflated test scores, which can affect the organization's ability to accurately distinguish potential successful from non-successful job candidates. Additionally, many applicants are then given the opportunity to retake cognitive tests (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007), which can lead to greater score increases (Hausknecht et al., 2007). These score increases may be the result of legitimate ability growth, but they also may be due to construct-irrelevant (i.e., non-ability) changes (Hausknecht, Trevor & Farr, 2002; Lievens, Reeve & Heggstad, 2007). The goal of this study is to examine the effects of test instruction and

preparation on one type of cognitive test (i.e., aptitude tests) and to examine possible sources for test-score increases upon retesting on such tests.

### **Cognitive Tests**

Cognitive tests measure general mental ability or general intelligence (Murphy & Davidshofer, 1998; Schmidt & Hunter, 1998). These tests assess individual variations in mental performances such as manipulation, retrieval, evaluation, and processing of information (Murphy & Davidshofer, 1998). Cognitive tests use this assessment of performance to establish a level of ability or development (Cohen et al., 1996). There are many different types of cognitive tests, such as achievement, aptitude, development, and intelligence tests. Aptitude tests are particularly popular in selection settings given their predictive quality (Catano et al., 2005), and therefore, they will be the focus of this study.

### **Aptitude Tests**

Aptitude tests assess informal learning or life experience—tapping into an innate knowledge set (Cohen et al., 1996). In other words, aptitude tests are designed to measure cumulative knowledge, skills, and abilities developed over an individual's life experience (Murphy & Davidshofer, 1998). They are used primarily for prognostic purposes to make predictions of future performance (Cohen et al., 1996; Goldstein & Hersen, 1990); this is unlike achievement tests, which measure specific accomplishments or past performance (such as a final exam taken at the end of a course). For example, an applicant for a law enforcement position might write an aptitude test to allow the organization to assess the applicant's general mental ability to infer potential training success and potential job performance.

Aptitude tests, among other cognitive tests, have been shown to be of great value in applied settings—such as selection, classification, placement, and diagnostic decisions (Cohen et al., 1996; Kuncel, Hezlett, & Ones, 2001). They are particularly strong predictors of job performance (Hunter & Hunter, 1984; Schmidt & Hunter, 1998). Multiple studies have found that in comparison to eight other selection methods (e.g., interviews, references checks, education, work experience), cognitive tests were superior in terms of having higher validity, lower adverse impact, and greater feasibility (Goldstein & Hersen, 1990; Hunter & Hunter, 1984). That is, aptitude tests are stronger predictors of job performance, are less biased towards demographic groups (e.g., visible minorities), and are easier to administer and score. In Hunter and Hunter (1984)'s meta-analysis, aptitude tests had a validity coefficient of  $r = .28$  when predicting job performance, suggesting that aptitude tests can explain approximately 8% of the variance in job performance. Employees who are selected through the use of an aptitude or other cognitive test have been found to be almost half a standard deviation higher on job performance than employees who were not selected with one (Schmidt, Hunter, McKenzie, & Muldrow, 1979). When organizations hire employees using a cognitive test, they can predict future productivity levels. This predictive ability can then amount to as great as 20% of total corporate profits (Hunter & Hunter, 1984). Although there are many advantages to using an aptitude test in selection settings, their use is not without drawbacks.

### **Coaching Effects**

One of the main drawbacks to using an aptitude test is that it can be susceptible to artificially inflated scores. Artificially inflated scores can occur when applicants attempt

to improve their score on aptitude tests during the selection process by various means of preparation, including the use of books, courses, and practice questions. When test scores increase due to the use of test instruction, preparation, and/or feedback, the result is termed a “coaching effect” (Hausknecht et al., 2007, p. 374). However, given that aptitude tests are designed to assess innate knowledge (Cohen et al., 1996), studying should not significantly impact the applicant’s results. For example, two applicants who have the same level of general mental ability should obtain the same aptitude test score regardless of whether one prepared in advance. Additionally, aptitude appears to exist regardless of the setting for which it was developed; therefore, having a particular level of aptitude from previous education should be equivalent to workplace aptitude (Kuncel et al., 2001). Consequently, studying as a means to become familiar with a different “type” of aptitude shouldn’t be necessary. Thus, in theory, coaching shouldn’t impact test scores significantly.

Similar levels of coaching effects have been found across many occupations, including law enforcement (Hausknecht et al., 2002). In Kulik, Bangert-Drowns, and Kulik’s (1984) meta-analysis, coaching effects on aptitude tests resulted in score increases of .40 standard deviations. Kulik, Bangert-Drowns, and Kulik’s (1984) meta-analysis only included experiments with pre- and post-test designs, making it highly unlikely that these effects existed by chance. Score increases of .40 standard deviations can have a large impact on raw score changes (Hausknecht et al., 2002) and therefore on selection decisions. In addition to coaching effects, test scores also may artificially increase when applicant are allowed to retake an aptitude test.

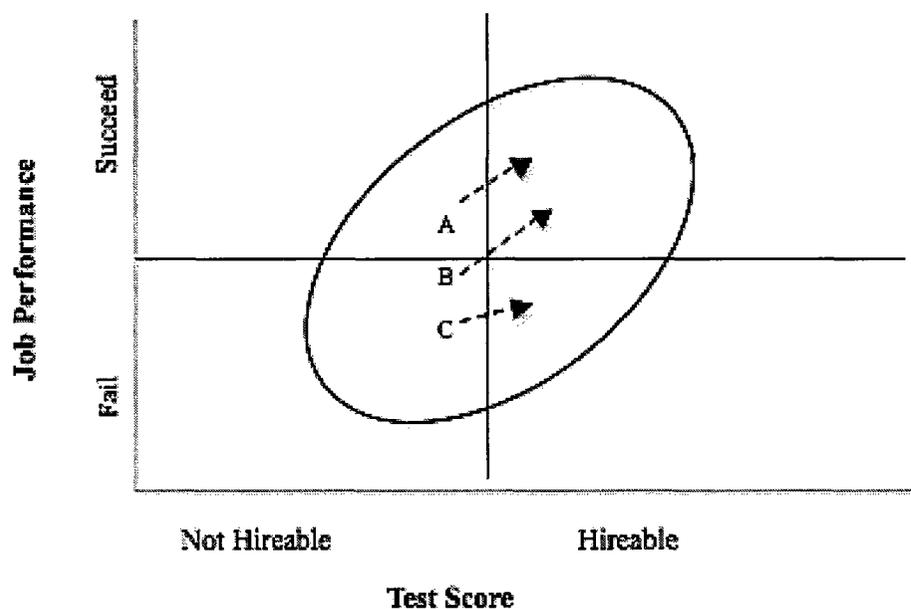
### Retesting Effects

In many employment settings, applicants are allowed to re-take an aptitude test if they fail. In fact, 25% to 50% of employees are retested on cognitive ability tests (Hausknecht et al., 2007). Employers usually have two rationales for retesting: (1) the initial assessment was inaccurate from individual issues or random measurement error (2) or the candidates have improved and should be allowed to demonstrate their new level of cognitive ability (Lievens, Buyse, & Sackett, 2005). When applicant scores increase on subsequent cognitive tests, it is called a “practice effect” or a “retesting effect” (Hausknecht et al., 2007). Retesting effects have been suggested to arise from a plethora of sources, such as: actual development of abilities, coaching effects, regression towards the mean, and debilitating construct-irrelevant factors (e.g., test-taking anxiety; Hausknecht et al., 2007; Lievens et al., 2007).

In a meta-analysis of coaching and retesting effects on cognitive ability tests, Hausknecht and colleagues (2007) found that retesting effects alone led to increases in test scores of approximately one-quarter of a standard deviation. Hausknecht and colleagues (2007) also calculated the approximate regression toward the mean, but it only accounted for .04 of a standard deviation. Kulik, Kulik, and Bangert (1984) found that when identical cognitive tests were used in retesting, scores increased by .42 standard deviations. And even when equivalent forms (i.e., identical tests, but not identical items) were used, scores still increased by .23 standard deviations (Hausknecht et al., 2002). Upon a third test administration, the difference between the first and third aptitude test was found to be a 0.76 standard deviation difference (Hausknecht et al., 2002).

In examining why retesting effects exist, research indicates that they may arise from three main sources: (1) increases in actual ability; (2) decreases in test-taking anxiety or stress; (3) and/or increases in test-taking familiarity (Hausknecht et al., 2002; Lievens et al., 2007). Figure 1 demonstrates the different types of score improvements in Figure 1.

*Different types of changes in aptitude test scores upon retesting (Modified from Messick, 1981)*



retesting situations. In this figure, each individual (A, B, and C) did not pass the hiring cut-off on first attempt of the aptitude test. All of the applicants do pass the hiring cut-off on the second test administration. However, as seen on the vertical axis, only A and B will be successful at the job.

In this illustration, Applicant A represents an individual who would have been initially successful if hired, but individual issues, such as high test anxiety or measurement error, impeded their test performance. When retesting effects arise from

decreases in test-taking anxiety or stress, scores are increasing from removing debilitating construct-irrelevant factors (Lievens et al., 2007). Retesting is ideal in this situation because the first aptitude score was not an accurate reflection of the applicant's ability (Messick, 1981). Researchers have found test-taking anxiety and stress are negatively related to aptitude test performance (Hembree, 1988; Zeidner, 1995). Test-taking anxiety is both a feeling of state anxiety (in that the current testing creates anxiety) and trait anxiety (in that tests in general generate anxiety; Arvey, Strickland, Drauden, & Martin, 1990). Test-taking stress, on the other hand, reflects current symptom manifestations of anxiety, such as increased heart rate (Day & Ziemer, 2003). An estimated meta-correlation of test-taking anxiety and aptitude test scores is  $r = -.18$  (Reeve, Heggestad, & Lievens, 2009). Although the relationship between test-taking anxiety/stress and test scores has been shown in the past, its specific contribution to retesting effects has not been established (Lievens et al., 2007). That is, the relationship has not been tested while controlling for other factors that potentially impact test scores (e.g., test-taking motivation, test-taking familiarity).

Not hiring Applicant B was a correct choice at the first test administration, as they would have performed poorly on the job. However, over time, they increased their actual ability and passed the test score cut-off point and would be successful at the job. When retesting effects arise from increases in actual ability, scores are increasing due to construct-relevant changes (Lievens et al., 2007). This effect is the only legitimate increase in test scores (Messick, 1981). In this situation, the aptitude test is unbiased and any score increase is the result of the individual increasing their cognitive mental ability, and no other factors have an impact (Messick, 1981).

Finally, Applicant C obtained a test score that allowed them to pass the hiring cut-off point, but they will not be successful in the job. This applicant represents someone who has learned techniques to improve their score on the aptitude test (e.g., test-taking wisdom, answer selection tricks), but did not increase their actual ability. Applicant C is the most problematic scenario (Messick, 1981). Hausknecht et al. (2002) believe that most score-increases are assumed to be from these non-construct related changes (i.e., not from increases in actual abilities or skills). When retesting effects arise from increased test-taking familiarity, scores are increasing due to non-ability related skills (Lievens et al., 2007). This type of increase is the most problematic (Messick, 1981). Test-taking familiarity can refer to general knowledge and comfort of testing situations, memorization of answers from previous test administrations, and test-wisdom (Hausknecht et al., 2007; Lievens et al., 2007; Reeve et al., 2009). In this retesting situation, the applicant has either learned to memorize answers from the previous testing or learned test-taking techniques to improve their scores (e.g., answer selection tricks; Messick, 1981). Although researchers have found that test-taking familiarity is positively related to aptitude test performance (meta-correlation of  $r = .11$ ; Anastasi, 1981; Kulik, Bangert-Drowns, & Kulik, 1984), its specific influence with retesting effects has also not been established (Lievens et al., 2007). That is, the relationship has not been tested while controlling for other factors that potentially impact test scores (e.g., test-taking motivation, test-taking stress).

Another individual difference that seems to impact aptitude test scores is test-taking motivation. Test-taking motivation is associated with individuals making an effort to answer test questions correctly (Arvey et al., 1990). In Wise and DeMars' (2005) meta-

analysis, motivation had an average effect size of  $g = .59$  on cognitive tests. Although test-taking motivation is not within the conceptualized definition of retesting effects, its direct influence on the retesting effect-cognitive test performance relationship needs to be examined more thoroughly (Hausknecht et al., 2007).

Overall, cognitive tests are susceptible to retesting and coaching effects (Hausknecht et al., 2002; Hausknecht et al., 2007; Kulik, Bangert-Drowns, & Kulik, 1984; Kulik, Kulik, & Bangert, 1984). These effects combined have been found to increase scores by three-quarters of a standard deviation (Kulik, Bangert-Drowns, & Kulik, 1984). After the first test administration, coaching strategies can become more efficient and more accurate as applicants understand the testing situation and what is being asked of them. Although there have been many meta-analyses on coaching and retesting effects, these studies, unfortunately, didn't disentangle the potential sources of retesting effects, thereby confounding coaching effects with other constructs, such as test-taking anxiety and test-taking familiarity. Additionally, these effects may have substantial impacts on raw test scores and selection decisions, and thus, need to be understood more thoroughly.

### **Summary and Hypotheses**

Although a great deal of research on retesting and coaching effects has been conducted, the question of why retesting effects occur remains to be answered (Hausknecht et al., 2002; Lievens et al., 2005). Lievens et al. (2005) recommended that laboratory studies were needed to disentangle why retesting effects occur. Therefore, the current study used an experimental approach to evaluate coaching and retesting effects on aptitude tests over two test administrations.

Two different aptitude tests will be examined: (1) a police selection aptitude test for a specific organization; and (2) a widely used, general aptitude test employed across many occupations and settings. This study uses a student sample, randomly assigned to a “coached” group who participated in a test instruction intervention or to a control group who did not receive any instruction.

First, it needs to be established that the combined effect of coaching and retesting on aptitude tests exists within a student experimental setting. This analysis will ensure that similar phenomena are occurring and, then, further hypotheses can be assessed. The experimental (coached) group will be given test instruction, thereby exposing them to both coaching and retesting.

*Hypothesis 1a:* Aptitude test scores will increase upon subsequent test administrations for the experimental (coached) group.

On the other hand, because the control group was not given any test instruction, any aptitude test score increases are the result of retesting effects. This effect will also be evaluated to ensure similar phenomena from past research are occurring within this sample.

*Hypothesis 1b:* Aptitude test scores will increase upon subsequent test administrations for the control group.

Additionally, the relationship between coaching and retesting effects needs to be explored (Hausknecht et al., 2007). This relationship will be examined by analyzing whether there is an interaction between test instruction and the aptitude tests. That is, coaching will strengthen the positive relationship between the score increases on second aptitude test administrations.

***Hypothesis 1c:*** Individuals who received test instruction (in the coached group) will significantly improve their aptitude test scores over individuals who did not receive instruction (in the control group).

Because there is a possibility that individuals prepared for the aptitude tests on their own (outside the instruction session), individuals may increase their test scores via coaching effects in the form of informal preparation. Therefore, it is expected that individuals who prepared more for the aptitude tests will increase their aptitude test scores.

***Hypothesis 2:*** Individuals who prepared for the aptitude tests will significantly improve their aptitude test scores, after controlling for demographic variables and retesting effects (test-taking anxiety, test-taking stress, test-taking familiarity, and test-taking motivation).

Assuming retesting effects exist, a further understanding of why and how they are related to aptitude test scores needs to be investigated (Hausknecht et al., 2002; Lievens et al., 2005; Lievens et al., 2007). Specifically, the relationships between the potential retesting effects sources (e.g., test-taking anxiety, stress, familiarity, and motivation) with aptitude test scores will be examined.

***Hypothesis 3:*** After controlling for demographics (a) test-taking anxiety will be negatively related to aptitude test scores; (b) test-taking stress will be negatively related to aptitude test scores; (c) test-taking familiarity will be positively related to aptitude test scores; (d) and test-taking motivation will be positively related to aptitude test scores. These relationships will exist at both aptitude test administrations, using cross-sectional data.

## Method

This study was approved by the Saint Mary's University Research Ethics Board (Appendix A).

### Sample

Fifty-two students were recruited from Saint Mary's University's campus. Participants were either compensated through the psychological bonus credit system or from recruitment posters for \$20 compensation. Twelve participants did not complete all three sessions of the study and one participant was deleted in the data cleaning process due to questionable responding (e.g., the participant skipped many questions and chose to quit the cognitive tests before the allotted time was complete). Therefore, there were usable data from 39 participants (N = 17 men; N = 22 women). The mean age of participants was 23.92 years old, ranging from 18 to 44 (SD = 6.56), and with a mean enrollment year of 2.76 (SD = 1.49). Over half of the respondents (53.84%) stated that their major was psychology. The sample was representative of Canada in terms of its ethnic demographics as 81.6% were Caucasian, 2.6% were African-Descent, 2.6% were Middle Eastern, 5.1% were South Asian, 2.6% were Hispanic, and 5.1% were Asian (Canada is approximately 86.6% Caucasian and 13.4% Visible Minorities; Statistics Canada, 2006). When identifying their first language, 84.6% of participants spoke English, and the remaining participants spoke French, Hindi, Japanese, Chinese, Arabic or Malayalam—this language makeup is representative of Halifax (Halifax Regional Municipality has 91.4% English speakers; Statistics Canada, 2010). Most participants (92.3%) were single, 5.1% were married and 2.6% were common-law.

## **Procedure**

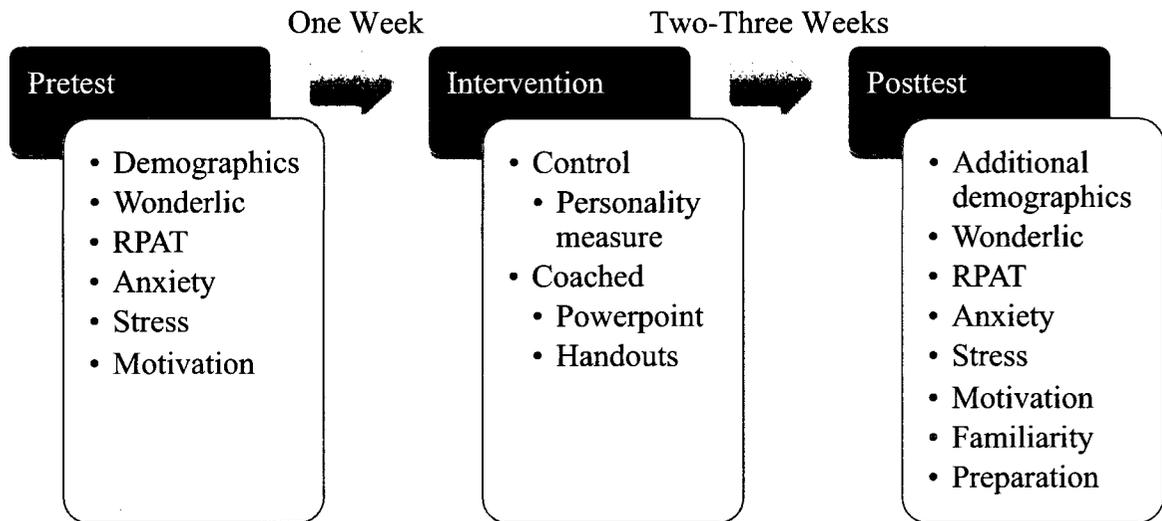
This study was advertised on the Saint Mary's University Psychology Bonus System and through recruitment flyers (see Appendix B) posted around Saint Mary's University campus. Recruitment was also conducted in psychology and human resources classes through verbal recruitment and/or a powerpoint slide (see Appendix B). The study was marketed as "Would You Get Hired? Examining an Applicant Selection Test." Participants received either 3% bonus marks toward a psychology class or \$20 for an incentive to partake. An additional incentive was offered for a draw for a \$100 pre-paid Visa gift card. The draw was used to simulate a hiring reward and, therefore, participants were only placed into the draw if they had passed the cut-off point to "be hired" on the selection tests. The study was composed of three approximately 45-minute sessions. All sessions were conducted face-to-face and all tests were administered with paper and pencil. To ensure anonymity, students created their own identification code to match their data across all three sessions. This code was made up of the first three letters of their middle name and the last three numbers of their phone number.

During the Pretest session, participants completed: demographic questions (Appendix C); the Royal Canadian Mounted Police Aptitude Test (RPAT; Appendix D); the Wonderlic Personnel Test; questions on test-taking anxiety, test-taking stress, and test-taking motivation (Appendix E).

The Intervention session took place five to ten days after Session One, usually occurring in a week's span. During the Intervention, participants went through one of two conditions. The first condition was the experimental manipulation. This manipulation involved the participants receiving a powerpoint presentation and handouts on how to study

Figure 2.

*Measures used within each study session*



for the aptitude tests (see Appendix F). This instruction was followed up, approximately a week later, with an e-mail containing a reminder to study for the Posttest, as well as a list of websites that could be used to study with (see Appendix F). Participants were also reminded of the incentive of being put in a draw, given that they were “hired”, for the Visa gift card. This draw was meant to increase participants’ motivation to succeed. The other condition was the control group. Instead of the instructional intervention, participants in this session completed a personality test (Appendix G). This personality measure was purely a distractor test and was not used in any analyses. The control group also received an e-mail reminding them to sign up for the Posttest, as well as a reminder of the draw incentive.

The Posttest session occurred two-three weeks after the Intervention. In this session, participants completed the same aptitude tests, and the same measures of test-taking

anxiety, stress, and motivation. Finally, the participants were asked questions about their test-taking familiarity, levels of test preparation, manipulation checks, and additional demographic questions (Appendix H). After completing the Posttest, the participants were given the chance to be put in the draw for the \$100 Visa gift card.

## Measures

### Pretest.

**Demographics.** Demographic questions asked included: gender, age, ethnicity, first/main language, marital status, socioeconomic class, estimated annual family income, year of study, and study major (see Appendix F). Socioeconomic class was measured on a 5-point categorical scale from “Lower Class” to “Upper Class”. Gender was measured with two categorical points of “Male” or “Female”. All the remaining demographic questions were open-ended.

**Cognitive abilities (aptitude).** Two aptitude tests were used in this study. An alternate, shortened version of the Royal Canadian Mounted Police Aptitude Test (RPAT) and the Wonderlic Personnel Test.

**RPAT.** Due to the secure nature of the RPAT, an alternate, shortened form of the RPAT was used. It consisted of items that are given to applicants to study with from a preparatory guide (RCMP Preparatory Guide, 2008). These items were originally developed for the full version of the RPAT, thereby are reflective of the actual items on the RPAT (RCMP, 2000). The alternate shortened RPAT is a 23-item measure of cognitive aptitude with a time limit of 28 minutes. Participants were asked to respond to six items about composition, three items about comprehension, two items about memory, three items about judgement, two items about observation, three items about logic, and four items

about computation. The items have four multiple choice options of potential answers. A score was created by adding the number of items correctly answered, with judgement questions weighted double. Cronbach's alpha for the full version of the RPAT ranges from .86 to .87. However, the shortened version used in this study had an alpha of .46. This low reliability is mainly from using only 23-items assessing seven different constructs. These constructs are very different in nature, and therefore, will not be strongly correlated with each other, particularly with a smaller sample size. To establish whether this shortened version of the RPAT was equivalent in terms of its reliability with the full-length RPAT, a correction for test length was conducted. The Spearman Brown prophecy formula was used to estimate the reliability of the a full length RPAT (150 items) based on this shortened version (Crocker & Algina, 1986). In the present study, the coefficient was estimated to be .85

*Wonderlic Personnel Test.* The Wonderlic Personnel Test is a 50-item measure of cognitive ability with a time limit of 12 minutes. It assesses aptitude from questions pertaining to general intelligence (Wonderlic, 2002). After the age of 30, participants' scores are corrected for, in that additional points are given to older test-writers. The Wonderlic is reported as having internal consistency of  $\alpha = .88$  to  $.94$  (Wonderlic, 2002). A specific alpha was not computed for this study because internal consistency is an inappropriate measure for speeded tests (Anastasi, 1982). Using equivalent forms reliability, however, is more appropriate for speeded tests (Crocker & Algina, 1986). Past research on the Wonderlic has found equivalent form reliability to be between  $.73$  and  $.95$  (Wonderlic, 2002).

***Test-taking anxiety.*** Test-taking anxiety was measured with the Comparative Anxiety subscale of the Test Attitude Survey from Arvey et al. (1990). It is a 10-item scale using a 7-point Likert-type scale (1 = Strongly Disagree to 7 = Strongly Agree). Sample items include, “I usually get very anxious about taking tests” and “I am not good at taking tests.” The internal consistency of this scale during this study was good ( $\alpha = .86$ ) and the item-total correlations ranged from .35 to .75. Although an item-total correlation of .35 may be problematic, the item “I dislike taking tests of this type” appears to be assessing a different facet of test-taking anxiety, and is therefore, pertinent for this scale.

***Test-taking stress.*** Test-taking stress was measured with 4-items modified from the Day and Ziemer (2003) questionnaire. Each of the items is rated using a 7-point Likert-type scale (1 = Strongly Disagree to 7 = Strongly Agree). Sample items include, “Writing this test made me feel nervous” and “I believe that my heart rate increased during this test.” This scale had good internal consistency ( $\alpha = .88$ ) within this study and the item-total correlations ranged from .61 to .81.

***Motivation.*** Test-taking motivation was measured with 10-items from the Motivation subscale of the Arvey et al. (1990) Test Attitude Survey. Each of the items is rated using a 7-point Likert-type scale (1 = Strongly Disagree to 7 = Strongly Agree). A sample item is “Doing well on this test is important to me.” The internal consistency of this scale was good for the Pretest ( $\alpha = .93$ ) and the item-total correlations ranged from .64 to .84.

***Intervention.*** Participants went through the experimental manipulation during the Intervention. Half the participants were given instruction on how to study for the test and

the other half completed a personality inventory that was not used for the purposes of this study.

**Posttest.**

*Cognitive abilities (aptitude).* The same shortened version of the RPAT was used during the Posttest. Alpha for the Posttest was .31. After correcting alpha with the Spearman Brown prophecy formula (Crocker & Algina, 1986) reliability was .75. The same Wonderlic was also re-administered.

*Test-taking anxiety.* The same test-taking anxiety scale from Arvey et al. (1990) was used in the Posttest. Internal consistency was good ( $\alpha = .91$ ) and the item-total correlations ranged from .50 to .83.

*Test-taking stress.* The same modified test-taking stress scale from Day and Ziemer (2003) was used in the Posttest. Internal consistency was good ( $\alpha = .89$ ) and the item-total correlations ranged from .61 to .84.

*Test-taking familiarity.* Test-taking familiarity was measured with four items created for this survey. Each of the items was rated using a 7-point Likert-type scale (1 = Strongly Disagree to 7 = Strongly Agree). Sample questions include “I have taken tests similar to this test before” and “I recognized the items on this test from the first study session.” The internal consistency of this scale was adequate ( $\alpha = .66$ ). The item-total correlations ranged from .56 to .61.

*Preparation.* Preparation was measured with three items. Two items were composed of the Preparation subscale of the Arvey et al. (1990) Test Attitude Survey. One additional item was created and added to measure test preparation. These three items were rated using a 7-point Likert-type scale (1 = Strongly Disagree to 7 = Strongly

Agree). Sample items include “I spent a good deal of time preparing for this test” and “I prepared a lot for this test”. These items had a good internal consistency ( $\alpha = .83$ ). The item-total correlations ranged from .56 to .85. In follow-up to the preparation scale, two questions were asked of “If you used outside materials to study for this test (a) how many sources (books, websites, etc.) did you use to study? (b) and how many hours did you study for?” These items were open-ended and were correlated  $r = .84$ .

***Test-taking motivation.*** The same 10-item test-taking motivation scale from Arvey et al. (1990) used in the Pretest was used again. Internal consistency was  $\alpha = .91$ , and the item-total correlations ranged from .41 to .80.

## Results

This study is a randomized control pre-post test experimental design. The manipulation, or independent variable, is whether the participants are instructed on the aptitude tests. The dependent variables are the scores on the aptitude tests. All statistical analyses were performed using SPSS Version 16.0 for Mac.

### Preliminary Analyses

Prior to analyses, statistical assumptions were checked. All preliminary analyses were conducted on the data separated by condition group. First, descriptives were run on all the variables to ensure the data was in the correct range of responses, had an average skew, and an average kurtosis. All variables were in appropriate ranges and normally distributed. A missing values analysis was conducted to see if there were any variables that had more than 10% missing data. The only variable missing more than 10% was estimated total family income. Little’s MCAR test was significant, indicating that the missing data were not random. Estimated Family Income had missing data which was not

missing at random as it significantly related to eight items in the Pretest Wonderlic, three items in the Pretest RPAT, one item in the Pretest anxiety scale, five items in the Posttest Wonderlic, one item in the Posttest RPAT, one item in the Posttest anxiety scale, one item in the Posttest Motivation scale, and one item on the Familiarity scale. Because the missing data were systematic (or nonrandom), estimation maximization was utilized. Additionally, no participant was missing more than 10% of their data. All variables were checked for univariate outliers. One outlier ( $z = 3.53$ ) was found. This participant indicated that they “Strongly Agreed” to using outside sources to prepare. This response is within the range of valid responses, and was not deleted or windsorized. Using a  $p < 0.001$  criterion for Mahalanobis distance (or a  $\chi^2 > 34.53$ ), no multivariate outliers were found. Normality was examined through histograms and through the plotting of standardized residuals against standardized predicted values (probability plots). Linearity and homoscedasticity were also examined through these probability plots. Homoscedasticity was also examined through box-plots. No violations were made in regards to normality, linearity, and homoscedasticity.

Independent-samples t-tests were conducted on participants who dropped out of the study and those who completed all three sessions. Participants who dropped out differed on age,  $t(47.06) = 2.43, p < .05$ , two-tailed, and anxiety,  $t(49) = 2.97, p < .01$ , two-tailed. Therefore, participants who dropped out were excluded from further analyses<sup>1</sup>. Finally, independent-samples t-tests were conducted on all the study variables, including demographics to ensure condition groups were equivalent. The experimental

---

<sup>1</sup> Analyses were also conducted including the 12 participants who dropped out. Inclusion of these participants changed the pattern of results. Implications for the inclusion and exclusion of these subjects are presented in the discussion section.

condition groups significantly differed on socioeconomic class,  $t(37) = 1.93, p < .05$ , one-tailed. Therefore, socioeconomic class was controlled for during the analyses.

A check was conducted on whether participants report that they read the e-mails sent to them on a scale of 1 = Strongly Disagree to 7 = Strongly Agree. Participants, in general, slightly agreed that they read the e-mails ( $M = 4.40, SD = 1.72$ ). Additionally, the participants were asked if they had applied to a police organization before (to ensure Pretest wouldn't examine potential retesting and coaching effects already), and no participant had applied.

### **Analyses**

Table 1 shows the means, standard deviations, and intercorrelations among the variables included in this study. The data was then split to examine any differences by condition group. Table 2 shows the means, standard deviations, and reliability coefficients split by condition group. Table 3 shows the intercorrelations among the variables, split by condition group<sup>2</sup>.

To examine whether relationships between variables were consistent among condition groups, a calculation of the test of difference between two independent correlation coefficients was conducted (Preacher, 2002). One particular trend was the differences of relationships involving Wonderlic test scores and test-taking anxiety between condition groups. This correlation was significantly different between the control and the coached group with Pretest Anxiety and the Pretest Wonderlic ( $z = 1.97$ ), Pretest

---

<sup>2</sup> No correlations are shown for amount of sources studied from and hours studied for the control group as participants did not prepare for the Posttest in this condition.

Table 1.  
Means, standard deviations, and intercorrelations among the study variables

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
Pretest.																				
1. Gender <sup>i</sup>	.44 (.50)	-																		
2. Age	.17	23.92 (6.56)	-																	
3. Study Year	.17	.16	-																	
4. Socioeconomic Class	.20	-.02	.08	-																
5. Family Income	94,696.36																			
6. Wonderlic	(56,672.14)	-.10	.19	.24	.34 <sup>a</sup>	-														
7. RPAT	26.35 (6.27)	.16	.36 <sup>a</sup>	.26	-.15	.32 <sup>a</sup>	-													
8. Test-Taking Anxiety	16.79 (3.00)	-.06	.15	.12	-.18	.18	.45 <sup>b</sup>	(.46)												
9. Test-Taking Stress	3.23 (1.09)	-.43 <sup>b</sup>	-.36 <sup>a</sup>	-.20	.04	-.11	-.53 <sup>c</sup>	-.40 <sup>b</sup>	(.86)											
10. Test-Taking Motivation	3.00 (1.48)	-.19	-.14	-.09	.05	-.06	-.34 <sup>a</sup>	-.19	.66 <sup>c</sup>	(.88)										
Intervention.																				
11. Instruction <sup>ii</sup>	5.14 (1.15)	.30 <sup>a</sup>	.18	-.17	-.06	.03	.28 <sup>a</sup>	.24	-.48 <sup>b</sup>	-.20	(.93)									
Posttest.																				
12. Wonderlic	.51 (.51)	-.18	-.10	.02	-.30 <sup>a</sup>	-.15	.12	.09	.02	-.06	-.17	-								
13. RPAT	26.08 (10.49)	.13	.31 <sup>a</sup>	.16	-.20	.31 <sup>a</sup>	.89 <sup>c</sup>	.50 <sup>b</sup>	-.54 <sup>c</sup>	-.30 <sup>a</sup>	.32 <sup>a</sup>	.13	-							
14. Test-Taking Anxiety	17.92 (2.60)	.31 <sup>a</sup>	.08	.27	.01	.02	.46 <sup>b</sup>	.43 <sup>b</sup>	-.39 <sup>b</sup>	-.16	.15	.21	.53 <sup>c</sup>	(.31)						
15. Test-Taking Stress	3.16 (1.16)	-.35 <sup>a</sup>	-.43 <sup>b</sup>	-.16	.13	-.00	-.41 <sup>b</sup>	-.37 <sup>b</sup>	.90 <sup>c</sup>	.63 <sup>c</sup>	-.53 <sup>c</sup>	.03	-.41 <sup>b</sup>	-.29 <sup>a</sup>	(.91)					
16. Test-Taking Motivation	2.94 (1.49)	-.16	-.14	-.11	.03	.01	-.10	-.09	.53 <sup>c</sup>	.69 <sup>c</sup>	-.33 <sup>a</sup>	.04	-.10	-.16	.70 <sup>c</sup>	(.89)				
17. Test-Taking Familiarity	5.14 (0.90)	.26	.05	.09	.13	.12	.34 <sup>a</sup>	.27 <sup>a</sup>	-.53 <sup>c</sup>	-.13	.72 <sup>c</sup>	-.15	.32 <sup>a</sup>	.17	-.51 <sup>b</sup>	-.16	(.91)			
18. Preparation	4.42 (1.19)	.00	.29 <sup>a</sup>	.05	-.02	.01	.11	-.02	-.28 <sup>a</sup>	-.27	.01	-.10	.18	.14	-.26	-.19	.09	(.66)		
19. Amount of Study Sources	1.79 (1.02)	.03	-.22	-.02	.29 <sup>a</sup>	-.02	-.21	-.44 <sup>b</sup>	.30 <sup>a</sup>	-.01	-.14	.20	-.27 <sup>a</sup>	-.12	.36 <sup>c</sup>	.25	-.10	.05	(.84)	
20. Hours Studied	.29 (.90)	.09	-.09	-.08	.11	-.21	-.15	-.21	.16	-.07	-.09	.32 <sup>a</sup>	-.13	-.07	.15	.10	-.03	.21	.71 <sup>c</sup>	-
	.16 (.55)	.22	-.06	.03	.16	-.27 <sup>a</sup>	-.07	-.19	.11	-.07	.04	.24	-.06	.06	.04	.04	.16	.71 <sup>c</sup>	.84 <sup>c</sup>	-

Note. N = 39. Reliability coefficients for each variable shown in parentheses along the diagonal

<sup>i</sup> Gender was coded 0 = Female, 1 = Male, <sup>ii</sup> Instruction was coded 0 = Control Group, 1 = Coached Group

<sup>a</sup>  $p < .05$ , <sup>b</sup>  $p < .01$ , <sup>c</sup>  $p < .001$ , one-tailed.

Table 2.

*Means, standard deviations, and reliability coefficients among the study variables by condition group*

Variable	Control (N = 19)			Coached (N = 20)		
	M	SD	$\alpha$	M	SD	$\alpha$
<b>Pretest</b>						
Gender	.53	.51	-	.35	.49	-
Age	24.58	7.07	-	23.30	6.17	-
Study Year	2.74	1.88	-	2.80	1.06	-
Socioeconomic Class	2.42	.69	-	1.95	.83	-
Family Income	102,995.74	62,745.06	-	86,811.95	45,978.64	-
Wonderlic	25.63	6.89	-	27.05	5.72	-
RPAT	16.53	3.08	.37	17.05	2.98	.53
Test-Taking Anxiety	3.21	.93	.81	3.26	1.25	.90
Test-Taking Stress	3.09	1.41	.83	2.91	1.58	.92
Test-Taking Motivation	5.34	.80	.86	4.95	1.41	.96
<b>Posttest</b>						
Wonderlic	28.84	6.34	-	30.40	6.39	-
RPAT	17.37	2.65	.20	18.45	2.50	.37
Test-Taking Anxiety	3.12	1.02	.87	3.20	1.31	.93
Test-Taking Stress	2.88	1.29	.83	3.01	1.70	.92
Test-Taking Motivation	5.27	.82	.89	5.01	.97	.92
Test-Taking Familiarity	4.08	1.34	.52	3.80	1.61	.72
Preparation	1.58	.75	.85	1.98	1.21	.84
Amount of Study Sources	.00	.00	-	.58	1.21	-
Hours Studied	.03	.11	-	.29	.75	-

*Note.* Gender was coded 0 = Female, 1 = Male

Table 3.  
Intercorrelations among the study variables split by condition group

Variable	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.
Pretest																			
1. Gender <sup>1</sup>	-	.16	.27	.12	-.13	.58 <sup>b</sup>	-.15	-.28	-.11	.30	-.04	.06	-.12	-.15	-.08	-.39 <sup>a</sup>	.13	-	-
2. Age	.16	-	.10	-.19	-.11	.27	.02	-.38	-.46 <sup>a</sup>	.01	.23	.09	-.49 <sup>a</sup>	-.32	-.15	.54 <sup>b</sup>	-.24	-	-
3. Study Year	.04	.32	-	-.12	.15	.31	.30	-.47 <sup>a</sup>	-.21	-.23	.25	.33	-.34	-.08	.20	.16	-.06	-	-
4. Socioeconomic Class	.18	.07	.41 <sup>a</sup>	-	.39	-.20	-.37	.21	.39	.18	-.19	-.06	.24	.17	.10	-.16	.32	-	-
5. Family Income	-.14	.59 <sup>b</sup>	.46 <sup>a</sup>	.25	-	.29	.26	.02	-.09	-.08	.34	.02	.11	.13	.12	.12	.30	-	-
6. Wonderlic	.33	.53 <sup>b</sup>	.16	-.04	.44 <sup>a</sup>	-	.50 <sup>a</sup>	-.30	-.30	.14	.93 <sup>c</sup>	.42 <sup>a</sup>	-.22	-.12	.19	-.06	-.15	-	-
7. RPAT	.06	.32	-.18	.00	.11	.39 <sup>a</sup>	-	-.14	-.20	-.14	.58 <sup>b</sup>	.30	-.08	.01	.07	.04	-.40 <sup>a</sup>	-	-
8. Test-Taking Anxiety	-.56 <sup>b</sup>	-.37	.09	-.04	-.25	-.76 <sup>c</sup>	-.61 <sup>b</sup>	-	.77 <sup>c</sup>	-.41 <sup>a</sup>	-.19	-.27	.92 <sup>c</sup>	.80 <sup>c</sup>	-.28	-.23	.02	-	-
9. Test-Taking Stress	-.28	.16	.08	-.23	-.06	-.37	-.18	.60 <sup>b</sup>	-	-.27	-.21	-.02	.78 <sup>c</sup>	.83 <sup>c</sup>	.02	-.33	-.07	-	-
10. Test-Taking Motivation	.29	.27	-.16	-.25	.06	.45 <sup>a</sup>	.49 <sup>a</sup>	-.52 <sup>b</sup>	-.20	-	.02	.07	-.40 <sup>a</sup>	-.58 <sup>b</sup>	.54 <sup>b</sup>	-.21	-.07	-	-
Posttest																			
11. Wonderlic	.34	.43 <sup>a</sup>	.01	-.17	.32	.85 <sup>c</sup>	.42 <sup>a</sup>	-.80 <sup>c</sup>	-.36	.54 <sup>b</sup>	-	.42 <sup>a</sup>	-.10	.04	.05	.03	-.20	-	-
12. RPAT	.12	.13	.18	.19	.09	.48 <sup>a</sup>	.55 <sup>b</sup>	-.51 <sup>a</sup>	-.27	.27	.62 <sup>b</sup>	-	-.10	.03	-.02	-.20	-.03	-	-
13. Test-Taking Anxiety	-.53 <sup>b</sup>	-.39 <sup>a</sup>	.04	.09	-.11	-.61 <sup>b</sup>	-.61 <sup>b</sup>	.89 <sup>c</sup>	.53 <sup>b</sup>	-.60 <sup>b</sup>	-.67 <sup>b</sup>	-.46 <sup>a</sup>	-	.83 <sup>c</sup>	-.29	-.27	.12	-	-
14. Test-Taking Stress	-.16	.00	-.18	-.03	-.09	-.10	-.16	.39 <sup>a</sup>	.61 <sup>b</sup>	-.23	-.21	-.32	.63 <sup>b</sup>	-	-.20	-.23	.08	-	-
15. Test-Taking Motivation	.51 <sup>a</sup>	.20	-.05	.09	.08	.53 <sup>b</sup>	.47 <sup>a</sup>	-.68 <sup>b</sup>	-.25	.81 <sup>c</sup>	.59 <sup>b</sup>	.39 <sup>a</sup>	-.65 <sup>b</sup>	-.13	-	-.14	-.09	-	-
16. Test-Taking Familiarity	.30	.05	-.10	.02	-.14	.30	-.06	-.30	-.23	.09	.32	.48 <sup>a</sup>	-.25	-.16	.21	-	-.22	-	-
17. Preparation	.04	-.19	-.00	.40 <sup>a</sup>	-.22	-.32	-.52 <sup>b</sup>	.42 <sup>a</sup>	.04	-.12	-.38 <sup>a</sup>	-.26	.48 <sup>a</sup>	.33	-.06	.21	-	-	-
18. Amount of Study Sources	.22	-.09	-.17	.30	-.29	-.31	-.36	.20	-.07	-.05	-.25	-.22	.18	.12	.02	.35	.81 <sup>c</sup>	-	-
19. Hours Studied	.36	-.04	.04	.30	-.42 <sup>a</sup>	-.19	-.27	.16	-.06	.06	-.13	-.02	.10	.06	.06	.30	.78 <sup>c</sup>	.84 <sup>c</sup>	-

Note. Coached Group N = 20, Control Group N = 19

Control group is above the diagonal and the coached group is below the diagonal. <sup>1</sup> Gender was coded 0 = Female, 1 = Male

<sup>a</sup>  $p < .05$ , <sup>b</sup>  $p < .01$ , <sup>c</sup>  $p < .001$ , one-tailed.

Anxiety and the Posttest Wonderlic ( $z = 2.60$ ), and Posttest Anxiety and the Posttest Wonderlic ( $z = 2.04$ ). Finally, there was a significant difference between the experimental groups for Test-Taking Familiarity and the Posttest RPAT ( $z = 1.98$ ). This finding is not surprising because one of the questions asked how familiar the test and items were to the participants; that is coached individuals would have felt that they were more familiar with the items because they were coached, and did better as a result.

I examined the mean aptitude tests scores on the with both groups across the testings (see Table 4). I predicted that aptitude test scores would increase upon

Table 4.

*Mean score differences over consecutive tests split by condition group*

	Pretest <i>M (SD)</i>	Posttest <i>M (SD)</i>	<i>d</i>	<i>CI (95%)</i>
<u>Control<sup>i</sup></u>				
Wonderlic	25.63 (6.89)	28.84 (6.34)	.48	.22 - .74
RPAT	16.53 (3.08)	17.37 (2.65)	.29	-.06 - .64
<u>Coached<sup>ii</sup></u>				
Wonderlic	27.05 (5.72)	30.40 (6.39)	.55	.22 - .88
RPAT	17.05 (2.98)	18.45 (2.50)	.51	.25 - .77

*Note.* <sup>i</sup>  $df = 17$ , <sup>ii</sup>  $df = 18$

\*  $p < .05$ , \*\*\*  $p < .001$

subsequent test administrations for the experimental (coached) group as well as for the control group (Hypothesis 1a/b). I also predicted that individuals who received test instruction would significantly improve their aptitude test scores over individuals who did not receive instruction (in the control group; Hypothesis 1c). To assess these hypotheses,

two repeated-measures ANOVAs were conducted with each aptitude test as the dependent variable. Test instruction, or condition group, was the between-subjects variable and session was the within-measures variable (see Table 5 and 6). Overall, Table 5.

*Source table for the repeated measures ANOVA on the Wonderlic*

Source	SS	df	MS	F	$\eta^2$
Randomized Groups					
Intervention	43.16	1	43.16	.57	.02
Error	2,812.84	37	76.02		
Repeated Measures					
Session	209.69	1	209.69	47.35***	.56
Session $\times$ Intervention	.10	1	.10	.02	.00
Intervention $\times$ Error	163.85	37	4.43		
Total	3,229.64	77			

\*\*\* $p < .001$

Wonderlic test scores significantly increased for the entire sample from first administration ( $M = 26.36$ ,  $SD = 6.27$ ) to second ( $M = 29.64$ ,  $SD = 6.33$ ),  $F(1,37) = 47.35$ ,  $p < .001$ ,  $\eta^2 = .56$ . RPAT test scores significantly increased from first administration ( $M = 16.79$ ,  $SD = 3.00$ ) to second ( $M = 17.92$ ,  $SD = 2.60$ ),  $F(1,37) = 5.31$ ,  $p < .05$ ,  $\eta^2 = .13$ .

Scores increased significantly for the coached group on the Wonderlic from the first administration ( $M = 27.05$ ,  $SD = 5.72$ ) to second ( $M = 30.40$ ,  $SD = 6.39$ ),  $t(19) = 4.46$ ,  $p < .001$ . Scores increased significantly for the coached group on the RPAT from

Table 6.

Source table for the repeated measures ANOVA on the RPAT

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	$\eta^2$
Randomized Groups					
Intervention	12.55	1	12.55	1.12	.03
Error	414.40	37	11.20		
Repeated Measures					
Session	24.49	1	24.49	5.31*	.13
Session $\times$ Intervention	1.52	1	1.52	.33	.01
Intervention $\times$ Error	170.66	37			
Total	623.62	77			

\*  $p < .05$

the first administration ( $M = 17.05$ ,  $SD = 2.98$ ) to second ( $M = 18.45$ ,  $SD = 2.50$ ),  $t(19) = 2.37$ ,  $p < .05$ . Therefore, Hypothesis 1a was supported.

Furthermore, scores significantly increased for the control group on the Wonderlic from first administration ( $M = 25.63$ ,  $SD = 6.89$ ) to second ( $M = 28.84$ ,  $SD = 6.34$ ),  $t(18) = 5.58$ ,  $p < .001$ . However, scores did not significantly increase for the control group on the RPAT from first administration ( $M = 16.53$ ,  $SD = 3.08$ ) to the second ( $M = 17.37$ ,  $SD = 2.65$ ),  $t(18) = 1.08$ , *ns*. Therefore, Hypothesis 1b was partially supported.

The between-subjects ANOVA on Instruction for the Wonderlic was not significant,  $F(1,37) = .57$ , *ns*. The interaction between Instruction and the Session for the Wonderlic was also not significant,  $F(1,37) = .02$ , *ns*, indicating that the coaching did not significantly improve aptitude test scores over retesting for the Wonderlic. The between-

subjects ANOVA for the RPAT on instruction was also not significant,  $F(1,37) = 1.12$ , *ns*. Additionally, the interaction between instruction and the RPAT was not significant,  $F(1,37) = .33$ , *ns*, also indicating that the coaching did not significantly improve aptitude test scores over retesting for the RPAT. Therefore, Hypothesis 1c was not supported.

I predicted that individuals who independently prepared for the aptitude tests would significantly improve their aptitude test scores, after controlling for demographic variables and retesting effects (Hypothesis 2). To test this hypothesis, two hierarchical regressions were conducted, one for the Wonderlic and one for the RPAT. Scores on the second aptitude test were the dependent variables. Gender, age, socioeconomic class, test instruction, test-taking anxiety, test-taking stress, test-taking familiarity, and test-taking motivation were entered in Step 1. The extent to which individuals prepared for the aptitude test was entered in Step 2 (see Table 7<sup>3</sup>).  $R^2$  for the regression at Step 1 was not statistically significant for the Wonderlic,  $F(8, 30) = 1.63$ , *ns* or for the RPAT,  $F(8, 30) = 1.08$ , *ns*. After Step 2,  $R^2$  was not significant for the Wonderlic,  $F(9, 29) = 1.59$ , *ns*. Test preparation did not contribute to a significant  $R^2$  change,  $\beta = -.20$ ,  $t(37) = 1.07$ , *ns*. After Step 2,  $R^2$  was not significant for the RPAT,  $F(9, 29) = 1.34$ , *ns*, and test preparation did not contribute to a significant  $R^2$  change,  $\beta = -.16$ ,  $t(37) = .82$ , *ns*. Overall, after controlling for gender, age, socioeconomic class, test instruction, test-taking anxiety, test-taking stress, test-taking familiarity and test-taking motivation, test preparation did not

---

<sup>3</sup> Two additional analyses were run: (1) with Estimated Annual Family Income instead of Socioeconomic Class; (2) and with neither Estimated Annual Family Income nor Socioeconomic Class. Similar patterns of results for all of these analyses were found. Additionally, analyses were run while controlling for the Pretest aptitude tests scores. A similar pattern of results was found regardless of whether or not pretest scores were included.

Table 7.

*Results of regression analysis of test preparation on the second administration of aptitude tests*

Step and Variable	Wonderlic		RPAT	
	$\beta$	$R^2\Delta$	$\beta$	$R^2\Delta$
Step 1.		.30		.22
Gender	.01		.28	
Age	.17		-.07	
Socioeconomic Class	-.14		.07	
Intervention <sup>i</sup>	.14		.30	
Test-Taking Anxiety <sup>ii</sup>	-.37		-.24	
Test-Taking Stress <sup>ii</sup>	.23		.06	
Test-Taking Familiarity	.07		.14	
Test-Taking Motivation <sup>ii</sup>	.19		.01	
Step 2.		.03		.04
Gender	.04		.33	
Age	.15		-.09	
Socioeconomic Class	-.09		.14	
Intervention <sup>i</sup>	.20		.37	
Test-Taking Anxiety <sup>ii</sup>	-.28		-.13	
Test-Taking Stress <sup>ii</sup>	.22		.06	
Test-Taking Familiarity	.12		.19	
Test-Taking Motivation <sup>ii</sup>	.21		.02	
Test Preparation	-.20		-.24	
Total R <sup>2</sup>		.33		.26

Note. N = 39; Gender was coded 0 = Female, 1 = Male

<sup>i</sup> Intervention was coded 0 = Control Group, 1 = Coached Group

<sup>ii</sup> These reflect values at the second test administration

contribute significantly to the aptitude test scores on second administration. Hypothesis 2 was not supported.

I predicted that after controlling for demographics (a) test-taking anxiety will be negatively related to aptitude test scores; (b) test-taking stress will be negatively related to aptitude test scores; (c) test-taking familiarity will be positively related to aptitude test scores; (d) and test-taking motivation will be positively related to aptitude test scores (Hypothesis 3). To assess retesting effects on aptitude test scores, four hierarchical regressions were conducted (see Table 8<sup>4</sup>). Aptitude test scores were the dependent variables. Only variables measured at each session were used to predict that particular administration of the aptitude test. Therefore, only cross-sectional information is being utilized. Demographics were entered in Step 1, and test-taking anxiety, stress, familiarity, and motivation were entered in Step 2. Because preparation and instruction occurred after the pre-test, and test-taking familiarity was only measured during the Posttest, these variables were not included in the regression to predict Pretest variables. For the Pretest,  $R^2$  for the regression at Step 1 was not significant for the Wonderlic,  $F(3, 35) = 2.36, ns$ , or the RPAT,  $F(3, 35) = .71, ns$ . After adding the potential sources of retesting effects in Step 2,  $R^2$  was significant for the Wonderlic,  $F(6, 32) = 2.64, p < .05, R^2_{adj} = .21$ . At Step 2,  $R^2$  was not significant for the RPAT,  $F(6, 32) = 1.90, ns$ . For the Posttest,  $R^2$  for the regression at Step 1 was nonsignificant for the Wonderlic,  $F(5, 38) = 1.70, ns$ , and for

---

<sup>4</sup> Two additional analyses were run: (1) with Estimated Annual Family Income instead of Socioeconomic Class; (2) and with neither Estimated Annual Family Income nor Socioeconomic Class. Similar patterns of results for all of these analyses were found. Additionally, analyses were run while controlling for the Pretest aptitude tests scores: Similar patterns of results were found regardless of whether or not pretest scores were included.

Table 8.

*Results of hierarchical regression analysis of the relationships between retesting effects and aptitude test scores*

Step and Variable	Pretest				Posttest			
	Wonderlic		RPAT		Wonderlic		RPAT	
	$\beta$	$R^2\Delta$	$\beta$	$R^2\Delta$	$\beta$	$R^2\Delta$	$\beta$	$R^2\Delta$
Step 1.		.17		.06		.20		.21
Gender	.13		-.05		.14		.36*	
Age	.34*		.15		.25		.01	
Socioeconomic Class	-.16		-.17		-.10		.11	
Independent Preparation	-		-		-.23		-.23	
Formal Instruction <sup>ii</sup>	-		-		.19		.36*	
Step 2.		.16		.21*		.17		.08
Gender	-.07		-.29		-.07		.25	
Age	.20		-.02		.12		-.11	
Socioeconomic Class	-.11		-.11		-.08		.12	
Independent Preparation	-		-		-.10		-.16	
Intervention	-		-		.16		.34	
Test-Taking Anxiety <sup>i</sup>	-.45		-.61*		-.46		-.35	
Test-Taking Stress <sup>i</sup>	.01		.18		.06		.16	
Test-Taking Motivation <sup>i</sup>	.04		.06		.12		-.00	
Test-Taking Familiarity	-		-		.05		.16	
Total R <sup>2</sup>		.33*		.26		.37		.29

Note. N = 39; \*  $p < .05$ ; <sup>i</sup> Cross-sectional variables were used

<sup>ii</sup> Instruction was coded 0 = Control Group, 1 = Coached Group

the RPAT,  $F(5, 33) = 1.79$ , *ns*. After adding the potential sources of retesting effects in Step 2,  $R^2$  was still not significant for the Wonderlic,  $F(9, 29) = 1.88$ , *ns*. At Step 2,  $R^2$  was not significant for the RPAT,  $F(9, 29) = 1.32$ , *ns*. The only retesting variable that was significantly associated with aptitude test scores was test-taking anxiety during the Pretest of the RPAT,  $\beta = -.47$ ,  $t(37) = 2.37$ ,  $p < .05$ . Retesting effects, combined, also added significant  $R^2$  change for the first Wonderlic. Overall, after controlling for gender, age, socioeconomic class, test preparation, and test instruction, test-taking anxiety, test-taking stress, test-taking familiarity and test-taking motivation did not significantly explain aptitude test scores. Hypotheses 3a, 3b, 3c, and 3d were not supported.

### Discussion

This study attempted to add to the literature by providing a better understanding of the effects of test instruction and preparation on aptitude test scores, as well as to examine possible sources of test-score increases upon retesting. Overall, when examining purely retesting effects, individuals improved their scores by .48 standard deviations for the Wonderlic and by .29 standard deviations for the RPAT. From these results, only increases on the Wonderlic were statistically significant. Previous meta-analyses have found retesting effects to be .26 (Hausknecht et al., 2007). Some researchers have argued that changes in test scores may just reflect regression towards the mean (Campbell & Kenny, 1999). Hausknecht et al. (2007) found, however, that less than ten percent of the effect size increase from retesting effects is actually due to regression toward the mean. However, given the experimental nature of this study (i.e., not just examining retesting effects of those who failed on the first attempt), if regression towards the mean was occurring, scores should both increase and decrease.

The combined effect of retesting and coaching effects increased aptitude test scores significantly. Individuals who were formally instructed on aptitude tests increased their test scores by .55 standard deviations on the Wonderlic and by .51 standard deviations on the RPAT. Previous studies have found effect sizes around .40 (e.g., Kulik, Bangert-Drowns, and Kulik, 1984). These findings either suggest that these particular aptitude tests are susceptible to retesting effects, or that the coaching provided was more beneficial, causing larger coaching effects. After extrapolating rough estimates, coaching effects for the Wonderlic might be around .07 standard deviations and .25 standard deviations for the RPAT. Coaching effects, however, did not significantly increase the aptitude test scores upon retesting. The moderating effect between test instruction and the aptitude test scores was nonsignificant. Preparation, also, did not significantly predict aptitude test scores. In fact, preparation was actually negatively related with the aptitude test scores. Preparation by participants, however, was done on a volunteer basis. In fact, only five participants out of the twenty in the coaching condition actually stated they prepared outside of the instruction intervention; this most likely limited the results. This result may indicate that coaching effects are not as problematic as expressed by the literature. However, retesting effects do appear to be particularly problematic. Given that retesting effects seem to have greater influence on aptitude test score increases, the possible sources of these effects were examined.

It was predicted that there would be a negative relationship between test-taking anxiety and aptitude test scores. Greater test-taking anxiety was related to decreased aptitude test scores. However, after controlling for gender, age, socioeconomic class, test preparation, test instruction, test-taking stress, test-taking familiarity and test-taking

motivation, test-taking anxiety only contributed unique variance to the RPAT at the Pretest. This finding suggests that test-taking anxiety can affect how individuals score on the RPAT on the first attempt. During the Posttest of both aptitude tests, and the Pretest of the Wonderlic, however, test-taking anxiety did not appear to significantly contribute to aptitude test scores. This lack of effect may be due to the fact that the Arvey et al. (1990) anxiety scale was a mix of state and trait anxiety questions. Perhaps current state anxiety may affect test scores, but general trait anxiety would have remained constant; thus, confounding results.

It was predicted that there would be a negative relationship between test-taking stress and aptitude test scores. Higher test-taking stress was related to lower aptitude test scores on the Wonderlic during the Pretest. However, test-taking stress wasn't related to the RPAT at either Pre- or Posttest or the Wonderlic at Posttest. Test-taking stress did not contribute significant variance after gender, age, socioeconomic class, test preparation, test instruction, test-taking anxiety, test-taking familiarity, and test-taking motivation were controlled for. The stress scale (modified by Day & Ziemer, 2003) mainly examined current test-taking stress (vs. general test-taking stress), suggesting that it would be more reflective of current test-taking attitudes. However, because it did not significantly relate to aptitude test scores, it may not be as influential as other variables.

It was hypothesized that there would be a positive relationship between test-taking familiarity and aptitude test scores. Higher levels of test-taking familiarity were not related to aptitude test scores. This finding was very surprising. One would assume that the more familiar one is with a test-taking, the easier the aptitude tests would be for the individual. Individuals with more "test-taking wiseness", or construct-irrelevant

knowledge, seem to not be influencing their aptitude test scores, adding greater validity to these aptitude tests.

Finally, it was predicted that there would be a positive relationship between test-taking motivation and aptitude test scores. Higher levels of test-taking motivation were related to Wonderlic test scores at both test administrations but not the RPAT test scores. However, after controlling for gender, age, socioeconomic class, test preparation, test instruction, test-taking anxiety, test-taking stress and test-taking familiarity, test-taking motivation was not related to the aptitude test scores. One possibility is that mediating or moderating variables may be affecting this relationship.

Overall retesting effects appear to dramatically increase aptitude test scores. However, in trying to disentangle the sources of these effects, only test-taking anxiety partially affected aptitude test scores. This finding suggests that the current theoretical definition of retesting effects needs to be re-examined. That is, other variables may be influencing retesting effects, or there may be potential moderators between these retesting effects sources and increases in test scores.

The combined effect of retesting and coaching effects was over half a standard deviation. This study supports that reconsiderations and cautions should be made around aptitude tests retesting and study materials. Given that retesting effects were found, there may be implications on the administration of the RPAT. Some researchers have found that retesting may enhance construct and predictive validity (e.g., Anastasi, 1981), whereas other researchers found that retesting might attenuate construct and criterion validity (e.g., Lubinski, 2000). Further analyses should evaluate whether there are higher predictive validities with individuals who write on the first test administration or the

second. This type of research should also be considered for individuals who use more study materials.

### **Practical Implications**

There are some practical implications from results from this study. In a meta-analysis by Hausknecht et al. (2002), the impact of retesting effects on selection decisions was fairly substantial. Test scores from the first to the third test administration were found to bring participants who ranked at the 50<sup>th</sup> percentile up to roughly the 75<sup>th</sup>-80<sup>th</sup> percentile. This increase may be the difference between hiring the applicant or not. Aptitude test scores can also increase by over thirty percentiles on just the second testing from the combined effect of retesting and coaching (Hausknecht et al., 2002; Kulik, Bangert-Browns, & Kulik, 1984). These re-testers are commonly pooled and compared with individuals who have taken the aptitude test only once. Dunlap and Snyder (1920) find this process problematic, as it is unequivocal and unfair. Lievens and colleagues (2005) found that mean regression lines tended to under-predict scores for first-time test-takers, but over-predict second-time test-takers. Therefore, there appears to be practical issues caused by retesting and coaching effects. Fortunately, some potential solutions are available.

Given that retesting effects exist, the ability to be retested could be banned. Although applicants can find this to be unfair (e.g., particular day or mood conditions), it has been implemented successfully with other occupations (Carretta, Zelenski, & Ree, 2000). However, even if retesting can be banned within one organization, retesting cannot be banned across all organizations. Therefore, there is a tricky balance between letting applicants feel they are in a fair selection procedure versus using invalid selection

assessment. Banning retesting is the most extreme fix; fortunately less extreme solutions exist. The time length between test administrations could also be adjusted. Salthouse, Schroeder, and Ferrer (2004), however, found that it takes a minimum of seven years before the positive effects of retesting are no longer in effect (even after controlling for cognitive declines in age). Readjusting length of time between retesting can be one method to establish stronger validity coefficients toward predicting future job performance. Past research has also found that cognitive tests that draw items from a large database tend to have less coaching effects (Kulik, Bangert-Drowns, & Kulik, 1984). Developing an item database for the RPAT or the Wonderlic could be a potential solution to coaching effects. Similarly, using Computer Adaptive Testing (CAT) also may reduce retesting and coaching effects. A CAT system selects questions from a pool of precalibrated items appropriate for the level of the specific individual (Economides & Roupas, 2007). At minimum, switching to a large database of items, instead of using identical or alternate forms can help assure that candidates are being chosen based on their true ability and not on their ability to memorize or be coached. Choosing appropriate candidates from valid cognitive tests will also lead to greater validity in predicting candidate performance on the job, which in turn, leads to greater productivity and, therefore, greater profits.

### **Limitations and Future Research**

The primary limitation of this study is its lack of power in certain analyses. An a priori power analysis suggested that 90 participants were needed to obtain full power for this study. In interpreting effects, therefore, nonsignificant effects for Hypothesis 1b, 1c and 2 cannot be determined to be null effects, but are to be deemed inconclusive. For

Hypothesis 1b, there was enough power to detect an effect for the Wonderlic, but not for the RPAT. Power of .27 was obtained. For Hypothesis 1c, power of .13 was obtained for the Wonderlic and .33 for the RPAT. For Hypothesis 2, power of .77 was observed with the Wonderlic for the omnibus test and .44 for the  $R^2$  change. For the RPAT, power of .59 was observed for the omnibus test, and .70 for the  $R^2$  change. Inadequate power may be the reason for having an inability to detect any of these effects or there may be a legitimate null effect. The main focus for future research should be to extend this study to obtain a larger sample size in order to have adequate power to test effects.

This same study should also be replicated with different aptitude tests to establish whether the same retesting and demographic variables influence retesting and coaching effects, or whether it depends on the specific test. Given that this study elapsed for approximately a month between the Pre- and the Posttest, future studies should also examine retesting and coaching effects at different time intervals.

It should also be noted that there is suppression occurring within Hypothesis 3. Test-taking stress is generally negatively related to aptitude test scores, but within these regressions its beta weights becomes positive, indicating partial suppression. This suppression may then amplify the relationship between other variables and aptitude test scores. Further research should examine whether both test-taking anxiety and stress need to be included, or whether one or a combined variable of both would be more appropriate.

Additionally, this study uses a student sample, making the results less generalizable. Although student samples are usually used out of convenience, this sample is advantageous because it poses as a potential applicant sample. Regardless of this

benefit, a student population still is limited in some of its generalizability. It would be ideal to replicate this study within an applicant sample.

Finally, the measures relied on self-report data. Self-reporting may be associated with mono-method bias. Replication of the current findings using alternative or multiple methods is recommended. Fortunately, Conway and Lance (2010) found that using only self-report measures in organizational settings is not as detrimental as widely believed. The authors argued that it is wrong to assume that variables are routinely upwardly biased (Conway & Lance, 2010). Nevertheless, it would be beneficial for future research to use an applicant sample and incorporate a criterion measure, such as external ratings of training or job performance.

There is also speculation about the extent to which individual differences may affect retesting effects on aptitude tests (e.g., Murphy & Davidshofer, 1998; Sackett, Borneman, & Connelly, 2008). Some individuals may benefit more from coaching than others (Murphy & Davidshofer, 1998; Sackett et al., 2008). For example, aptitude test scores and socioeconomic status appear to be related (Briggs, 2001; Sackett et al., 2008). Zwick, Brown, and Sklar (2004) found that family income added significant variance when predicting college performance, even when controlling for Scholastic Aptitude Test (SAT) scores. In selection settings, Fozard and Nuttall (1971) found that men with lower socioeconomic status had lower aptitude scores on eight different dimensions. Sackett et al. (2008) found that coaching effects might be more prevalent in affluent participants because they can afford more coaching materials and tutoring. That is, individuals with higher socioeconomic status (e.g., more money, less need for second jobs, have jobs with flex-time, etc.) would have more money to buy preparation materials and have more time

to be able to study or be tutored (Sackett et al., 2008). Although some research on the relationship between socioeconomic status and aptitude test scores has been conducted, overall, more clarification is needed (Sackett et al., 2008).

Some researchers (e.g., Hausknecht, 2010) have also begun to examine whether personality impacts coaching or retesting effects. For example, even though motivation can impact aptitude test scores and on whether applicants re-test, Hausknecht (2010) found that conscientiousness did not predict retesting effects. Additionally, Hausknecht (2010) found that individuals who passed on the first test, only somewhat improved upon retesting. Participants increased their quantitative score by .15 standard deviations and abstract reasoning by .19, but did not change their personality profile. Individuals who failed, however, improved their quantitative score by .46, abstract reasoning by .45, *and* personality by .40-.60 standard deviations. Future research should further examine personality and its role with coaching and retesting effects.

Additionally, there were some relationships found within this study that might be worthwhile investigating. For example, some cross-sectional and longitudinal relationships between test-taking motivation and test-taking anxiety were significantly negatively related. Understanding whether increased motivation decreases anxiety, increased anxiety decreases motivation, or if there is a third variable influencing this relationship may help understand the intricacies of retesting effects. Additionally, test-taking anxiety, at both test administrations was related to test preparation. Therefore, it seems as though those with high test-taking anxiety within the Pretest prepared more for Posttest. Similarly, those who prepared more for Posttest still seems to be anxious during

Posttest testing. This relationship should be tested to see if there are any impacts or changes in the scores on the aptitude tests.

Finally, because there were differences among participants who dropped out and those who completed the study (see Table 9), future research should examine the

*Table 9.*

*Differences among participants who dropped out or completed the study*

	Dropouts (N = 12)	Complete (N =39)	<i>t</i>
	M (SD)	M (SD)	
Age	21.17 (1.47)	23.92 (6.56)	2.43*
Gender	.17(.39)	.44 (.50)	1.95
Socioeconomic Class	2.38 (.48)	2.18 (.79)	.81
Family Income	94,700.84 (12,447.17)	92,426.13 (47,987.67)	-.27
Wonderlic	22.92 (4.36)	26.36 (6.27)	1.77
RPAT	15.42 (3.18)	16.79 (3.00)	1.37
Test-Taking Anxiety	4.29 (1.01)	3.24 (1.09)	2.97**
Test-Taking Stress	3.48 (1.48)	3.00 (1.48)	1.00
Test-Taking Motivation	4.58 (1.21)	5.14 (1.15)	1.45

Note. *df* = 23.39 – 49

\*  $p < .05$ , \*\*  $p < .01$

differences between individuals who drop out not only within other studies, but also within applicant pools. This issue is particularly important because individuals who are dropping out have higher levels of anxiety, potentially relating to retesting effects.

Moreover, the pattern of relationships between those who drop out and those who completed the entire study may differ between retesting effects and aptitude test scores. With the inclusion of participants who dropped out, the hierarchical regression for the Pretest Wonderlic and RPAT changed. For the Pretest Wonderlic regression, Step 1 and 2 became significant. However, for the Pretest RPAT regression, the beta-weight for test-taking anxiety became nonsignificant. Therefore, the inclusion of dropout participants changes the pattern of results. These findings warrant further investigation.

### **Conclusion**

Overall, with the exception of retesting effects on the RPAT, retesting was associated with increases in aptitude test scores, but coaching did not add to these effects. I examined some factors that may be associated with these increased scores. However, test-taking anxiety, stress, familiarity, and motivation did not uniquely add to this understanding. Future research should examine other potential sources of retesting effects and possible moderators or mediators between retesting effects and aptitude test scores.

## References

- Anastasi, A. (1981). Coaching, test sophistication, and developed abilities. *American Psychologist*, 36(10), 1086-1093. doi:10.1037/0003-066X.36.10.1086.
- Anastasi, A. (1982). *Psychological testing*. (5th ed.), New York: Macmillan.
- Arvey, R., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, 43(4), 695-716. doi:10.1111/j.1744-6570.1990.tb00679.x.
- Briggs, DC (2001). The effect of admissions test preparation: Evidence from NELS: 88. *Chance*, 14(1), 10–21
- Campbell, D., & Kenny, D. (1999). A primer on regression artifacts. New York, NY US: Guilford Press. Retrieved from PsycINFO database.
- Carretta, T., Zelenski, W., & Ree, M. (2000). Basic Attributes Test (BAT) retest performance. *Military Psychology*, 12(3), 221-232.  
doi:10.1207/S15327876MP1203\_4
- Catano, V. M., Wiesner, W. H., Hackett, R. H., & Methot, L. L. (2005). Selection II: Testing. In M. Belcourt (Ed.), *Recruitment and selection in Canada* (3rd ed., pp. 327-389). Human Resources Management. Toronto: Thomas Nelson.
- Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement* (3rd ed.). Mountain View, CA: Mayfield. (Original work published 1988)
- Conway, J., & Lance, C. (2010). What Reviewers Should Expect from Authors Regarding Common Method Bias in Organizational Research. *Journal of Business &*

Psychology, 25(3), 325-334. doi:10.1007/s10869-010-9181-6.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*.

Orlando, FL: Harcourt Brace Jovanovich, Inc.

Day, A. L. & Ziemer, K. (2003). Subjective and objective measures of task control and stress. Presentation at the annual meeting of the Administrative Sciences

Association of Canada, Halifax, NS, June 16.

Dunlap, K., & Snyder, A. (1920). Practice Effects in Intelligence Tests. *Journal of Experimental Psychology*, 3(5), 396-403. doi:10.1037/h0073468.

Economides, A., & Roupas, C. (2007). Evaluation of computer adaptive testing systems.

*International Journal of Web-Based Learning and Teaching Technologies*, 2(1), 70-87. Retrieved from PsycINFO database.

Fozard, J., & Nuttall, R. (1971). General Aptitude Test Battery scores for men differing in age and socioeconomic status. *Journal of Applied Psychology*, 55(4), 372-379.

doi:10.1037/h0031530.

Goldstein, G., & Hersen, M. (Eds.). (1990). *Handbook of psychological assessment* (2nd ed.). Elmsford, NY: Pergamon Press.

Hausknecht, J. (2010). Candidate persistence and personality test practice effects:

Implications for staffing system management. *Personnel Psychology*, 63(2), 299-324. doi:10.1111/j.1744-6570.2010.01171.x.

Hausknecht, J., Halpert, J., Di Paolo, N., & Moriarty Gerrard, M. (2007, March).

Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92(2), 373-385,

doi:10.1037/0021-9010.92.2.373

- Hausknecht, J., Trevor, C., & Farr, J. (2002, April). Retaking ability tests in a selection setting: Implications for practice effects, training performance, and turnover. *Journal of Applied Psychology, 87*(2), 243-254, doi:10.1037/0021-9010.87.2.243
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47-77. doi:10.2307/1170348.
- Hunter, J., & Hunter, R. (1984, July). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*(1), 72-98, doi:10.1037/0033-2909.96.1.72
- Kulik, J., Bangert-Drowns, R., & Kulik, C. (1984, March). Effectiveness of coaching for aptitude tests. *Psychological Bulletin, 95*(2), 179-188, doi:10.1037/0033-2909.95.2.179
- Kulik, J., Kulik, C., & Bangert, R. (1984, June). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal, 21*(2), 435-447, doi:10.2307/1162453
- Kuncel, N., Hezlett, S., & Ones, D. (2004). Academic Performance, Career Potential, Creativity, and Job Performance: Can One Construct Predict Them All?. *Journal of Personality and Social Psychology, 86*(1), 148-161. doi:10.1037/0022-3514.86.1.148.
- Lievens, F., Buyse, T., & Sackett, P. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58*(4), 981-1007. doi:10.1111/j.1744-6570.2005.00713.x.
- Lievens, F., Reeve, C., & Heggstad, E. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology, 92*(6), 1672-1682. doi:10.1037/0021-9010.92.6.1672.

- Lubinski, D. (2000). Scientific and social significance of assessing individual differences: 'Sinking shafts at a few critical points.'. *Annual Review of Psychology*, 51, 405-444. doi:10.1146/annurev.psych.51.1.405.
- Messick, S. (1981). The controversy over coaching: Issues of effectiveness and equity. In B. F. Green (Ed.), *Issues in testing: Coaching, disclosure and ethic bias* (Vol. 11, pp. 21-53). *New directions for testing and measurement*. San Francisco: Jossey-Bass.
- Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: Principles and applications* (4th ed.). Upper Saddle River, NJ: Prentice Hall. (Original work published 1988)
- Preacher, K. J. (2002, May). Calculation for the test of the difference between two independent correlation coefficients [Computer software]. Available from <http://www.quantpsy.org>.
- Reeve, C., Heggstad, E., & Lievens, F. (2009). Modeling the impact of test anxiety and test familiarity on the criterion-related validity of cognitive ability tests. *Intelligence*, 37(1), 34-41. doi:10.1016/j.intell.2008.05.003.
- Royal Canadian Mounted Police. (2000). *RCMP Police Aptitude Test technical report: Unclassified version*.
- Royal Canadian Mounted Police. (2008, July). *Preparatory guide for the RCMP Police Aptitude Test* from [http://www.rcmp-grc.gc.ca/recruiting-recrutement/documents/RPATprepguideE\\_08-07-10.pdf](http://www.rcmp-grc.gc.ca/recruiting-recrutement/documents/RPATprepguideE_08-07-10.pdf)

- Sackett, P., Borneman, M., & Connelly, B. (2008, May). High-Stakes Testing in Higher Education and Employment. *American Psychologist*, 63(4), 215-227. Retrieved September 10, 2009, doi:10.1037/0003-066X.63.4.215
- Salthouse, T., Schroeder, D., & Ferrer, E. (2004). Estimating Retest Effects in Longitudinal Assessments of Cognitive Functioning in Adults Between 18 and 60 Years of Age. *Developmental Psychology*, 40(5), 813-822. doi:10.1037/0012-1649.40.5.813.
- Schmidt, F., & Hunter, J. (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. doi:10.1037/0033-2909.124.2.262.
- Schmidt, F., Hunter, J., McKenzie, R., & Muldrow, T. (1979, December). Impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology*, 64(6), 609-626, doi:10.1037/0021-9010.64.6.609
- Statistics Canada. (2006). *2006 Canada census: Special interest profiles*. Retrieved June 25, 2010, from <http://www12.statcan.gc.ca/census-recensement/2006/dp-pd/prof/sip/Rp-eng.cfm?LANG=E&APATH=3&DETAIL=0&DIM=0&FL=A&FREE=0&GC=0&GID=0&GK=0&GRP=1&PID=97614&PRID=0&PTYPE=97154&S=0&SHOWALL=0&SUB=0&Temporal=2006&THEME=80&VID=0&VNAMEE=&VNAMEF=>
- Statistics Canada. (2010, February 5). *2006 community profiles: Halifax*. Retrieved June 25, 2010, from <http://www12.statcan.ca/census-recensement/2006/dp-pd/prof/92->

591/details/page.cfm?Lang=E&Geo1=CSD&Code1=1209034&Geo2=PR&Code2=13&Data=Count&SearchText=halifax&SearchType=Begins&SearchPR=01&B1=Custom&Custom=1000,7000,8000

Wise, S., & DeMars, C. (2005). Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. *Educational Assessment*, 10(1), 1-17.

doi:10.1207/s15326977ea1001\_1.

Wonderlic, Inc. (2002). *Wonderlic Personnel Test & Scholastic Level Exam User's Manual*. Libertyville, IL.

Zeidner, M. (1995). Coping with examination stress: Resources, strategies, outcomes. *Anxiety, Stress & Coping: An International Journal*, 8(4), 279-298.

doi:10.1080/10615809508249379.

Zwick, Rebecca, Brown, Terran, & Sklar, Jeffrey C. (2004). *California and the SAT: A Reanalysis of University of California Admissions Data*. UC Berkeley: Center for Studies in Higher Education. Retrieved from:

<http://www.escholarship.org/uc/item/9hf3t044>

## Appendix A

**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:

<b>Principal Investigator:</b>	FLORKO, Lauren (Student)
<b>Faculty Supervisor:</b>	DAY, Arla
<b>Name of Research Project:</b>	Evaluating the Royal Canadian Mounted Police Aptitude Test for Coaching and Practice Effects.
<b>REB File Number:</b>	09-250

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 (Annual Report) to the REB at the end of the year and request an extension. You are also required to submit Form #5 (Completion of Research) upon completion of your research.

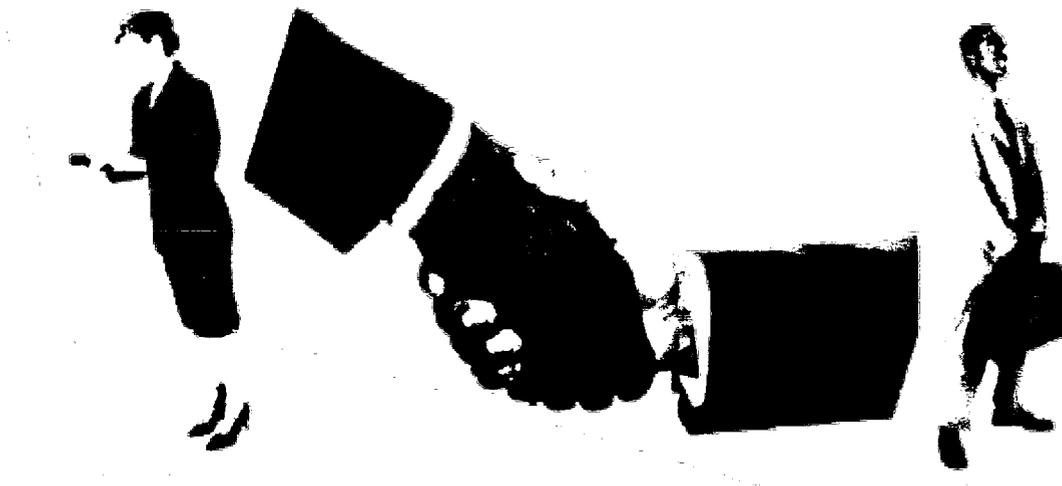
**Date:** 25 January 2010

**Signature of REB Acting Chair:**

Appendix B

# Would You Get Hired?

Examining an Applicant Selection Test



## Receive 3 Bonus Points in your Psychology Class & A Chance to Enter a Draw for a \$100 Visa Gift

This is a three-part study, with each session taking no more than 45 minutes. You may sign up for this study via the SMU Psychology Survey System or by e-mailing me at the e-mail address below. Approved by Saint Mary's University Research Ethics Board, File #09-250.

[lauren.florko@yahoo.ca](mailto:lauren.florko@yahoo.ca)

**Would You Get Hired?**  
[lauren.florko@yahoo.ca](mailto:lauren.florko@yahoo.ca)



## Would You Get Hired? Examining an Applicant Selection Test

- This is a three-part study, with each session taking no more than 45 minutes. You may sign up for this study via the SMU Psychology Sona System or by e-mailing me. Approved by Saint Mary's University Research Ethics Board: File #09-250

[lauren\\_florko@yahoo.ca](mailto:lauren_florko@yahoo.ca)



[www.smupsych.sona-systems.com](http://www.smupsych.sona-systems.com)

Receive **\$20** After the 3 Study Sessions &  
A Chance to Enter a Draw for a **\$100 Visa Gift Card!**

## Would You Get Hired? Examining an Applicant Selection Test

- This is a three-part study, with each session taking no more than 45 minutes. You may sign up for this study via the SMU Psychology Sona System or by e-mailing me. Approved by Saint Mary's University Research Ethics Board: File #09-250



■ [www.smupsych.sona-systems.com](http://www.smupsych.sona-systems.com)

■ [lauren\\_florko@yahoo.ca](mailto:lauren_florko@yahoo.ca)

Receive **3 Bonus Points** in your Psychology Class &  
A Chance to Enter a Draw for a **\$100 Visa Gift Card!**

Appendix C

**Please read and answer the following questions:**

Age \_\_\_\_\_ Gender:  Female  Male

Ethnicity (e.g., Caucasian, Middle Eastern)? \_\_\_\_\_

First (main) language? \_\_\_\_\_

Marital status (e.g., Single, Married)? \_\_\_\_\_

What socioeconomic class did you grow up in (circle one)?

- Lower Class
- Lower Middle Class
- Middle Class
- Upper Middle Class
- Upper Class

What is your estimated total family income? \_\_\_\_\_

What year of study are you currently enrolled in?

- First year
- Second year
- Third year
- Fourth year
- Fifth year
- Sixth year
- Seventh year
- Eighth year or higher

What is your major? \_\_\_\_\_

Appendix D

This is an example mug shot. Please memorize this information, as you will be tested on it at the end of the test.<sup>5</sup>



Name:	David McClury
Gender:	Male
Age:	27
Eye Colour:	Green
Hair Colour:	Brown
Identifying Features:	Tattoo of skull on upper left Bicep
Crime wanted for:	Murder

---

<sup>5</sup> This will handed to the participant on a separate sheet and they will be given two minutes to memorize this information

1. Both women have made previous complaints, none of which were followed through because of insufficient evidence or strong fear of retaliation by the suspect. The suspect is attempting to intimidate the women and is known for irrationale behaviour.

Which word in the sentences above is misspelled?

- a) insufficient
  - b) retaliation
  - c) intimidate
  - d) irrationale
2. On the night of the accident, coeficient of friction testing was conducted using police transport on a cycloidal skid mark.

Which word in the sentence above is misspelled?

- a) accident
  - b) coeficient
  - c) friction
  - d) cycloidal
3. Which word completes the next sentence?

Once she became a public figure, she had to give \_\_\_\_\_ her anonymity.

- a) about
  - b) back
  - c) in
  - d) up
4. Which words complete the next sentence?

My partner and \_\_\_\_\_ went to the home and \_\_\_\_\_ knocking on the door.

- a) I, began
- b) me, began
- c) me, begun
- d) I, begun

5. Which of the following is the best definition of the word **amendment**?
- a) cancellation by making invalid or outdated
  - b) identification by comparison and elimination
  - c) improvement by revision or correction
  - d) protection by establishing rules and laws
6. Which of the following words can be defined as: “**The process of deriving general principles from particular instances**”?
- a) concoction
  - b) deduction
  - c) induction
  - d) reduction

Questions 7 to 9 refer to the following passage:

Throughout its long history, China has seen the rise and fall of a large number of secret organizations. Originally, most of them had no criminal intent. They were simply brotherhoods based on shared political purposes and mutual assistance. However, some of them fell under the control of dishonest people. They then drifted into crime and ended up posing a threat to the social order. This was the case for two of the largest secret organizations of Chinese modern history, the Tsing and Hong organizations.

Many criminal organizations such as the Tsing and Hong and many others controlled brothels, opium dens, casinos and drug trafficking operations. They forced business owners to give them a percentage of their profit. They also committed more serious crimes such as abduction, the trade of women and children, and even assassinations. To achieve impunity from the law, they corrupted government and colonial officials. It was not uncommon for leaders of these organizations to occupy legitimate positions in companies or even in government agencies. They used these jobs as a front for their illegal activities. Each organization had its own system of laws and punishments. The internal law of the Hong organization had five possible punishments: capital punishment, corporal punishment, caning, degradation and banishment. A strict hierarchy existed and obedience to superiors was mandatory members, called apprentices, were placed under a master and were at his service. In the golden age of the Tsing and Hong organizations, a powerful leader could recruit thousands of apprentices.

7. According to the previous passage, which one of the following statements is true?
- a) Chinese secret organizations have always threatened the social order.
  - b) Financial assistance is encouraged in Chinese criminal organizations.
  - c) Few criminal organizations are known in China; only two are known today.
  - d) Some secret Chinese organizations have not turned to crime.

8. According to the previous passage, which one of the following statements is true?
- a) Criminal organizations bribed government employees and senior officials to avoid facing consequences of their misdeeds.
  - b) All leaders of these organizations had legitimate positions in businesses or in the government.
  - c) The trade of women and children was the main activity of Chinese criminal organizations.
  - d) Prostitution is one of the rare illegal businesses in which Chinese criminal organizations did not participate.
9. Which one of the following statements is supported by the previous passage?
- a) In the Hong organization, only apprentices had to obey their superiors.
  - b) In the Hong organization, a person's rank could not be lowered as a form of punishment.
  - c) In the Hong organization, the penalty for disobedience was corporal punishment.
  - d) In the Hong organization, the leader recruited many new members.
10. You are a police officer in a small town. You are on patrol at 2:00 a.m. when you see a car leaving town and moving very fast. You turn on the lights and sirens of your police car and attempt to pull the car over. The car does not pull over and a brief high-speed chase results. After a short distance, the driver loses control of the car and it skids into the ditch. Four men jump out of the car and run into the forest. Of the following, which is the best initial course of action to take?
- a) Call for backup and immediately chase the suspects into the forest.
  - b) Call for backup and request police dog(s). Upon their arrival, begin the chase of the suspects into the forest.
  - c) Determine the registered owner of the car. The following day, go to the registered car owner's home and question him about the incident.
  - d) Draw your gun and order the men to stop. If they do not comply, fire a warning shot into the air

11. You are a police officer in a large municipality. Your coworker and friend, Constable Grey, is in constant financial trouble as the result of loans that she obtained before joining the RCMP. Constable Grey is married and has recently taken on a second job to pay her bills. You have noticed that since she has had this second job, the quality of her work has gone down considerably. Of the following, which is the best initial step to take?
- a) Suggest to Constable Grey that she speak to a financial consultant.
  - b) Lend Constable Grey some money.
  - c) Speak to Constable Grey's spouse to identify possible solutions to her financial difficulties.
  - d) Speak to Constable Grey and tell her about your concerns.
12. At a meeting, your supervisor tells all the members on shift about a new RCMP policy. The policy states that police officers are to wear their hats at all times while on duty in the downtown area. The rationale for the policy is that the wearing of hats projects a professional policing image and allows the public to easily identify the person as a police officer. Later that evening, you and your partner receive a call that a young female was just sexually assaulted by two males who are both well-known to you. The two males are to be arrested. You believe that the two male suspects are at a fountain in the downtown area that is only approachable on foot. There is a music festival being held in the area that is attracting large crowds of people. You and your partner feel that, under these circumstances, it would be inappropriate to wear your hats; the suspects would likely see you approaching and may flee. Of the following, which is the best course of action to take?
- a) Call your supervisor, explain the situation, and request permission not to wear your hats. Follow the decision given.
  - b) Proceed to the area without wearing your hats. Explain to your supervisor afterwards the reason for your action.
  - c) Wear your hat while in the area, consistent with the RCMP policy.
  - d) Do not attempt to arrest the suspects at this time as their identity is known. Wait for another time to arrest them.

For the following questions, please circle the potential match from the given mug shot of the suspected criminal

13.



Suspected criminal

Potential matches

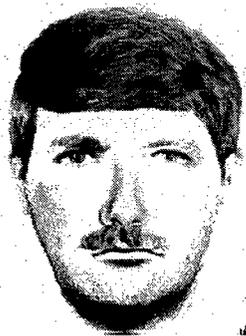
a)



b)



c)



d)



14.



**Suspected criminal**

**Potential matches**

a)



b)



c)



d)



15. You are preparing a report concerning a car accident. The following five pieces of information are to be included in the report:
1. Constable MacInnis explained that the car was heading east bound on 84th Avenue in the first lane and had struck a young girl who was running across 84th Avenue.
  2. The statement was not obtained at this time, as the driver of the car was too emotionally upset.
  3. Constable Smith attended the scene of a car accident and was briefed by Constable MacInnis of what had occurred.
  4. The driver was then turned over to Constable Smith for a statement.
  5. Constable MacInnis then stated that he had detained the driver of the car and that he had read him his rights.

In which of the following orders should the information in the report be presented?

- b) 1, 5, 2, 4, 3
- c) 3, 5, 1, 4, 2
- d) 1, 2, 4, 3, 5
- e) 3, 1, 5, 4, 2

Questions 16 and 17 refer to the following information:

You are investigating a serious accident involving five cars. The car at the front of the accident is labelled 1 and the car at the back of the accident is labelled 5. The remaining cars are labelled according to their location in the chain of cars. The information related to the accident is sketchy and somewhat disorganized. You know the following:

- I. The drivers involved in the accident are Paul, Kim, Jennifer, Kevin, and Sarah.
- II. Kevin was in one of the cars ahead of Kim
- III. Sarah and Jennifer were in car 1 and 5, although not necessarily in that order.

16. Which one of the following statements must be false?
- a) Kim is in a car ahead of Sarah
  - b) Kim is in car 2
  - c) Paul is in car 2
  - d) Kim is in a car behind Paul
17. If Kevin is in car 3, which one of the following statements must be true?
- a) Paul is in a car behind Kim
  - b) Kim is in a car ahead of Sarah
  - c) Sarah is in car 1
  - d) Paul is in a car ahead of Kevin

Questions 18 to 19 refer to the following information:

The RCMP has developed strategic partnerships with various financial institutions, both nationally and internationally, in an effort to reduce the counterfeiting of currency and credit cards. The following table reflects the amount of success they have had over a five year period. For questions 18 and 19, assume that the two currencies are at par (i.e., CDN \$1 = US \$1).

**Counterfeiting Seizure Statistics**

Year	Canadian Currency	American Currency
1996	\$5,121	\$72,500
1997	\$3,211	\$850,000
1998	\$2,127	\$63,427
1999	\$7,500	\$22,500
2000	\$6,835	\$54,073

18. What percentage of the total funds seized in 1999 was Canadian?
- a) 3%
  - b) 4%
  - c) 25%
  - d) 33%
19. What percentage of all the seized American funds were seized in 1997?
- a) 29%
  - b) 78%
  - c) 80%
  - d) 99%

Questions 20 and 21 refer to the following scenario:

After a drug investigation, police officers seized the following items from three arrested persons:

Arrested person	Amount of cocaine	Number of \$100 bills	Number of \$50 bills	Number of \$20 bills
Smith	2.6 kg	111	45	34
Hasek	0.5 kg	35	21	26
Ramji	1.8 kg	20	135	5

Assume 1 gm of cocaine is worth \$90.

20. How much cash did the officers take from suspect Hasek?
  - a) \$4,070
  - b) \$5,070
  - c) \$13,940
  - d) \$49,070
21. What was the value of all the items taken from all the arrested persons?
  - a) \$378,750
  - b) \$459,905
  - c) \$468,950
  - d) \$514,440
22. Where was the suspect's tattoo? (From the mug shot previously memorized)
  - a) Upper left bicep
  - b) Upper left thigh
  - c) Upper right bicep
  - d) Upper right thigh
23. What colour were the suspect's eyes?
  - a) Brown
  - b) Green
  - c) Blue
  - d) Hazel

## Appendix E

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

**Arvey et al., (1990) Anxiety**

I probably didn't do as well as most of the other people who took this test \_\_\_\_\_

I am not good at taking tests \_\_\_\_\_

During the testing, I often thought about how poorly I was doing \_\_\_\_\_

I usually get very anxious about taking tests \_\_\_\_\_

I usually do well on tests \_\_\_\_\_

I expect to be among the people who score really well on this test \_\_\_\_\_

My test scores don't usually reflect my true abilities \_\_\_\_\_

I very much dislike taking tests of this type \_\_\_\_\_

During this test, I found myself thinking of the consequences of failing \_\_\_\_\_

During the testing, I got so nervous I couldn't do as well as I should have \_\_\_\_\_

**Day & Ziemer (2003) Anxiety**

Writing the test made me feel nervous. \_\_\_\_\_

I felt very comfortable writing this test \_\_\_\_\_

I believe that my heart rate increased during this test \_\_\_\_\_

Overall, I felt anxious performing this test \_\_\_\_\_

## Appendix F

1	2	3	4	5	6	7
Strongly Disagree	Disagree	Slightly Disagree	Neutral	Slightly Agree	Agree	Strongly Agree

**Arvey et al., (1990) Test-Taking Motivation**

Doing well on this test is important to me \_\_\_\_\_

I want to be among the top scorers on this test \_\_\_\_\_

I wanted to do well on this test \_\_\_\_\_

I tried my best on this test \_\_\_\_\_

I tried to do the very best I could do on this test \_\_\_\_\_

While taking this test, I concentrated and tried to do well \_\_\_\_\_

I pushed myself to work hard on this test \_\_\_\_\_

I was extremely motivated to do well on this test \_\_\_\_\_

I just didn't care how I did on this test \_\_\_\_\_

I didn't put much effort into this test \_\_\_\_\_

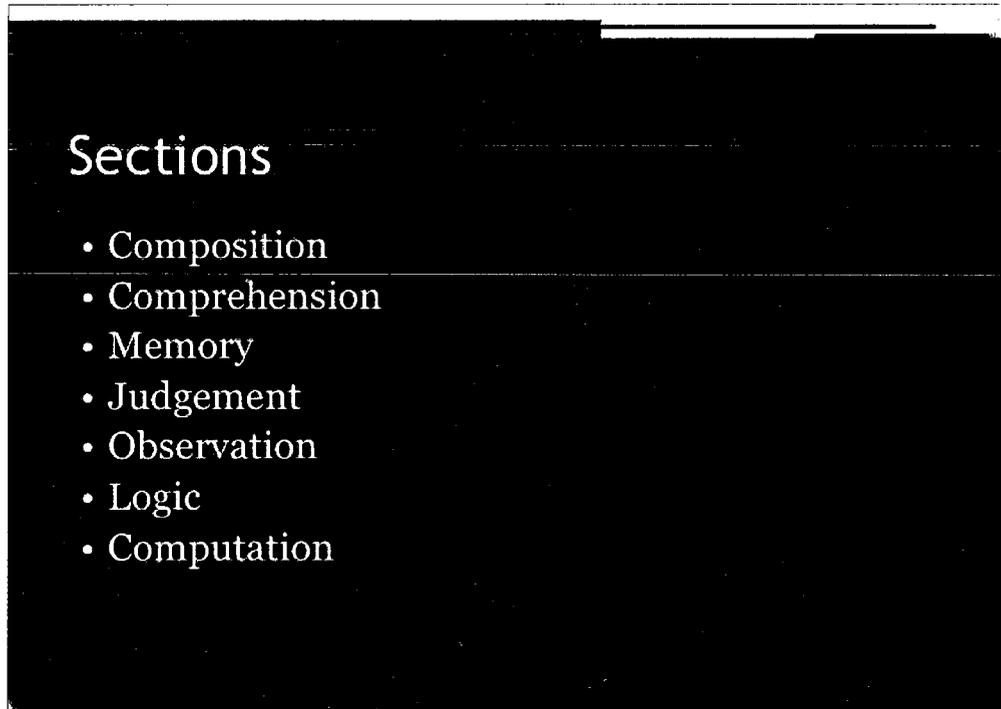
Appendix G

## How to Practice for the Tests



### Sections

- Composition
- Comprehension
- Memory
- Judgement
- Observation
- Logic
- Computation



## Composition

- Composition measures your knowledge of grammar, spelling, and vocabulary.
- How to improve:
  - Read! Look up words if you don't know them
  - Try to spot spelling mistakes
  - Familiarize yourself with dictionaries and thesauruses
  - Use memory aids to help you remember the correct spelling.
    - For example, "I before e except after c"
  - Word a day
  - Practice writing/spelling with a friend dictating
  - Do crossword puzzles

## Comprehension

- Comprehension measures your ability to read and accurately interpret written material
- How to improve:
  - Read!
    - Challenge your reading level
  - Do book club-style analysis on what you have read with others

## Comprehension

In 1992, approximately \$131 billion was spent on American advertisement. The Advertising Standards Authority (ASA) recommends to the media that advertisements must follow the British Code of Advertising Practice. If ads do not follow this code they are not to be published. Advertisement is also seen by economists as either a boost or a hindrance to perfect competition, since it attempts to make illusory distinctions between essentially similar products. Money spent on advertising has increased dramatically in recent years. In 2007, spending on advertisements has been estimated at \$150 billion in the United States and \$385 billion worldwide. Top spending industries in the United States are automobile (\$3.6 bn), food (\$1.78 bn), toiletries (\$537 m), entertainment (\$1.1 bn), telephone (\$733 m), alcohol (\$552 m), retailers (\$2.9 bn) and medicines (\$525 m).

- Some economists dislike advertising because:
- In 1992, the amount spent on global advertisement was:

## Comprehension

In 1992, approximately \$131 billion was spent on American advertisement. The Advertising Standards Authority (ASA) recommends to the media that advertisements must follow the British Code of Advertising Practice. If ads do not follow this code they are not to be published. Advertisement is also seen by economists as either a boost or a hindrance to perfect competition, since it attempts to make illusory distinctions between essentially similar products. Money spent on advertising has increased dramatically in recent years. In 2007, spending on advertisements has been estimated at \$150 billion in the United States and \$385 billion worldwide. Top spending industries in the United States are automobile (\$3.6 bn), food (\$1.78 bn), toiletries (\$537 m), entertainment (\$1.1 bn), telephone (\$733 m), alcohol (\$552 m), retailers (\$2.9 bn) and medicines (\$525 m).

- In the United States, the industry which spends most on advertisement (from the selection) is:
  - Food
  - Telephone
  - Finance
  - Entertainment
- In the United States, the industry which spends the least on advertisement (from the selection) is:
  - Toiletries
  - Telephone
  - Alcohol
  - Medicine

## Memory

- Memory measures your ability to remember over a period of time.
- How to improve:
  - Read the newspaper (particularly people's photographs)
    - Test yourself 30 minutes later
- Draw direct links between what you are trying to memorize and things that are familiar to you.

## Judgement

- Judgement measures your ability to use strategies to reach sound decisions
  - (i.e., common sense)
- These questions are worth double!
- How to improve:
  - Think about the possible consequences of taking/not taking each actions listed.
  - Take note of any legal situations in the news
    - Notice the actions taken, the rationale for these actions, and the public responses to the actions.
- Good judgement is hard to improve
  - Expand your horizons so that you may understand the consequences of a set of plausible actions
  - Once you have improved your ability to predict possible consequences, improvement in common sense will follow.

## Judgement

### Example

Everyone in your work group has received a new computer except you. What would you do?

1. Assume it was a mistake and speak to your supervisor.
2. Confront your supervisor regarding why you are being treated unfairly.
3. Take a new computer from a co-worker's desk.
4. Complain to human resources.
5. Quit

## Judgement

### Example

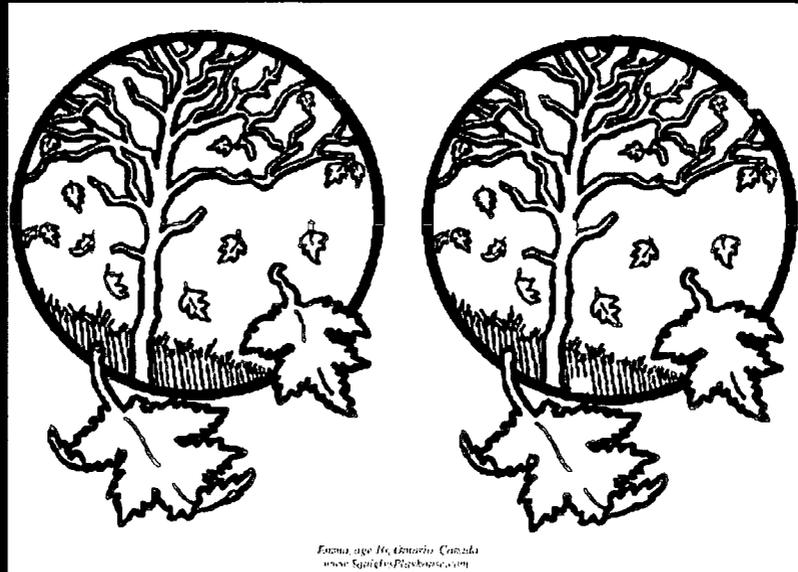
You have been assigned lead responsibility for two weeks in the absence of your supervisor. On your first day in this role, one of your new employees comes into your office and complains that they were sexually harassed by the security guard when they entered the building. They ask that the situation be kept confidential. What would be your first action in response to this situation?

1. Contact the security guard and conduct an interview to obtain all the facts
2. Assure the employee you will look into the situation but cannot guarantee confidentiality
3. Contact your supervisor to obtain instruction on next steps
4. Conduct informal interviews with your other employees to determine if they have been harassed

## Observation

- Observation measures your ability to attend to details in visual material
- How to improve:
  - Focus on features that are unlikely to be altered without plastic surgery (e.g., face and eye shape)
  - Use a process of elimination
  - “Spot the difference” puzzles
  - “Find Waldo” puzzles
  - Word finds
- Though these activities are not exactly the same, they allow you to practice using your observation skill.

## Observation



## Logic

- Logic measures your ability to analyze problems and situations using deductive and inductive processes
- How to improve:
  - Solve one piece of the puzzle at a time
  - In some instances, creating a diagram of the information on a piece of scrap paper may help in organizing your thoughts
  - Logic puzzle magazines
  - Adventure/Puzzle computer games
  - Read mystery novels and try to solve the crime before the hero does

## Logic

- James is taller than Kate and Carly. Sammy is shorter than Kate. Natalie is shorter than Kate and Sammy, however Sammy is shorter than Carly. Who is the shortest?
- Liam is younger than Mandy but older than Tara. Blake is older than Ruby who is older than Liam. Mandy is older than Blake. Who is the second youngest?

## Computation

- Computation measures your ability with basic computations
  - Add, subtract, multiply and divide
- How to improve:
  - Reread a high school math textbook
  - Practice math on a daily basis
    - (e.g., when paying for a product, calculate in your head how much change you should expect)
  - Do math *without* a calculator

## Additional Activities

- CBC's Test the Nation
- Computer games with english/math questions
- List of websites

- Computation
  - A website that allows you to set time limits and practice basic math
  - <http://www.shodor.org/interactivate/activities/ArithmeticQuiz/Logic/>
- Computation
  - A quiz with word math problems
  - <http://www.funtrivia.com/playquiz/quiz306193845b0.html>
- Logic
  - A quiz using inductive and deductive reasoning
  - <http://www.funtrivia.com/playquiz/quiz99127b5b6d0.html>
- Observation
  - A gallery of "Spot the Difference" pictures
  - <http://puzzles.about.com/od/opticalillusions/ig/SpotTheDifferencePuzzles/>
- Reading Comprehension
  - A timed quiz for reading a composition
  - <http://web2.uvcs.uvic.ca/courses/elc/studyzone/570/pulp/hemp3.htm>
- Reading Comprehension
  - Another timed quiz for reading a composition
  - <http://web2.uvcs.uvic.ca/courses/elc/studyzone/490/reading/tattoos2-reading.htm>
- Composition
  - A fun choose-the-correct-spelling game
  - [http://www.bbc.co.uk/hardspell/starspell\\_game.shtml](http://www.bbc.co.uk/hardspell/starspell_game.shtml)
- Composition
  - A database of help for all different areas of composition
  - <http://www.bbc.co.uk/skillswise/words/spelling/>
- Composition
  - Like Battleship but with commonly misspelled words
  - [http://www.quia.com/ba/1476.html?AP\\_rand=1243326807](http://www.quia.com/ba/1476.html?AP_rand=1243326807)

## Handout

Some general test taking strategies

- a. Listen to the instructions carefully. It is important that you clearly understand what you are expected to do. If you are unclear, ask questions before you start the exam. Not following the instructions provided will probably result in a lower obtained score.
- b. Read the questions carefully. Make sure you understand what the question is asking before answering.
- c. Try to arrive at an answer before looking over the choices. However, read all of the offered choices before selecting your final answer.
- d. If you are unsure of an answer, eliminate the options you know are wrong. Even if you can only eliminate one option, you will at least limit your guess to fewer options and, therefore, increase your chance of guessing correctly.
- e. Do not get hung up on any one question. If you are having difficulties with one particular question, make your best guess at the answer and move on. You can always come back later and spend more time with the question. If you waste time on a question to which you do not know the answer, you might not reach some

questions to which you do know the answer and, therefore, will not get these points.

- f. Never leave a question unanswered. Guess at any questions to which you do not know the answer. There is no penalty for a wrong answer.
- g. Bring a watch to the test. While the test administrator will occasionally announce the time remaining, it is better to have a time piece available so that you can better judge your time. Time management is a valuable test taking skill.
- h. Do not panic. Exams can be very stressful events. If you feel yourself getting too anxious during the test, take a couple of minutes to collect yourself. When you are ready, proceed with the test. It may be a good idea to move on from a question that is making you especially anxious and come back to it later.

### The Aptitude Test's components

#### 1) Composition

- This section will examine your ability to articulate, in a written format, complex thoughts in a clear and concise manner understandable to others. Specifically, this section will examine your knowledge of grammar, spelling, and vocabulary. Some of the sources used to create this section include *ITP Nelson Canadian Dictionary of the English Language: An Encyclopedic Reference (1997)*; *Webster's Ninth New Collegiate Dictionary (1984)*; Shaw's (1993) *Errors in English and Ways to Correct them*; and Strunk Jr. & White's (1979) *The Elements of Style*. Using these references may be helpful in improving your performance.
- In all, there will be three types of English Composition questions on the test: spelling, grammar, and vocabulary questions.
- Here are some activities that can help improve your English composition performance.
  - a. Read, read, and then read some more.
  - b. Familiarize yourself with the use of dictionaries and thesauruses. When reading, identify any words that you do not know and look up the definition of these words in a dictionary.
  - c. Use memory aids to help you remember the correct spelling. For example, mnemonics such as "I before e except after c" can be very useful.
  - d. Make it a point to learn at least one new word every day.
  - e. Practice with a friend by having him/her verbally dictate short written passages to you while you attempt to write out the passage. Afterwards, check your spelling.

- f. Try to spot spelling mistakes in newspapers or other documents. Confirm your findings by looking up the correct spelling in a dictionary.
- a. Do crossword puzzles found in newspapers, magazines, and puzzle books. Continue to challenge yourself by progressively completing more difficult puzzles.

## 2) Comprehension

- This section is designed to evaluate your ability to read and accurately interpret written material. You will be presented with a series of short passages. For each of these passages, two or three questions will be asked to determine if you have correctly interpreted the content of the particular passage.
- Here are some activities that can help improve your English comprehension performance:
  1. As with the composition exercises, read, read, and then read some more.
  2. It is important to challenge yourself to the reading level expected to adequately perform the job of police officer. A selection of magazines that have similar levels of reading as that on the RPAT include Canadian Geographic, Saturday Night, and some of the longer articles in Maclean's. Your local library should have a selection of these and similar magazines.
  3. Have a friend read an article or passage that you have also read. Discuss the contents of the text to confirm your interpretation of the message.

## 3) Memory

- This section will test your ability to memorize pictorial and textual materials over a period of time. You will be presented with a mug shot of an individual, along with their names, descriptions, and the crimes for which they are wanted. Your task is to memorize all of this information.
- Here are some activities that can help improve your memory
  1. Spend time memorizing textual material such as newspapers. Test yourself on what you remember 30 minutes later.
  2. Spend time memorizing the pictures of individuals in the newspaper, and the names of the people in the pictures. Test yourself after 30 minutes to see how well you did.

3. Have a friend note the makes, colour, and licence plates of a few cars in a parking lot. Memorize this information (giving yourself about 2 minutes to do so), then test yourself 30 minutes later.
4. Try to draw direct links between what you are trying to memorize and things that are familiar to you. Whenever possible, use imagery to help remember the information (i.e., form a picture of the information in your mind).

#### 4) Judgement

- This section will test your ability to use appropriate resources and strategies to achieve objectives. You demonstrate good judgement by reaching sound decisions and taking the appropriate courses of action.
- The term judgement is used synonymously with the concept of common sense, the Judgement questions are each worth 2 points. Keep this in mind when you are deciding how much time to assign to the Judgement questions.
- In responding to the questions, think about the possible consequences of taking/not taking each of the actions listed. Choose the option that you believe has the most positive set of consequences, or perhaps the least negative set of consequences.
- Here are some activities that can help improve your judgement performance:
  2. Take note of any policing situations that you may read about in the newspaper. Notice the actions taken, the rationale for these actions, and the public responses to the actions.
  3. The skill of good judgement or common sense is a difficult one to improve. The key is to expand your horizons so that you may understand the consequences of a set of plausible actions. Once you have improved your ability to predict possible consequences, improvement in common sense will follow.

#### 5) Observation

- This section will test your ability to attend to details in visual material. For each question, you will examine a set of sketches of faces. For each set of drawings, one of the drawings labelled “a”, “b”, “c”, or “d” will have near identical facial features as that identified as the original photo. Your task is to identify this drawing. In making your decision, you are to assume that no plastic surgery has occurred.
- To complete the task accurately, focus on features that are unlikely to be altered without plastic surgery. For example, examine the shape of the eyes, the size and

appearance of the nose, the shape of the face, etc. Do not be distracted by features that can easily be altered, such as hair style and colour, facial hair, and clothing.

- Use a process of elimination when completing the task. Eliminate any pictures where you notice a difference in a feature that should not change. Once three pictures have been eliminated, the remaining picture should be the correct answer.
- Here are some activities that can help improve your observation performance:
  1. “Spot the difference” puzzles often found in newspapers.
  2. “Find Waldo” puzzles.
  3. Word finds.
- Though these activities are not exactly the same as on the test, they do allow you to practice using your observation skill.

## 6) Logic

- This section will examine your ability to identify and to analyze problems and situations using deductive processes (ability to apply general rules to a problem and arrive at a logical answer) and inductive processes (ability to combine information in order to form general rules). This will be investigated using a wide variety of questions. You will be asked to complete tasks such as ordering pieces of information in a logical sequence, following directions on a map, determining the pattern that exists in a set of data, and solving problems.
- To solve many of the presented puzzles, it is important that you organize your thoughts. Solve one piece of the puzzle at a time: it is unlikely that a solution can be obtained by trying to examine all of the data at once. It is also important to organize the order in which each piece of the puzzle should be solved. The solutions to some aspects of the puzzles cannot be obtained without the solution to some other piece of the puzzle. In some instances, creating a diagram of the information on a piece of scrap paper may help in organizing your thoughts.
- Here are some activities that can help improve your logic performance:
  1. There are logic puzzle magazines that provide puzzles similar to those found on the RPAT.
  2. There are many computer games (e.g., adventure games, puzzle games) that make use of logic skills to solve them.
  3. Read mystery novels such as Arthur Conan Doyle’s Sherlock Holmes series. Try to solve the crime before the hero does.

## 7) Computation

- This section will evaluate your ability with basic computations. The section will examine if you know when to and how to add, subtract, multiply and divide. It will also evaluate some very basic algebra skills. The knowledge of mathematics

required to do well in this section is no higher than a grade nine level. You will not be allowed to use a calculator.

- Here are some activities that can help improve your computation performance:
  1. Reread a high school math textbook, which often can be obtained through a local library (no higher than the grade 9 level is required). Similarly, there are educational computer programs that focus on these math skills.
  2. Practice solving a variety of math problems on a daily basis (e.g., when paying for a product, calculate in your head how much change you should expect before the cashier gives it, or try to calculate the amount of tax that needs to be paid for a product prior to arriving to the cashier).

## Appendix H

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you.

1 Very Inaccurate	2 Moderately Inaccurate	3 Neither Accurate Not Inaccurate	4 Moderately Accurate	5 Very Accurate
-------------------------	-------------------------------	---	-----------------------------	--------------------

1.	I am the life of the party.	_____
2.	I insult people.	_____
3.	I am always prepared.	_____
4.	I get stressed out easily.	_____
5.	I have a rich vocabulary.	_____
6.	I often feel uncomfortable around others.	_____
7.	I am interested in people.	_____
8.	I leave my belongings around.	_____
9.	I am relaxed most of the time.	_____
10.	I have difficulty understanding abstract ideas.	_____
11.	I feel comfortable around people.	_____
12.	I am not interested in other people's problems.	_____
13.	I pay attention to details.	_____
14.	I worry about things.	_____
15.	I have a vivid imagination.	_____
16.	I keep in the background.	_____
17.	I sympathize with others' feelings.	_____
18.	I make a mess of things.	_____
19.	I seldom feel blue.	_____
20.	I am not interested in abstract ideas.	_____
21.	I start conversations.	_____
22.	I feel little concern for others.	_____

23. I get chores done right away.	
24. I am easily disturbed.	
25. I have excellent ideas.	
26. I have little to say.	
27. I have a soft heart.	
28. I often forget to put things back in their proper place.	
29. I am not easily bothered by things.	
30. I do not have a good imagination.	
31. I talk to a lot of different people at parties.	
32. I am not really interested in others.	
33. I like order.	
34. I get upset easily.	
35. I am quick to understand things.	
36. I don't like to draw attention to myself.	
37. I take time out for others.	
38. I shirk my duties.	
39. I rarely get irritated.	
40. I try to avoid complex people.	
41. I don't mind being the center of attention.	
42. I am hard to get to know.	
43. I follow a schedule.	
44. I change my mood a lot.	
45. I use difficult words.	
46. I am quiet around strangers.	
47. I feel others' emotions.	
48. I neglect my duties.	
49. I seldom get mad.	
50. I have difficulty imagining things.	
51. I make friends easily.	
52. I am indifferent to the feelings of others.	
53. I am exacting in my work.	
54. I have frequent mood swings.	
55. I spend time reflecting on things.	
56. I find it difficult to approach others.	
57. I make people feel at ease.	
58. I waste my time.	
59. I get irritated easily.	
60. I avoid difficult reading material.	
61. I take charge.	
62. I inquire about others' well-being.	
63. I do things according to a plan.	
64. I often feel blue.	

65.	I am full of ideas.	_____
66.	I don't talk a lot.	_____
67.	I know how to comfort others.	_____
68.	I do things in a half-way manner.	_____
69.	I get angry easily.	_____
70.	I will not probe deeply into a subject.	_____
71.	I know how to captivate people.	_____
72.	I love children.	_____
73.	I continue until everything is perfect.	_____
74.	I panic easily.	_____
75.	I carry the conversation to a higher level.	_____
76.	I bottle up my feelings.	_____
77.	I am on good terms with nearly everyone.	_____
78.	I find it difficult to get down to work.	_____
79.	I feel threatened easily.	_____
80.	I catch on to things quickly.	_____
81.	I feel at ease with people.	_____
82.	I have a good word for everyone.	_____
83.	I make plans and stick to them.	_____
84.	I get overwhelmed by emotions.	_____
85.	I can handle a lot of information.	_____
86.	I am a very private person.	_____
87.	I show my gratitude.	_____
88.	I leave a mess in my room.	_____
89.	I take offense easily.	_____
90.	I am good at many things.	_____
91.	I wait for others to lead the way.	_____
92.	I think of others first.	_____
93.	I love order and regularity.	_____
94.	I get caught up in my problems.	_____
95.	I love to read challenging material.	_____
96.	I am skilled in handling social situations.	_____
97.	I love to help others.	_____
98.	I like to tidy up.	_____
99.	I grumble about things.	_____
100.	I love to think up new ways of doing things.	_____

Appendix I

1 Strongly Disagree	2 Disagree	3 Slightly Disagree	4 Neutral	5 Slightly Agree	6 Agree	7 Strongly Agree
---------------------------	---------------	---------------------------	--------------	------------------------	------------	------------------------

**Manipulation Check**

I read the e-mails sent to me about this study thoroughly \_\_\_\_\_

**Arvey et al. (1990) Preparation**

I spent a good deal of time preparing for this test \_\_\_\_\_

I prepared a lot for this test \_\_\_\_\_

I used outside sources to study for this test \_\_\_\_\_

**Test-Taking Familiarity**

I have taken tests similar to this test before \_\_\_\_\_

I take tests similar to this test often \_\_\_\_\_

I have taken an employment test before \_\_\_\_\_

I am familiar with different types of test formats \_\_\_\_\_

I recognized the items on this test from the first study session \_\_\_\_\_

**Please read and answer the following questions**

If you used outside materials to study for this test

a) How many sources (books, websites, etc.) did you use to study? \_\_\_\_\_

b) And how many hours did you study for? \_\_\_\_\_

Have you ever applied to any police organization?

Yes                      No

If yes, how many times? \_\_\_\_\_

Are you considering applying to any police organization?

Yes                      No