Canadian Forces Aptitude Test: Repeated Assessment and Practice Effect

Alla Skomorovsky
Selection and Assessment
Directorate Military Personnel Operational Research and Analysis

Defence R&D Canada
Director General Military Personnel Research & Analysis

Chief Military Personnel

National Defence   Défense nationale

Canadian Forces Aptitude Test: Repeated Assessment and Practice Effect

Alla Skomorovsky
Selection and Assessment
Directorate Military Personnel Operational Research and Analysis

Director General Military Personnel Research & Analysis

Technical Memorandum
DGMPRA TM 2009-003
May 2009

Author

(Original signed by)

Alla Skomorovsky, PhD

Approved by

(Original signed by)

Catherine Campbell, MASc

Section Head – Military Personnel Operational Research and Analysis

Approved for release by

(Original signed by)

Kelly Farley, PhD

Chief Scientist – Director General Military Personnel Research and Analysis

The opinions expressed in this paper are those of the authors and should not be interpreted as the official position of the Canadian Forces, nor of the Department of National Defence.

Abstract

The Canadian Forces Aptitude Test (CFAT) retest policy required by Personnel Psychology Directive (PPD) 203 was changed on January 2007. Specifically, the length of the test-retest interval following the initial assessment was reduced from three months to seven days. Previous research suggests that when the length of the test-retest interval is short, an increase in test score on retest may occur due to criterion-unrelated variance (i.e., practice effect). This study examines the impact of reducing the length of the test-retest interval to seven days. Results demonstrate a significant increase in CFAT scores seven days after the initial assessment, which is greater than the increase occurring three months following the initial assessment. It is recommended that a minimum three-month period be set between initial selection testing and retest.

Résumé

La politique de reprise du Test d'aptitude des Forces canadiennes (TAFC) exigée dans la Directive de psychologie du personnel (DPP) 203 a été modifiée en janvier 2007. Plus précisément, la longueur de l'intervalle test-retest après l'évaluation initiale est passée de trois mois à sept jours. Selon des recherches antérieures, lorsque l'intervalle test-retest est court, les scores obtenus à la reprise peuvent augmenter à cause d'une variance non liée à un critère (c.-à-d. effet lié à la pratique). Dans la présente étude, nous avons examiné l'impact de la réduction à sept jours de l'intervalle test-retest. Comme les résultats le montrent, les scores au TAFC augmentent grandement sept jours après l'évaluation initiale, et cette augmentation est supérieure à celle observée trois mois après le test initial. Nous recommandons de retenir l'intervalle minimal de trois mois entre le test initial et la reprise afin d'éviter l'effet lié à la pratique.

This page intentionally left blank.

Executive summary

Canadian Forces Aptitude Test: Repeated Assessment and Practice Effect:

Alla Skomorovsky; DGMPRA TM 2009-003; Defence R&D Canada – DGMPRA; May 2009.

There is a necessity in organizational settings to implement a cognitive test-retest policy for personnel selection purposes. According to Personnel Psychology Directive (PPD) 203 (1996), a policy change decision was made, such that candidates are eligible for a Canadian Forces Aptitude Test (CFAT) retest three months following the initial assessment rather than one year after the initial assessment. The rationale for this policy change was not related to the potential change in cognitive functioning within such a short period of time, but to allow a candidate to demonstrate his or her true abilities if there were certain transient limitations on original testing (e.g., the candidate's illness) (PPD 203, 1996).

In January 2007, the decision was made to reduce the length of the test-retest interval to seven days. Such a reduction of the length of test-retest interval has raised a concern about the validity of CFAT scores on retest. According to previous research, allowing a retest brings a potential problem of practice effect into play. Specifically, candidates may perform better on retest due to criterion unrelated variance (e.g., learning tricks, memorizing items) rather than to any actual improvement in cognitive abilities. This paper provides a review of research in the area of retesting and practice effect and presents the results of a study that examined the impact of a shorter test-retest interval on retest scores.

In this study, CFAT scores at initial testing were compared to those obtained on retest. Some individuals were retested at three months or later, while others were retested less than three months following the initial assessment. Furthermore, it was examined whether performance on retest was a function of time interval following initial assessment. Finally, an analysis was conducted to determine whether poorer performance at initial testing predicted better performance on retest.

The results of the study were consistent with previous research. Although there is an increase in CFAT scores on retest at both short and longer (three months or more) time intervals, the increase is larger when retest occurs earlier than three months following the initial assessment. Furthermore, the increase in the scores is greatest for candidates who took the retest exactly seven days after the initial assessment, as compared to the candidates who took the retest in the period between seven days and three months. These results suggest the presence of a practice effect particularly at the seven-day mark that decreases with the passage of time following initial testing.

The follow up regression analyses demonstrated that the length of the interval following initial assessment significantly predicted performance on retest for the candidates who had the retest less than three months after the initial assessment only. Length of time following the initial assessment did not significantly predict performance on retest for the individuals who were retested three months or more following the initial assessment. This initial increase in CFAT

scores, followed by a gradual decrease with the passage of time following initial assessment, can be explained by a practice effect that is stronger immediately following the initial assessment. These findings are consistent with previous research conducted in the area of retest for selection purposes. Specifically, Falleti, Maruff, Collie, and Darby (1996) also found an increase in test scores on retest one week after the initial assessment.

Finally, regression analyses indicated that individuals with lower scores on CFAT at initial testing benefit more from being retested (i.e., improve their scores more) than do those who obtained higher scores. It is unlikely that this improvement reflects changes in actual abilities among individuals with lower test scores, especially since the improvement in test scores was found to diminish with the passage of time. It seems that the improvement reflects a practice effect, which diminishes as time passes.

The current study's findings demonstrate that test scores on the CFAT increase upon retest. Moreover, the shorter the time interval following the initial assessment, the greater the increase in scores, and the more likely it is that the increase is related to criterion unrelated variance (i.e., practice effect). Specifically, the increase in CFAT scores seven days after the initial assessment is greater and involves greater practice effect than does the retest three months or more after initial assessment. When the retest is administered shortly after the initial assessment, individuals with lower scores on the initial cognitive testing were found to benefit the most from practice effect. It is likely that if the CFAT is re-administered within a short time following the initial test, particularly at the seven-day point, greater numbers of individuals with low true cognitive test scores may be accepted into the CF.

To conclude, the CFAT is used to predict future training success and job performance.

When candidates take a retest too soon, their scores are inflated (i.e., their ability appears to be higher than it actually is because of the practice effect). If large numbers of candidates are selected into the CF based on the inflated erroneous retest scores, failure rates on training will increase, resulting in increased training costs for the CF and an increase in the time it takes to meet CF training establishments. While a three-month interval still inflates scores to some degree (a one-year interval would be ideal), it does not have as serious inflation as a one-week interval would. It is highly recommended, therefore, for reasons of Force structure, that the interval between first test and retest be set at a minimum of three months.

Sommaire

Canadian Forces Aptitude Test: Repeated Assessment and Practice Effect:

Alla Skomorovsky; DGMPRA TM 2009-003; R & D pour la défense Canada – DRASPM; Mai 2009.

En milieu organisationnel, il est nécessaire de mettre en œuvre une politique de réévaluation pour les tests de la capacité cognitive à des fins de sélection du personnel. Selon la Directive de psychologie du personnel (DPP) 203 (1996), une décision a été prise concernant le changement de politique : les candidats sont maintenant admissibles à une reprise du Test d'aptitude des Forces canadiennes (TAFC) trois mois après l'évaluation initiale plutôt qu'un an après. Ce changement de politique ne se fondait pas sur un changement possible dans le fonctionnement intellectuel au cours d'une période aussi restreinte, mais voulait permettre à un candidat de démontrer ses vraies aptitudes au cas où celles-ci auraient été limitées temporairement lors du test original (p. ex. maladie du candidat) (DPP 203, 1996).

En janvier 2007, on a décidé de réduire la longueur de l'intervalle test-retest à sept jours. Une telle réduction de l'intervalle test-retest a suscité des inquiétudes concernant la validité des scores à la reprise du TAFC. Selon des recherches antérieures, le fait de permettre une reprise du test soulève le problème de l'effet lié à la pratique. Plus précisément, les candidats peuvent avoir une meilleure performance à la reprise du test à cause d'une variance non liée à un critère (p. ex. trucs d'apprentissage, items mémorisés) plutôt que d'une amélioration réelle de leur capacité cognitive. Le présent document passe en revue les recherches dans le domaine de la réévaluation et de l'effet lié à la pratique et présente les résultats d'une étude portant sur l'impact d'un plus court intervalle test-retest sur les scores obtenus au deuxième test.

Dans cette étude, les scores au TAFC initial ont été comparés à ceux obtenus lors de la reprise. Certaines personnes ont subi un nouveau test après trois mois ou plus, alors que d'autres ont été réévaluées moins de trois mois après le test initial. Nous avons en outre examiné si la performance à la reprise était fonction de l'intervalle entre les deux tests. Enfin, nous avons effectué une analyse pour déterminer si une moins bonne performance au test initial était un prédicteur d'une meilleure performance à la reprise.

Les résultats de l'étude concordaient avec les conclusions de recherches antérieures. Bien que les scores au TAFC aient augmenté lors de la reprise après un intervalle court et long (trois mois ou plus), l'augmentation était plus importante lorsque le deuxième test avait lieu moins de trois mois après l'évaluation initiale. De plus, l'augmentation des scores était la plus élevée chez les candidats qui avaient repris le test sept jours exactement après l'évaluation initiale, comparativement aux candidats qui avaient attendu entre sept jours et trois mois. Ces résultats évoquent l'existence d'un effet lié à la pratique, notamment sept jours après le test, effet qui diminue plus l'intervalle test-retest est long.

Les analyses ultérieures de régression ont montré que la longueur de l'intervalle après l'évaluation initiale est un prédicteur significatif de la performance à la reprise du test uniquement chez les candidats qui avaient été réévalués moins de trois mois après le test initial. Le laps de

temps écoulé après l'évaluation initiale n'était pas un prédicteur significatif de la performance à la reprise chez les personnes qui avaient été réévaluées trois mois ou plus après le test initial. L'augmentation initiale des scores au TAFC, suivie d'une diminution graduelle avec le temps, peut s'expliquer par un effet lié à la pratique, qui est plus puissant tout de suite après l'évaluation initiale. Ces conclusions concordent avec celles de recherches antérieures dans le domaine de la réévaluation à des fins de sélection. Plus précisément, Falleti, Maruff, Collie et Darby (1996) ont eux aussi constaté une augmentation dans les scores lors de la reprise d'un test une semaine après l'évaluation initiale.

Enfin, les analyses de régression ont indiqué que les personnes ayant obtenu des scores plus faibles au TAFC initial obtenaient de meilleurs scores à la reprise que celles qui avaient obtenu des scores plus élevés au départ. Il est peu probable que cette amélioration soit due à des changements dans les capacités réelles des personnes ayant obtenu des scores plus faibles, vu notamment que l'amélioration des scores faiblissait avec le temps. Il semble que l'amélioration soit attribuable à un effet lié à la pratique, qui s'atténue au fil du temps.

En conclusion, les résultats de la présente étude montrent que les scores obtenus au TAFC augmentent lors de la reprise. Par ailleurs, plus l'intervalle est court entre les deux tests, plus l'augmentation dans les scores est importante et plus il est probable que l'augmentation soit due à une variance non liée à un critère (c.-à-d. effet lié à la pratique). En particulier, l'augmentation des scores au TAFC sept jours après l'évaluation initiale est plus marquée et l'effet lié à la pratique est plus important qu'à la reprise du test trois mois ou plus après l'évaluation initiale. Lorsque le deuxième test est administré peu après le test initial, les personnes ayant obtenu des scores plus faibles au test cognitif initial profitaient le plus de l'effet lié à la pratique. Il est probable que si le TAFC est administré une deuxième fois peu de temps après le test initial, notamment sept jours plus tard, un nombre plus élevé de personnes ayant de faibles scores réels au test cognitif risquent d'être acceptées dans les FC.

Nous recommandons de maintenir l'intervalle minimal de trois mois entre le test initial et la reprise afin d'éviter l'effet potentiel lié à la pratique.

Table of contents

List of tables

1   Introduction

The field of organizational psychology has developed a growing literature dealing with the validation of standardized cognitive testing. It is common that a cognitive test is administered to the same individual on more than one occasion. For instance, cognitive tests are administered multiple times to examine short-term cognitive changes associated with diseases or surgeries. In the area of personnel selection, there is a growing need for cognitive tests to be administered more than one time. According to Lievens, Buyse, and Sackett (2005), there are at least two reasons to install a retesting policy in an organization. The first reason concerns the "transient characteristics of the applicant at the time of testing (e.g., illness, disability)", whereas the second concerns the transient characteristics of the testing situation (e.g., deviations from standardized test administration procedures) or random measurement error[1]. It is possible that a person did not demonstrate his or her true abilities to the full extent the first time and the person is given another chance to do that. Therefore, most organizations have installed retesting policies in the personnel selection sphere.

There are several conceptual concerns that arise from implementing a retest policy for personnel selection. If the candidate improves his/her score on retesting, which set of scores (initial assessment or retest/s) is most related to the criterion of interest (i.e., job performance) (Lievens *et al.*, 2005)? This question in turn leads to practical concerns: which score should be used for selection decisions and how is the candidate ranked for selection purposes? Researchers started to raise these concerns when their findings demonstrated that repeated assessment regularly leads to a change or an improvement in performance. The researchers' question is, however, whether the improvement in performance is due to true change in cognitive functioning or due to the repeated assessment itself (i.e., practice effect). While there is a continuing need for information regarding test-retest usefulness and concerns about the performance improvements on retest, there is very little information on the estimates of a practice effect when considering longer or shorter test-retest intervals (Dikmen, Heaton, Grant, and Temkin, 1999). This paper discusses the reasons for score changes on retesting in general and their implications for the validity of a test. In addition, this paper reports the results of a study conducted to examine the impact of different test-retest intervals on performance on retest of the Canadian Forces Aptitude Test (CFAT).

1.1   Reasons for Score Changes on Retesting and Implications for Validity: Practice Effect

According to Lievens *et al.* (2005), there are several underlying reasons for changes in performance. The first reason, measurement error, may result in either higher or lower scores on retesting. Second, increases in test scores may reflect a true improvement of the person's cognitive characteristics. Anastasi and Urbina (1997) argue that cognitive tests are more difficult to improve than knowledge tests in a short period of time. This argument suggests that true improvement of the candidate's standing during the short interval between two administrations is more likely to be the case in knowledge tests than in cognitive

---

[1]   Measurement error is the variation between measurements of the same quantity on the same individual that is regularly assessed as within-subject standard deviation (Bland & Altman, 1996)

abilities tests. Third, an individual's deficit, stress or other negative circumstances, present at initial testing, may not be present on retest. And finally, practice effect (i.e., memorizing items, learning tricks, recall of repeated items) may be the fourth reason for the change in the scores on retesting. These reasons have different impacts on test-retest validity (see Table 1).

Table 1: Score Changes on Retesting

Underlying Reasons for Score Changes on Retesting	Effects on Validity (within-person effects only)
1. Measurement error	Equal validity for initial test and retest
2. True change in the construct of interest	Higher validity for retest (excluding the cases where the change is constant for all test takers and the validity is unchanged)
3. Criterion-related change (reduction of stress or disability)	Higher validity for retest than for initial test
4. Criterion-irrelevant change (practice effect)	Lower validity for retest (excluding the cases where change is constant for all test takers, in which validity is unchanged)

Adapted from Lievens *et al.* (2005)

When a change in score occurs on retesting, in most of the cases, retest has equal or higher validity than an initial test. Nevertheless, if the change on retest occurs due to criterion unrelated changes (e.g., the candidate remembered some questions or learned a strategy, which helped him/her on retest), the validity for retest is lower than for the initial test. Furthermore, in this case, retest has lower validity for individuals who repeated the test, as compared to one-time test takers. Indeed, in most selection situations, individuals who do well the first time do not retake the test, while individuals who take the retest regularly performed more poorly on the initial test. Such a comparison negatively influences the selection decisions not only for those who took the retest but also for those who took it only once. In other words, increases in performance on a cognitive ability test happen due to practice or coaching rather than true performance improvement. This is referred to as *practice effect*. Practice effect may inflate or obscure meaningful changes on retest and, therefore, is an important factor to consider in making retesting policy (Theisen, Rapport, Axelrod, and Brines, 1998).

There have been multiple studies that demonstrate the presence of a practice effect with cognitive tests. Kulik, Kulik, and Bangert (1984) examined practice effect in cognitive testing, demonstrating a medium effect size of .42 for identical tests. More recently, Lievens *et al.* (2005) demonstrated that retaking a test would lead to significantly higher scores on all tests, including cognitive tests, and found the same effect size of .42. Matarazzo, Carmody, and Jacobs (1980) conducted a meta-analysis of cognitive functioning (measured by Wechsler Adult Intelligence Scale [WAIS]) with test-retest interval ranging between one week and 10 years. The meta-analysis findings suggested that an average practice effect of five points is expected on retesting. Researchers concluded that repeated assessment with the same test would lead to an improvement in performance on retest (Bornstein, Baker, and Douglass, 1987; Goldstein and Watson, 1989; Johnson, Hoch, and Johnson, 1991). Such improvement, according to the researchers, would be due to a practice effect (Lievens *et al.,* 2005; Temkin, Heaton, Grant, and Dikmen, 1999).

The extent of a practice effect is a function of several factors. According to Bornstein, Baker, and Douglass (1987) and Lezak (1995), tests that require an unfamiliar or infrequently practiced response, as well as tests that have a single solution, are likely to show a larger practice effect. Furthermore, research demonstrated that practice effect is greater when a cognitive test involves discovery of a strategy (Lowe and Rabbit, 1998).

There is some controversy over whether individuals with lower cognitive ability scores benefit more or less on retest as compared to individuals with higher cognitive ability scores. One set of research demonstrated that individuals with lower scores on cognitive ability tests gained more on retest than individuals with higher scores (e.g., Lowe and Rabbit, 1998). Conversely, other researchers (Rapport, Brines, Axelrod, and Theisen, 1997) found that individuals with average or higher than average scores on cognitive ability tests made greater gains on repeated testing than did those with lower than average scores.

There has been data collected on test-retest reliability of many instruments used to assess cognitive performance; however, only the reliability coefficients were reported (McCaffrey, Ortega, Orsillo, and Nelles, 1992). While this is useful information from a psychometric perspective, it does not differentiate between the true improvement and the score change due to a practice effect. For instance, if a job candidate received a score of 50 on the first test and a score of 60 on retest, the reliability coefficient does not explain whether this 10-point difference between the original test and retest was due to improvement in the individual's cognitive ability or solely to repeated testing. Therefore, information on test-retest reliability is not sufficient to detect or manage a practice effect.

1.2 Methods to Reduce Practice Effect

There have been several ways proposed to manage a potential practice effect. One method for reducing effects of practice is to develop alternative forms of the same cognitive test. The practice effect should be lower on a different version of the test, when individuals have not had experience with the test items (McCaffrey, Ortega, Orsillo, Nelles, and Haase, 1992). Anastasi (1988), however, found some improvements on retest as compared to original testing using parallel forms of the same test. It seems that if a cognitive test requires an individual to learn a strategy or rule (e.g., learning a synonym method in the verbal subscale of the CFAT), even alternative forms may not protect against practice effect (Basso, Bornstein, and Lang, 1999; Kay and Kane, 1991; Lowe and Rabbitt, 1998). Kulik, Kulik, and Bangert (1984) examined practice effect in cognitive testing, demonstrating an effect size of .23 for parallel tests[2]. Several studies demonstrate that practice effect may occur even if parallel forms are used for testing, since the format of the instrument remains the same and familiarity with task demands and cognitive strategies employed are generalizable (Anastasi, 1988; Crook, Youngjohn, and Larabee, 1992; Youngjohn and Crook, 1993). Uchiyama, D'Elia, Dellinger, and Becker (1995) and Watson, Pasteur, Healy, and Hughes (1994) concluded that a second administration of an alternative form of the measure is likely to result in improved performance.

---

[2]   An effect size of .23 is considered to be small in the literature (Cohen, 1988).

Another alternative proposed to counter the effect of practice is to administer the entire cognitive test twice to every individual on a regular basis (McCaffrey *et al.,* 1992). According to this method, the score obtained on the second testing should be used to reflect the individual's true cognitive ability. The main limitation of this method is that it is a costly strategy, which requires a great amount of researchers' time to administer tests and analyze results.

Another approach developed to counter the effect of practice is to utilize an adjustment for repeated administration (Temkin, Heaton, Grant, and Dikmen, 1999; Bruggemans, Van de Vijver, and Huysmans, 1997). According to this method, if the observed change is lower than a certain adjustment point it is believed to be a practice effect, while if it is higher than this point, it is considered a true change. Shatz (1981) proposes to use the standard error of measurement to set up confidence intervals around an individual's score in order to partial out practice effect from a retest score. The major limitation with this method is that the magnitude of the practice effect is not stable and varies as a function of the difficulty of a test (Basso, Bornstein, and Lang, 1999).

Moreover, a practice effect is not stable for the same test and depends on the length of the test-retest interval used (Benedict and Zdaljardic, 1998). While other variables, such as the general ability level at the time of initial testing, can also influence the magnitude of the practice effect, the length of the test-retest interval seems to be the main influencing factor (Dikmen, Heaton, Grant, and Temkin, 1999).

There has been very little research examining the effects of a short test-retest interval. Nevertheless, the available research demonstrates that the shorter the interval between the first and second test, the greater the magnitude of a practice effect. While a practice effect was not observed on simple tasks, it was found to be significant on the tasks that were more difficult to do (Falleti, Maruff, Collie and Darby, 2006). Falleti *et al.* (2006) examined repeated assessment of cognitive functioning among healthy young adults (18-40 years) and found that the practice effect observed between the first two assessments reflected the extent to which the individuals "were able to acquire, understand and adhere to the requirements of the different tests rather than reflecting any improvement in the cognitive functioning measured" (p. 1107). Falleti *et al.* (2006) demonstrated that while the practice effect was moderately high when retest occurred one week after initial testing, no significant practice effect was observed when the test-retest interval had been increased to one month. In other words, when individuals are retested one month after the initial assessment, improvement in performance reflects their actual levels of cognitive ability.

Carretta, Zelenski and Ree (2000), however, came to a more conservative conclusion examining the impact of different test-retest intervals on the magnitude of a practice effect. Healthy young adults ($N = 477$) received a test battery of cognitive abilities that contributes to a U.S. Air Force pilot selection composite known as the Pilot Candidate Selection Method. These individuals were retested two weeks, three months, and six months following initial testing. While 70% of the individuals tested demonstrated some improvements on retest regardless of the length of the test-retest interval, the magnitude of the practice effect diminished as the length of the test-retest interval increased. According to Carretta Zelenski and Ree (2000), retest on a cognitive battery could be permitted no earlier than six months after initial testing. The results of these studies demonstrated that while the length of the test-retest interval is the main factor influencing the magnitude of the practice effect, the exact length recommended for cognitive testing is not clear and may vary across tasks.

2   The Canadian Forces Aptitude Test

Currently, a psychometric approach to assessing cognitive abilities is highly accepted in the employment and selection area, as it can provide valuable information regarding the potential job performance of a candidate (Ree, Earles and Teachout, 1994). Similarly, cognitive testing is prevalent in the military sphere, playing an important role in selection and placement (Carretta, Zalenski, and Ree, 2000). In the CF, the cognitive abilities of potential recruits are tested by the Canadian Forces Aptitude Test (CFAT). The CFAT was found to predict occupational performance in numerous studies (Girard, 2004; Hodgson, 2005; MacLennan, 1997; Scholtz, 2004; Woychesin, 1999).

The CFAT is a 60-item standardized test of general cognitive ability. It is a timed test arranged in ascending order of difficulty. It is comprised of three subscales: verbal skills (15 items), spatial abilities (15 items), and problem-solving abilities (30 items). The verbal skills, spatial abilities, and problem-solving abilities subscales of the test were found to have moderate-high internal consistency reliability, where verbal skills alphas ranged between .78 and .87, spatial abilities alphas between .64 and .88, and problem-solving abilities alphas between .88 and .91 (Black, 1999; Vanderpool, 2003). The verbal skills scale assesses a candidate's ability to comprehend text and understand the use of words. The spatial abilities scale is a non-verbal measure that evaluates a candidate's ability to deal with complex geometrical figures. The problem-solving scale measures a candidate's ability to use mathematical skills in solving problems (Vanderpool, 2003). To be selected into the CF, non-commissioned members and officer applicants must achieve a minimum cut off score. Furthermore, to be classified into a given military occupation, applicants must achieve the specific minimum score for that particular occupation.

According to Personnel Psychology Directive 203 (PPD, 1996), a policy change decision was made such that candidates are eligible for a CFAT retest three months, rather than one year, after the initial assessment. The rationale for this policy change was not related to the potential change in cognitive functioning within such a short period of time. Rather, the rationale for the change was to allow a candidate to demonstrate his or her true abilities if there were certain transient limitations on original testing (PPD 203, 1996).

More recently, in January 2007, due to the increasing need for additional personnel in the CF, the retest policy was changed, the decision was made to reduce the length of the test-retest time interval to seven days. Because previous research, examining the impact of a short test-retest interval, has suggested that reducing the test-retest interval to seven days might result in an increased practice effect (e.g., Carretta Zelenski and Ree, 2000; Falleti *et al.,* 2006), such a reduction in the length of the test-retest interval has raised a concern about the usefulness of CFAT scores on retest. This study was conducted to compare the potential changes in scores among CF candidates who took CFAT retest since the seven-day policy came into place.

2.1　Hypotheses

Based on the previous research, the hypotheses of this study were:

    a.   CFAT scores on retest will be higher than those on initial testing.

    b.   The increase in CFAT scores when test-retest interval is short (less than three months following initial testing) will be greater than when test-retest interval is longer than three months.

    c.   CFAT scores on retest will be a function of the length of the interval following initial testing.

    d.   Individuals who perform more poorly at the initial test will gain more from retest than those who perform better.

3   Methods

3.1   Participants

CFAT data for candidates who might have received a retest since the retest policy change took place (since January 17th 2007) were analyzed. There were 16,847 entries in the dataset. Only entries for candidates who took the CFAT at least twice were left for further analyses. Finally, after deleting five wrong entries (double service numbers, wrong dates), 708 entries remained in the dataset. Among them there were 599 candidates who took CFAT retest less than three months since the initial assessment and 111 candidates who took CFAT retest 90 days or more since the initial test. There were only 25 candidates who took retest exactly seven days after the initial assessment, while the days for other candidates ranged between eight and 89 days ($M = 29.6$[3], $SD = 19.2$). Given such a small number of individuals who were retested exactly seven days after the initial assessment, this group was combined with the group of individuals who took the retest in the interval between seven days and three months following the initial assessment. A greater number of individuals with the CFAT retest within the short interval allowed greater confidence in the identification of the trends of change in CFAT scores.

There was also a range of days for the second group of candidates, who were retested three months or more following the initial assessment, although most of the candidates were retested four months (or 120 days) after the initial assessment ($M = 141.6$, $SD = 43.9$). Among the candidates who identified their gender, there were 451 males (68.4%) and 257 females (31.5%). In addition, 451 candidates chose to do the CFAT in English and 257 candidates chose to do the CFAT in French.

In order to examine potential differences in cognitive abilities between individuals who were retested earlier than three months or three months or more since the initial assessment, independent samples t-tests were conducted. There were no significant differences between the two groups on verbal, spatial, problem-solving abilities, or total CFAT scores (Table 2). Individuals who took CFAT three months or more following initial assessment were not different on cognitive ability from those who took CFAT earlier than three months following initial assessment.

Table 2: Independent Samples. Comparison of Initial Scores
among Individuals from Two Groups

	Retest Earlier than Three Months	Retest Three Months or more	T Test
CFAT Subscale	**Mean (SD)**	**Mean (SD)**	
Verbal skills	6.9 (2.7)	6.6 (3.0)	0.6
Spatial abilities	7.8 (2.8)	7.2 (2.8)	1.8
Problem-solving	10.8 (5.0)	9.8 (5.0)	1.5
Overall CFAT	25.5 (7.8)	23.5 (8.3)	1.8

---

3   Descriptive statistics for the means and standard deviations are given in days.

3.2   Data Analysis

In order to examine the impact of a short-term interval between initial test and retest on the retest performance, a series of analyses were conducted. Specifically, in order to examine within-person retest effects, paired samples t-tests were conducted, in which CFAT scores on initial assessment were compared to the CFAT scores on retest for two groups: 1) test-retest interval of less than three months and 2) test-retest interval of three months or more. Within-person retest effects refer to effects associated with the same group of individuals who retake an identical test (or an alternate form of the test) (Lievens, Buyse, and Sackett, 2005). The paired samples t-test examines whether there is a significant difference between test means of the same individuals across two examinations. Follow-up regression analyses were conducted to examine the potential link between CFAT performance on retest and the length of the test-retest interval. For this purpose, CFAT scores on retest were regressed onto the length of the interval following the initial assessment.

4   Results

4.1   Paired Samples T-Tests

In order to assess the first hypothesis, which stated that CFAT scores on retest would be higher than those on the initial test, the goal was to conduct two sets of paired samples t-tests: first, for those individuals who took the retest three months or less after the initial assessment, and second, for those who took the retest three months or more after the initial assessment. In addition, although the sample size for individuals who took the retest exactly seven days after the initial test was low ($N = 25$), the t-test analyses were conducted for exploratory purposes.

4.1.1   Seven-day Test-retest Interval

For individuals who took the retest exactly seven days following the initial test ($N = 25$), an improvement of scores on every subscale of CFAT was significant (Table 3).

Table 3: Paired Sample. Seven-day Test-retest Interval

	Initial Test	Retest	T Test
CFAT Subscale	**Mean (SD)**	**Mean (SD)**	
Verbal skills	7.3 (3.1)	8.9 (3.0)	2.7*
Spatial abilities	8.2 (2.4)	9.4 (2.7)	3.4**
Problem-solving	11.4 (4.4)	16.2 (5.7)	4.9***
Overall CFAT	27.0 (7.0)	34.5 (9.0)	5.4***

*$p<.05$; **$p<.01$; ***$p<.001$

4.1.2   Less than Three Months Test-retest Interval

For individuals who took the retest less than three months following the initial test ($N = 597$), an improvement of scores on every subscale of the CFAT was significant (Table 4). Nevertheless, comparison of the mean differences in Tables 3 and 4 demonstrated that an improvement following exactly seven days after the initial assessment was greater than following a longer interval between seven days and three months.

Table 4: Paired Sample. Less than Three Months Test-retest Interval

	Initial Test	Retest	T Test
CFAT Subscale	**Mean (SD)**	**Mean (SD)**	
Verbal skills	7.2 (2.9)	7.7 (3.0)	6.1***
Spatial abilities	7.7 (2.7)	8.8 (2.8)	11.7***
Problem-solving	10.4 (4.5)	13.6 (5.6)	20.0***
Overall CFAT	25.4 (7.1)	30.2 (8.6)	21.7***

***$p<.001$

4.1.3 Three Months or more Test-retest Interval

For individuals who took the retest three months or more following the initial test ($N = 111$), an improvement of scores on every subscale of the CFAT was significant (Table 5). However, comparison of the mean differences in Tables 4 and 5 demonstrate that an improvement following three months or more after the initial assessment was smaller than following a shorter interval between seven days and three months.

Table 5: Paired Samples. Three Months or more Test-retest Interval

	Initial Test	Retest	T Test
CFAT Subscale	Mean (SD)	Mean (SD)	
Verbal skills	6.6 (3.0)	7.3 (3.0)	3.7***
Spatial abilities	7.1 (2.8)	7.8 (2.8)	2.7**
Problem-solving	9.8 (5.0)	12.0 (5.6)	6.7***
Overall CFAT	23.5 (8.3)	27.1 (9.2)	7.1***

**p<.01; ***p<.001

4.1.4 Paired Samples T-Test Summary

The paired-samples t tests confirmed the first hypothesis of the study. Overall, performance on retest was significantly greater than performance at initial testing. Nevertheless, CFAT scores on retest were higher when the retest took place less than three months following the initial assessment. Analyses examining whether the differences between these groups were significant are presented in the next section of this paper.

4.2 Multivariate Analyses of Variance (MANOVA)

In order to examine the second hypothesis of the study, which stated that the gain in CFAT scores for the short test-retest interval is greater than the gain for the test-retest interval of three months or longer, MANOVA were conducted. For these purposes, the difference between the initial CFAT scores and retest scores was calculated and compared for individuals who were retested within three months and those who were retested at three months or more. Given the large difference in the sample sizes of the two groups ($N1 = 599$, $N2 = 110$), which can attenuate the results, 110 cases were randomly selected from the first group of 599 cases[4]. The selected 110 cases from the first group (for which the retest was taken less than three months following the initial test) were compared with the 110 cases from the second group (those who took the retest three months or more following the initial test).

---

[4]   MANOVA analysis was also conducted to compare the two original groups with unequal sample sizes and the same pattern of results was found.

In order to ensure homogeneity of variance, multivariate and univariate analyses were conducted. Box's M test of equality of covariance matrices was not significant at an alpha level of .001, demonstrating homogeneity of variance. In addition, the univariate tests for homogeneity of variance for each of the dependent measures were conducted. Levene's test of equality of error variances was not significant for the overall CFAT, $F (1, 218) = 2.37$, ns, the CFAT verbal, $F (1, 218) = 0.00$, ns, spatial, $F (1, 218) = .08$, ns, or problem-solving, $F (1, 218) = 3.58$, ns subscales, indicating that the homogeneity of variance assumption has not been violated.

The multivariate test demonstrated that there were significant differences between the group that took retest less than three months following the original test and the group that took retest three months or more following the original test, Wilk's $\lambda = .931$; $F (3, 216) = 5.32$, $p<.01$.

Examination of the univariate F tests for each dependent variable demonstrated that there were significant differences between the two groups on the spatial abilities, $F (1, 219) = 13.7$, $p<.05$, problem-solving abilities, $F (1, 219) = 8.72$, $p<.05$, and overall CFAT, $F (1, 219) = 12.5$, $p<.05$. There were no significant differences between the two groups on the verbal abilities domain, $F (1, 219) = 3.00$, ns. These findings suggest that those who took the retest earlier than three months following the initial test had significantly higher scores on spatial abilities and problem-solving subscales as well as on the overall CFAT (Table 6).

Table 6: Descriptive Statistics. CFAT Scores for Two Groups

	Less than Three Months Test-retest Period (*N*=110)		**Three Months or more Test-retest Period (*N*=110)**	
CFAT subscale	**Mean (SD)**		**Mean (SD)**	
Verbal skills	8.0	(2.7)	7.3	(3.0)
Spatial abilities	9.2	(2.6)	7.7	(2.5)
Problem-solving	14.3	(4.8)	12.0	(3.5)
Overall CFAT	31.5	(9.2)	27.1	(9.2)

4.3   Regression Analyses: Increase in CFAT Scores as a Function of the Test-retest Interval

In order to examine Hypothesis 3, which stated that performance on the retest would be a function of the test-retest time interval, multiple regression analyses were conducted. CFAT scores on retest were regressed onto the length of the interval following the initial assessment.

4.3.1   Less than three-months Interval

As can be seen in the Table 7, there was a significant negative correlation between test-retest interval and CFAT scores, suggesting that the greater the test-retest interval following the initial assessment, the lower the performance on CFAT retest.

The length of the interval between initial assessment and retest (i.e., number of days passed since the initial assessment) significantly predicted performance in verbal skills, $R^2 = .017$, $F(1, 595) = 10.54$, $p<.01$, spatial skills, $R^2 = .008$, $F(1, 595) = 4.90$, $p<.05$, problem-solving skills, $R^2 = .036$, $F(1, 595) = 21.91$, $p<.001$, and total CFAT score, $R^2 = .038$, $F(1, 595) = 23.36$, $p<.001$.

Table 7: Pearson Correlations between Test-retest Interval and CFAT Scores on Retest within Three Months following the Initial Assessment

	Verbal Skills	Spatial Abilities	Problem-solving	Overall CFAT
Test-retest interval	-0.13**	-0.09*	-0.19**	-0.19**
Verbal skills	—	0.12**	0.44***	0.66**
Spatial abilities		—	0.38***	0.60***
Problem-solving			—	0.91***

*$p<.05$; **$p<.01$; ***$p<.001$

4.3.2   More than three-months Interval

There was no significant correlation between test-retest time interval and CFAT scores (Table 8), suggesting that the length of the time interval between initial assessment and retest is not associated with CFAT performance when the test-retest time interval is three months or more. In addition, the length of the interval between initial assessment and retest did not significantly predict verbal skills, $R^2 = .012$, $F(1, 109) = 1.28$, *ns,* spatial skills, $R^2 = .000$, $F < 1$, problem-solving skills, $R^2 = .006$, $F < 1$, or overall CFAT score, $R^2 = .007$, $F < 1$.

Table 8: Pearson Correlations between Test-retest Interval and CFAT Scores on Retest Three Months or more following the Initial Assessment

	Verbal Skills	Spatial Abilities	Problem-solving	Overall CFAT
Test-retest interval	-0.11	0.00	-0.08	-0.08
Verbal skills	—	0.28**	0.52***	0.73**
Spatial abilities		—	0.45***	0.67***
Problem-solving abilities			—	0.92***

**$p<.01$; ***$p<.001$

4.3.3   Increase in CFAT Scores as a Function of the Test-retest Interval: Summary

Regression analyses demonstrated that performance on CFAT at the time of retest was a function of the passage of time following the initial assessment when the retest occurred less than three months following the initial assessment. On the contrary, when the retest occurred three months or more following the initial assessment, the test-retest interval did not predict performance on retest. The evidence that the length of time following initial assessment predicted performance on retest for shorter time interval suggests the presence of a practice

effect. Moreover, the data demonstrated that the practice effect gradually decreased as length of time following the initial assessment increased, and it finally disappeared three months following the initial assessment.

4.4   Regression Analyses: Increase in CFAT Scores as a Function of Performance at Initial Testing

The fourth hypothesis of the study stated that individuals who performed more poorly on the initial test would gain more on retest than would those who performed better on the initial test. Performance at initial testing was correlated with improvement on CFAT subscales but not with overall CFAT score[5] (Table 9). It is possible that some individuals improve on one scale but not on others, so that the calculation of the overall CFAT score cancels out the improvements. Indeed, while the improvement on the problem-solving subscale was significantly correlated with the improvement on both spatial abilities ($r = .14$, $p<.001$) and verbal skills subscales ($r = .12$, $p<.01$), the improvement on the verbal skills subscale was significantly correlated with the improvement on the problem-solving subscale ($r = .12$, $p<.01$), but not with the improvement on the spatial abilities subscale ($r = .03$, ns). This evidence suggests that some individuals who improve on the spatial abilities subscale would not necessarily improve on the verbal skills subscale and vice versa.

The correlations between performance on initial test and improvements on CFAT domains were negative, meaning that the greater the performance at initial testing, the smaller the improvement on CFAT domains. In other words, it was found that the lower the performance on the initial test, the greater the improvement on retest on every CFAT domain.

Table 9: Correlations between Initial CFAT Scores and Improvement on CFAT

	Overall CFAT Initial	Verbal Skills Subscale Initial	Spatial Abilities Subscale Initial	Problem-solving Subscale Initial
Overall CFAT improvement	**-0.06**	-0.03	-0.09*	-0.02
Verbal skills subscale mprovement	-0.08	**-0.32****	0.01	0.06
Spatial abilities subscale improvement	-0.07	0.06	**-0.40****	0.09*
Problem-solving abilities subscale improvement	0.01	0.10*	0.11**	**-0.12****

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

---

5   An improvement on a CFAT scale was calculated by subtracting the initial score from the relevant CFAT scale retest score.

In order to assess this hypothesis, regression analyses were conducted, in which the improvement on CFAT domains was regressed onto the CFAT scores obtained on initial test. Specifically, three hierarchical regression analyses were conducted[6], in which improvement on each of the CFAT subscales was regressed onto the relevant CFAT domains at initial testing, statistically controlling for the length of time following the initial assessment. Cognitive abilities at initial CFAT testing, entered in the second block of the regression equations, significantly predicted improvement in verbal abilities, $R^2_{change} = .101$, $F(1, 705) = 79.31$, $p<.001$, spatial abilities, $R^2_{change} = .162$, $F(1, 705) = 137.57$, $p<.001$, and problem-solving abilities, $R^2_{change} = .018$, $F(1, 705) = 12.79$, $p<.001$, while statistically controlling for the effect of interval length following initial assessment. These results demonstrated that, consistent with previous research (e.g., Carretta Zelenski and Ree, 2000) and with Hypothesis 4 of this study, individuals who obtain lower scores on CFAT initially benefit more from retest than do those individuals who obtain higher scores, even when the length of interval following the initial assessment is statistically controlled.

Table 10: Multiple Regression Analyses Assessing the Relationships between Performance at Initial Testing and CFAT Improvement

	Pearson r	*β*	*R^2 Change*
Verbal skills improvement			
Number of days following initial assessment	-0.02	-0.05	0.000
Verbal skills subscale initial	-0.32***	-0.32***	0.101***
Spatial abilities improvement			
Number of days following initial assessment	-0.10**	-0.13***	0.010**
Spatial abilities subscale initial	-0.40***	-0.40***	0.162***
Problem-solving improvement			
Number of days following initial assessment	-0.12**	-0.14***	0.015***
Problem-solving subscale initial	-0.12**	-0.13***	0.018***

**$p<.01$; ***$p<.001$

---

[6] Given the improvement on the overall CFAT was not correlated with the CFAT initial score, the regression analysis for the overall CFAT was not conducted.

5   Conclusion

There is a necessity in organizational settings to implement a cognitive test-retesting policy for personnel selection purposes. The main purpose of providing a retest in an organizational setting is to assess a potential improvement in cognitive functioning. In addition, retesting allows for the demonstration of a candidate's true ability if the candidate had certain transient limitations (e.g., illness) when taking the initial test.

In the CF, the cognitive ability of potential recruits is tested using the CFAT. Given that it is necessary to ensure that selection decisions are fair and equitable, and that the CFAT is one of the few resources in the CF selection system to compare candidates objectively and fairly, fairness and objectivity of the CFAT procedures are vital issues. It is essential to ensure that testing and retesting procedures have been standardized and that all candidates are subjected to the same challenge, in order to demonstrate individual performance against a valid and/or pertinent selection standard.

Until January 2007, the CFAT retest policy stated that a candidate is eligible for a retest three months after the initial assessment (PPD 203, 1996). Nevertheless, due to the increasing need for additional personnel in the CF, the retest policy was changed and the length of the test-retest interval following the initial assessment was reduced to seven days. Because previous research, examining the impact of a short test-retest interval, has suggested that reducing the test-retest interval to seven days might result in an increased practice effect (e.g., Carretta, Zelenski and Ree, 2000; Falleti *et al.,* 2006), this study was conducted to compare the potential changes in scores among CF candidates who took CFAT retest since the seven-day policy came into place.

Previous research (Basso, Bornstein, and Lang, 1999; Falleti, Maruff, Collie, and Darby, 1996; Kay and Kane, 1991; Lowe and Rabbitt, 1998) demonstrated that allowing a retest has the potential problem of a practice effect. Specifically, individuals may perform better on retest due to criterion unrelated variance (e.g., learning tricks, memorizing items) rather than due to true improvement in cognitive abilities. Hausknecht, Halpert, Di Paolo, and Moriarty (2007) conducted a meta-analysis to summarize the results of 50 studies of practice effects for tests of cognitive ability examining 107 samples and 134,436 participants and revealed a significant practice effect with an adjusted overall effect size of .26..Therefore, researchers in the area of testing and selection warn of the danger of administering a test without adequate knowledge of a practice effect. Among other factors influencing the magnitude of the practice effect, the length of the test-retest interval was found to be the key factor (e.g., Benedict and Zdaljardic, 1998). According to Dikmen, Heaton, Grant, and Temkin (1999), while such factors as the individual's age and general ability level at the time of initial testing can influence the magnitude of a practice effect to some degree, the length of the test-retest interval is the main factor determining the occurrence and magnitude of a practice effect.

Given that practice effect is an important concern interfering with the validity of testing results, in this paper, the literature in the area of retesting and practice effect was reviewed. Overall, previous research demonstrated that the length of the test-retest interval had a direct impact on the magnitude of the practice effect. Specifically, retests with short intervals (especially if less than one month) would have a higher practice effect magnitude, gradually decreasing with the passage of time.

The study presented in this paper was conducted to examine the impact of a shorter test-retest interval on CFAT scores on retest among CF candidates. In this study, analyses were conducted to compare CFAT scores on initial test and on retest for two groups of individuals: those who were retested three months or more and those who were retested less than three months following initial testing. In addition, an analysis was conducted to examine performance on retest as a function of passage of time following the initial assessment. Finally, the factors associated with better performance on retest were assessed. Specifically, the analyses were conducted to examine whether poorer performance at initial testing predicted better performance on retest, or, in other words, whether individuals with lower CFAT scores benefit the most from retest.

The results of the study were consistent with previous research (e.g., Carretta Zelenski and Ree, 2000; Falleti *et al.* 1996). Although there was an increase in CFAT scores on retest at both shorter and longer (three months or more) time intervals, the increase was larger when retest occurred less than three months following the initial assessment. Furthermore, the increase in the scores was greater among candidates who took the retest exactly seven days after initial assessment as compared to the candidates who took the retest in the period between seven days and three months. MANOVA analyses demonstrated significant differences between CFAT scores for individuals who took the retest three months or more as compared to those who took the retest less than three months following initial testing. These results demonstrated that a practice effect was smaller when the length of time following initial testing increased. Consistent with previous research (Carretta Zelenski and Ree, 2000), practice effects diminished as the length of the retest interval increased.

The follow up regression analyses demonstrated that the length of interval following the initial assessment significantly predicted performance on retest only for the candidates who took the retest less than three months after the initial assessment. The length of time following the initial assessment did not significantly predict performance among those who took retest three months or more following the initial assessment. This initial increase in CFAT scores, and the gradual decrease with the passage of time following initial assessment, can be attributed to a practice effect that was stronger immediately following initial assessment.

These findings are consistent with previous research conducted in the area of retest for selection purposes. Specifically, Falleti, Maruff, Collie, and Darby (1996) also found an increase in cognitive functioning scores on retest one week after the initial assessment. Carretta Zelenski and Ree (2000) recommended a six-month test-retest period for cognitive testing for military organizations, especially for highly cognitively loaded jobs (e.g., pilot). The results of the current study suggested that the practice effect was non-significant when the retest was administered to CF candidates three months or more following the initial testing.

Finally, regression analyses indicated that individuals who had lower scores on initial testing benefit more from retest than do those with higher scores. It seems unlikely that the change in the CFAT scores was due to a true improvement in cognitive abilities among individuals with poorer performance on a cognitive test. Taking into account other information obtained in this study, it seems that this improvement reflected a practice effect, which diminished with the passage of time, as candidates remembered less the initial test. This finding is consistent with previous research demonstrating that individuals who do more poorly on the initial cognitive test benefit the most from retest (Carretta Zelenski and Ree, 2000). A smaller improvement among

individuals who performed better can be attributed to a 'ceiling effect', when an improvement in the cognitive functioning is too small to be statistically significant.

Previous research and current study findings demonstrate that scores on cognitive tests always increase on retest. However, the shorter the interval following the initial assessment, the greater the increase in scores and the more likely it would be due to criterion unrelated variance (i.e., practice effect). Specifically, the increase in CFAT scores seven days after the initial assessment is greater and involves greater practice effect than the increase observed three months or more following initial assessment. While the length of interval following initial assessment is critical within the first three months, and especially the first month following initial assessment, its role in retest performance is less critical three months or more after initial assessment. Finally, when retest is administered shortly after the initial assessment, individuals with low scores on the initial test benefit the most from practice effect. It seems likely that when the retest is administered within a short time interval following the initial assessment, more individuals with lower cognitive abilities may be accepted into the CF.

To conclude, the CFAT is used to predict future training success and job performance. When candidates take a retest too soon, their scores are inflated (i.e., their ability appears to be higher than it actually is because of the practice effect). If large numbers of candidates are selected into the CF based on the inflated erroneous retest scores, failure rates on training will increase, resulting in increased training costs for the CF and an increase in the time it takes to meet CF training establishments. While a three-month interval still inflates scores to some degree (a one-year interval would be ideal), it does not have as serious inflation as a one-week interval would. It is highly recommended, therefore, for reasons of Force structure, that the interval between first test and retest be set at a minimum of three months.

6   Recommendation

Based on the previous research, the results of this study, and the harmful effects a shorter retest period would have on training success and Force structure, it is recommended that a minimum three-month period be set between initial selection testing and retest.

References

[1] Anastasi, A. (1988). *Psychological Testing* (6th ed), New York: Macmillan.

[2] Anastasi, A. and Urbina, S. (1997). *Psychological Testing* (7th ed.), Upper Saddle River, NJ: PrenticeHall.

[3] Basso, M.R., Bornstein, R.A., and Lang, J.M. (1999). Practice effects on commonly used measures of executive function across twelve months. *The Clinical Neuropsychologist, 13*, 283-292.

[4] Benedict, R.H.B. and Zdaljardic, D.J. (1998). Practice effects during repeated administration of memory tests with and without alternative forms. *Journal of Clinical and Experimental Neuropsychology, 20*, 339-353.

[5] Bland, J.M. and Altman, D.G. (1996). Measurement error, *British Medical Journal, 313,* 744.

[6] Bornstein, R.A, Baker, G.B., and Douglass, A.B. (1987). Short-term retest reliability of the Halstead-Reitan Battery in a normal sample. *Journal of Nervous and Mental Disease, 175*, 229-232.

[7] Bruggemans, E.F., Van de Vijver, F.J.R., and Huysmans, H.A. (1997). Assessment of cognitive deterioration in individual patients following cardiac surgery: Correcting for measurement error and practice effects. *Journal of Clinical and Experimental Neuropsychology, 19*, 543-559.

[8] Carretta T.R., Zelenski and Ree, M.J. (2000). Basic Attributes Test (BAT) Retest Performance. *Military Psychology, 12*, 221-232.

[9] Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.

[10] Crook, T.H., Youngjohn, J.R., and Larabee, G.J. (1992). Multiple equivalent forms of a computerized everyday memory battery. *Archives of Clinical Neuropsychology, 7*, 221-232.

[11] Dikmen, S.S., Heaton, R.K., Grant, I., and Temkin, N.R. (1999). Test-retest reliability and practice effects of expanded Halstead-Reitan neuropsychological test battery. *Journal of International Neuropsychological Society, 5*, 346-356.

[12] Falleti, M.G., Maruff, P., Collie, A., and Darby, D.G. (2006). Practice effects associated with the repeated assessment of cognitive function using the cogstate battery at 10-minute, one week, and one-month test-retest intervals. *Journal of Clinical and Experimental Neuropsychology, 28*, 1095-1112.

[13] Girard, M. (2004). *Validation of the CFAT for Vehicle Technician Selection.* Technical Note 2004-02. Director Human Resources Research and Evaluation, National Defence Headquarters, Ottawa.

[14] Goldstein, G. and Watson, J.R. (1989). Test-retest reliability of a new form of the Auditory Verbal Learning Test (AVLT). *Archives of Clinical Neuropsychology, 9*, 303-316.

[15] Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T., and Moriarty G.M.O. (2007). Retesting in Selection: A Meta-Analysis of Coaching and *Practice* Effects for Tests of Cognitive Ability. *Journal of Applied Psychology, 92*, 373-385.

[16] Hodgson, K. (2005). *Validation of the CFAT and Establishment of Cutoff Scores for Military Police Selection.* Technical Note 2005-03. Director Human Resources Research and Evaluation, National Defence Headquarters, Ottawa.

[17] Johnson, B.F., Hoch, K., and Johnson, J. (1991). Variability in psychometric test scores: The importance of the practice effect in patient study design. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 15*, 625-635.

[18] Kay, G. and Kane, R.L. (1991). Repeated measures in neuropsychology: Use of serial testing to measure changes in cognitive functioning. *Journal of Clinical and Experimental Neuropsychology, 13*, 49-54.

[19] Kulik J.A., Kulik, C.C., and Bangert, R.L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal, 21*, 435-447.

[20] Lezak, M.D. (1995). Neuropsychological Assessment (3rd Ed), New York: Oxford University Press.

[21] Lievens, F., Buyse, T, and Sackett, P.R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58*, 981-1007.

[22] Lowe, C. and Rabbitt, P. (1998). Test/re-test reliability of the CANTAB and ISPOCD neuropsychological batteries: Theoretical and practical issues. *Neuropsychologia, 36*, 915-923.

[23] MacLennan, R.N. (1997). *Validity generalization across military occupational families.* Technical Note 00-97. Personnel Research Team, Ottawa, Ontario, Canada.

[24] Matarazzo, J.D., Carmody, T.P., and Jacobs, L.D. (1980). Test-retest reliability and stability of the WAIS: A literature review with implications for clinical practice. *Journal of Clinical Neuropsychology, 2*, 89-105.

[25] McCaffrey, R.J., Ortega, A., Orsillo, S.M., Nelles, W.B., and Haase, R.F. (1992). Practice effects in repeated neuropsychological assessments. *The Clinical Neuropsychologist, 6*, 32-42.

[26] Personnel Psychology Directive 203 (1996), Canadian Forces, D Pers Pol 6-3, 21 1810Z AUG 96, Ottawa, Ontario.

[27] Rapport, L.J., Brines, D.B., Axelrod, B.N. and Theisen, M.E. (1997). Full scale IQ as mediator of practice effects: The rich get richer. *Clinical Neuropsychologist, 11*, 375-380.

[28] Ree, M.J., Earles, J.A., and Teachout, M. (1994). Predicting job performance: Not much more than g, *Journal of Applied Psychology, 79*, 518-524

[29] Scholtz, D. (2004). *Validation of the CFAT and Establishment of Cutoff Scores for Steward Selection*. Technical Note 2004-01. Director Human Resources Research and Evaluation, National Defence Headquarters, Ottawa, Ontario, Canada.

[30] Shatz, M.W., (1981). WAIS practice effects in clinical neuropsychology. *Journal of Clinical Neuropsychology, 3*, 171-179.

[31] Temkin, N.R., Heaton, P.K., Grant, I., and Dikmen, S.S. (1999). Detecting significant change in neuropsychological test performance: A comparison of four models. *Journal of the International Neuropsychological Society, 5*, 357-369.

[32] Theisen, M.E., Rapport, L.J., Axelrod, B.N., and Brines, D.B. (1998). Effects of practice in repeated administrations of the Wechsler memory scale-revised in normal adults. *Assessment,* 5, 85-92.

[33] Uchiyama, C.L., D'Elia, L.F., Dellinger, A.M., and Becker, J.T. (1995). Alternate forms of the Auditory-Verbal Learning Test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology, 10*, 133-145.

[34] Vanderpool, M.A. (2003). *Determining if the Canadian Forces Aptitude Test is Adversely Impacting Canadian Aboriginal Peoples*. Technical Note 2003-03. Director Human Resources Research and Evaluation, National Defence Headquarters, Ottawa, Ontario, Canada.

[35] Watson, F.L., Pasteur, M.A.L., Healy, D.T., and Hughes, E.A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology, 9*, 51-61.

[36] Woycheshin, D.E. (1999). *Validation of the Canadian Forces Aptitude Test against QL3 course performance*. Technical Note 99-11. Director Human Resources Research and Evaluation, National Defence Headquarters, Ottawa, Ontario, Canada.

[37] Youngjohn, J., and Crook, T. III (1993). Stability of everyday memory in age-associated memory impairment: A longitudinal study. *Neuropsychology, 7*, 406-416.

This page intentionally left blank.

Distribution list

Document No.: DGMPRA TM 2009-003

    LIST PART 1: Internal Distribution by Centre
1   CMP List 1
1   ADM (S&T)
1   DGLCD
1   DG Air Pers
1   DG Air FD
1   RMC (Kingston)
1   CFC (Toronto)
1   DG CORA
1   DRDC CORA Chief Scientist
2   DRDC CORA Library
2   DRDKIM Library
1   DRDC/DGSTO/DSTP
1   DGMPRA
1   DGMPRA – Chief Scientist
1   DGMPRA – Deputy DG
1   DGMPRA – Personnel Generation Research – Section Head
1   DGMPRA – Personnel and Family Support Research – Section Head
1   DGMPRA – Organizational and Operations Dynamics – Section Head
1   DGMPRA – Team Leaders
1   CMS/D Mar Strat 2-6
1   SJS DOSS Pers Ops
1   DRDC (Toronto)
1   DMP Pol 2
1   VAC LO
26   TOTAL LIST PART 1

    LIST PART 2: External Distribution by DRDKIM
1   Library and Archives Canada

1   TOTAL LIST PART 2

27   TOTAL COPIES REQUIRED

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) DGMPRA 101 Colonel By Drive Ottawa, Ontario K1A 0K2	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED (NON-CONTROLLED GOODS) DMC A REVIEW:GCEC JUNE 2010

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

Canadian Forces Aptitude Test: Repeated Assessment and Practice Effect:

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Skomorovsky, A.

5. DATE OF PUBLICATION (Month and year of publication of document.) May 2009	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 36	6b. NO. OF REFS (Total cited in document.) 37

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

DGMPRA
101 Colonel By Drive
Ottawa, Ontario K1A 0K2

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DGMPRA TM 2009-003	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

The Canadian Forces Aptitude Test (CFAT) retest policy required by Personnel Psychology Directive (PPD) 203 was changed on January 2007. Specifically, the length of the test-retest interval following the initial assessment was reduced from three months to seven days. Previous research suggests that when the length of the test-retest interval is short, an increase in test score on retest may occur due to criterion-unrelated variance (i.e., practice effect). This study examines the impact of reducing the length of the test-retest interval to seven days. Results demonstrate a significant increase in CFAT scores seven days after the initial assessment, which is greater than the increase occurring three months following the initial assessment. It is recommended that a minimum three-month period be set between initial selection testing and retest.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

DEFENCE **R&D** DÉFENSE

DRDC CORA

www.drdc-rddc.gc.ca