The Influence of Candidate Cognitive Ability and Personality on Assessment Center Ratings

By

Nadia A. Worthington

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.

August, 2018, Halifax, NS

Approved: Dr. N. Roulin
                 Supervisor

Approved: Dr. D. Powell
                 Examiner

Approved: Dr. V. Catano
                 Examiner

Approved: Dr. J. O'Brien
                 Examiner

Date: 19 July 2018

Table of Contents 2

Abstract

The Influence of Candidate Cognitive Ability and Personality
on Assessment Center Ratings

By Nadia A. Worthington

Abstract: The purpose of this study was to advance our understanding of assessment center (AC) functioning. This research examined (a) AC construct validity, (b) the moderating role of cognitive ability on the relationship between personality and performance, and (c) the existence of potential adverse impact. Over a period of 12 years, performance data were collected on 299 AC candidates who attended a four-day military AC. Ten dimensions were assessed across eight exercises. Multitrait-multimethod analyses examined convergent and discriminant validity and revealed an exercise effect. Results obtained by comparing exercises that varied in trait activation potential provided partial support for application of the trait activation theory framework for understanding construct validity. Comparisons of AC candidates across groups demonstrated significant group differences for the selection decision and assessment ratings based on first official language, and not sex. Results shows that Francophone candidates achieved higher overall assessment ratings and were more frequently found suitable.

July 19th, 2018.

Following a request from the organization that provided the data (i.e., DGMPRA), the present version of the thesis is a "redacted" version. The real names of the Assessment Center exercises have been replaced by letters (e.g., Exercise A) throughout the thesis, to ensure that confidential information about exercise content was not made public. However, the version of the thesis that was defended and approved by the committee included all the information.

Acknowledgements

What follows in this manuscript did not result in isolation. First, I feel I should thank my supervisor, whose steadfast encouragement and insight were key to the continuation and completion of this project. I am also indebted to both Dr. Victor Catano and Dr. James O'Brien for the many thoughtful discussions that helped me focus throughout this process. Finally, Dr. Debra Powell's thoughtful, thorough, and valued feedback was immensely appreciated. Additionally, I would like to thank Kathy Berlinguette, whose guidance and support throughout this process was invaluable.

I'd also like to thank Jennifer Price and Alex Duval. Although we only had one year together, your thoughtful discussions, professionalism, and support came to me at a critical time. I only hope that I can return the kindness when our paths cross again.

To my cohort: on many occasions, each one of you has graciously given me your time when I needed technical (read: statistics) and emotional support. For this I am eternally thankful. Mostly, I am grateful for the time we had together because being around all of you lifted my spirits and broadened my understanding of the many pressing issues we face today.

Finally, I thank my husband for both his emotional and editorial support. Also, my family and friends have encouraged me through these past two years and my career. I would not be writing this today without their encouragement to take on new challenges.

The Influence of Cognitive Ability and Personality on Assessment Center Ratings

Assessment centers (ACs) are a commonly used personnel selection tool. This popularity stems from the method's ability to elicit specific behaviors from candidates (Arthur, Day, & Woehr, 2008) and the fact that it produces minimal adverse impact (Iles, 1992). In addition to personnel selection, ACs are used for evaluating managerial potential (Schleicher, Day, Mayes, & Riggio, 2002), leader development (Meriac, Hoffman, Woehr, & Fleisher, 2008), retention of specific employees during cutbacks as well as internal and external selection in organizations and their respective sub-units (Thornton, & Gibbons, 2009). The AC method is used by the Canadian Armed Forces (CAF) to select applicants for various military occupations.

A selection process must be reliable, valid, legally defensible (Catano, Hackett, & Wiesner, 2016). Specifically, ACs should follow the set of protocols outlined by the International Task Force on Assessment Center Guidelines (Rupp et al., 2015). Traditionally, construct validity is considered evidence that demonstrates a selection method's validity (Society for Industrial and Organizational Psychology, 2003). Establishing construct validity for AC presents challenges that are partially attributable to findings that ACs tend to measure methods, or exercises, and not the behavioral dimensions, as intended (Jackson, Michaelides, Dewberry, & Kim, 2016). This finding, dubbed the *exercise effect*, was considered error and a critical issue early in AC research (e.g. Sackett & Dreher, 1982). Further research brought attention to the fact that ACs, despite the exercise effect, successfully predicted performance causing researchers to reconsider the finding (Lance, Foster, Gentry, & Thoresen, 2004).

Lievens (2002) suggested that the variation in results across exercises occurs

because candidates change their behavior in response to the situation (i.e. exercise). This suggestion is supported by the fact that exercises are designed to elicit behaviors relating to job tasks and competencies identified as important for the target occupation or role (Howard, 2008). Thus, it is likely that candidates will perform differently across exercises, performing better in some than in others (Arthur et al., 2008). Limited research has explored how exercises impact the dimensions (e.g. Spector, Schneider, Vance & Hezlett, 2000; Shore, Thornton, & Shore, 1990) and, from a practical standpoint, there is little understanding and guidance on the design of exercises that efficiently elicit target behaviors (Lievens, Tett, & Schleicher, 2009).

Further study in this area of AC research is especially important given that dimensions play a key role in the design, scoring and interpretation of ACs (Meriac, Hoffman, & Woehr, 2014). Lievens and colleagues (2006) suggested the use of trait activation theory (TAT) to further advance research in the AC field. TAT suggests that traits will lead to behaviors, but only when the situation cues the relevant trait (Tett & Burnett, 2003). Therefore, within the AC, dimensions are no longer viewed as stable traits (Lievens, 2009). TAT also highlights the importance of building multiple stimuli into AC exercises that are explicitly designed to have strong situation relevance (Lievens & Schollaert, 2011) which in turn leads to increased behavioral observability.

In addition to the traits that are activated during exercises, cognitive ability plays an important role because both cognitive ability and personality can influence individual functioning (Aschwanden, Martin, & Allemand, 2017). Both have been established as valid predictors of a number of job-relevant outcomes, such as performance (e.g. Barrick, Mount, & Judge, 2001; Schmidt, & Hunter, 1998). Considerably less research has

investigated the interaction between these two predictors. Wright, Kacmar, McMahan, and Deleeuw (1995) found that the interaction between cognitive ability and personality explained a significant amount of incremental variance in job performance. In the AC context, less is known about how the interaction of cognitive ability and personality influence candidate performance. A study by Spector and colleagues (2000) investigated cognitive ability and personality correlates of AC exercises, but not the interaction between both.

To further our understanding of AC functioning, the current research aims first to examine the construct validity of an operational military AC. It is expected that a lack of convergent and discriminant validity will be observed, which leads us to examine candidate performance from a different perspective. Therefore, the second aim is to use TAT to further examine the effectiveness of specific exercises at eliciting relevant traits. Thirdly, we investigate the influence of cognitive ability and personality on candidate performance. By considering both together, we can more clearly understand how individual differences influence candidate performance during ACs. Finally, we examine group differences in AC performance in order to determine if any potential biases exist. Exploring potential differences for AC performance for male and female candidates, as well as Anglophone and Francophone candidates, will lead to a better understanding of any potential selection bias associated with this selection method.

The Assessment Center Method

The overall goal of an AC is to evaluate a candidate's performance in order to predict their ability to be successful in a target role (Thornton, & Gibbons, 2009). The Guidelines and Ethical Considerations for Assessment Center Operations establishes a

procedure for the creation and administration of ACs (Rupp et al., 2015). Initially, a job analysis is required to identify the relevant dimensions, job-related simulations, knowledge, skills, abilities and tasks. Once conducted, the job analysis is used as a foundation throughout the building process.

AC candidates are assessed by a team of trained assessors who observe and record behaviors on pre-determined behavioral dimensions (Jackson et al., 2016). Common sources of evidence include written test scores, group exercises, in-basket exercises and interviews (Shore, et al., 1990). Candidate behavior is ultimately aggregated and integrated in a meaningful fashion that enables the selection of candidates (Rupp et al., 2015).

Ratings from across the AC are often combined into one overall assessment rating (OAR) which is then used for decision-making. However, AC performance can be captured in other meaningful ways. Post-exercise dimensions ratings are created by combining the dimensions ratings obtained during AC exercises. More recently, researchers have also used exercise ratings as meaningful units of performance (e.g. Jackson, Barney, Stillman, & Kirkley, 2007).

The Military Police Officer Assessment Center

The Military Police Officer Assessment Center (MPOAC) was implemented in 2005 as the final selection stage for applicants applying to the military police officer (MPO) occupation. Prior to being considered as a candidate for the MPOAC, applicants must complete several selection steps which include, but are not limited to: (a) the Canadian Forces Aptitude Test (CFAT) – a generalized test of cognitive ability; (b) the Trait Self-Descriptive Personality Inventory (TSD-PI); and (c) a structured employment

interview. The CFAT and TSD-PI are administered before the candidates complete the interview; however, the period of time between test administrations, interviews, and AC attendance varies for each candidate.

Applicants that do not meet the minimum score of the 30th percentile on the total CFAT score are considered ineligible for the MPO occupation. There are currently no minimum cut-off scores for the TSD-PI. Once the interview is completed, applicants are assigned an overall score which is used to create a top-down selection list. Applicants are chosen from this list and offered a position based on the number of available positions.

A job analysis identified core competencies and attributes critical to the MPO occupation. The relevant job behaviours were classified into 10 competencies (henceforth referred to as dimensions). The assessed dimensions include: integrity, conscientiousness, performance under stress, professional demeanor, personal impact, problem solving, initiative, flexibility, resource management, and written communication skills (see Table 1 for detailed definitions).

Multiple tests and tasks were designed to elicit the dimensions derived from the job analysis. Multiple exercises, which incorporated different techniques, were developed (see Dimension by Exercise Matrix, Table 2) to assess 10 of these dimensions. Eight exercises are used to measure the dimensions listed above.

This process resulted in the implementation of the MPOAC, a four-day AC comprised of eight exercises that assessed 10 dimensions. Multiple assessors are used for each iteration of the MPOAC. Assessors must be either (1) senior military police members or MPOs or (2) personnel selection officers. Assessors were randomly assigned

Table 1

Military Police Officer Assessment Center Dimensions

Dimension	Definition
Integrity	Refers to the individual's ability to be honest, trustworthy, and truthful. Individuals who act with integrity are able to be trusted, and not likely to steal, cheat or lie. They should have strong moral principles, which allow them to resist temptations of an unethical or illegal nature.
Resource Management	Refers to the ability of the individual to manage their resources. These individuals produce detailed, clearly defined, short- and long-term plans that achieve desired results while concurrently considering potential limitations, future problems and impacts. They also coordinate and monitor the execution of tasks/projects, including the scheduling of human, material and financial resources to ensure that key objectives are met.
Problem Solving	Refers to the individual's ability to solve problems. They should be able to identify, analyze and understand information they are given and make rational, realistic and sound choices/decisions based on available information, knowledge and skills. They should seek all possible relevant information, probe the facts carefully, and critically analyze the issue from different perspectives. They should be aware of their surroundings and recognize changing situations. Individuals are able to understand the dynamics of organizations, including the formal and informal cultures and decision-making processes.
Personal Impact	Relates to the effectiveness of the individual in terms of the first impression they create and the confidence they display. Includes the ability to command attention and respect, their willingness and ability to effectively take charge of people/situations providing a clear sense of direction and purpose. It includes expressing confidence in their abilities and judgement when dealing with challenging circumstance or when facing their limitations. Corresponds to an individual's confidence when dealing with or first approaching others.

Professional Demeanour	Effectiveness of the individual in their interaction with others, and their ability to interact with other people in a meaningful, appropriate, respectful and effective fashion. Implies a willingness and ability to listen and use attending skills when interacting with others. Individual's show sensitivity, compassion and sincerity, be tactful yet diplomatic in their dealings with others, and be able to reduce tension in potentially hostile situations. Implies a willingness to recognize and respect the beliefs or practice of others, regardless of the difference between them, in order to effectively and respectfully interact with individuals of different backgrounds, personality, attitudes, opinions and values.
Conscientiousness	The ability of the individual to meet or exceed given standards and deadlines, persistently strive for excellence, even in difficult situations, and be efficient, thorough, hardworking and dependable.
Performance Under Stress	Relates to the ability of the individual to keep their emotions under control and show restraint when provoked, when faced with opposition or hostility from others, or when working under stressful conditions. The individual should be able to effectively manage stress to prevent it from negatively affecting performance.
Initiative	Relates to the ability of an individual to act without prompting. The individual who displays initiative is self-motivated and self-directed in identifying and addressing important issues. They are willing to actively influence events rather than passively respond to them. In addition, individuals who display initiative should be able to generate new and imaginative ideas or present novel solutions.
Flexibility	An individual's willingness and ability to be responsive and adaptable to change. These individuals are willing and able to adjust to changing demands and conditions (i.e., crisis situations and repetitive environments), and to adapt their approach when dealing with other people or situations. They are receptive to innovation and willing to embrace change.
Written Communication Skills	The degree to which the candidate accurately and concisely describes events and presents conclusions in a written format, organizing the material in a clear, logical manner using appropriate grammar, style and language.

Table 2

Dimension by Exercise Matrix

	Exercise							
Dimension	Exercise A	Exercise B	Exercise C	Exercise D	Exercise E	Exercise F	Exercise G	Exercise H
Integrity		x	x				x	x
Resource Management				x	x			
Problem Solving	x		x	x	x	x		
Personal Impact	x	x	x		x	x	x	x
Professional Demeanour	x	x	x	x	x	x	x	x
Conscientiousness	x	x	x	x	x	x		
Performance Under Stress	x	x	x		x	x	x	x
Initiative	x	x	x	x	x	x		
Flexibility	x	x	x	x	x	x		

| Written Communication | x | x |

to candidates. All assessors were required to attend 3 days of frame-of-reference training immediately preceding the AC. The training involved defining the dimensions, reviewing behavioral examples of the dimensions, and practicing assigning behavioral ratings. Raters are given constructive feedback and opportunities to practice delivering exercise instructions, scoring, and assessor scripts.

The MPOAC is an expensive selection method and occurs several times throughout a given fiscal year. To date, no research has examined the reliability or validity of this method. However, research has been conducted on the Military Police Assessment Center (MPAC), which assess the suitability of applicants for the Military Police (MP) occupation. MP members performing policing tasks including, but not limited to, emergency response, the investigation and reporting of incidents involving military or criminal offenses, and the provision of community services. MPOs perform managerial functions and are leaders for teams of MP members who enforce laws and regulations across the Canadian Armed Forces.

The MPAC and MPOAC have similarities in terms of the dimensions that are assessed and the exercises used to elicit relevant behaviors. For example, Exercise A, Exercise B, Exercise C, and Exercise F are used in both ACs. Hodgson (2006) attempted to examine the construct and predictive validity of the MPAC. Findings from a principle components analysis and multitrait-multimethod (MTMM) analysis showed that the MPAC lacked construct validity, a consistent finding for AC research using these methods (e.g., Sackett & Dreher, 1982). However, Jalbert (n.d.) found support for both the construct and predictive validity of the MPAC. The MPAC and MPOAC are different methods designed to assess distinct populations. Although Jalbert's (n.d.) research found

support for construct validity, the MPOAC's properties remain to be examined

independently from other CAF assessment centers.

Linking MPOAC Dimensions with Personality Traits and Cognitive Ability

The dimensions evaluated in the MPOAC were created using the findings from a

job analysis, therefore the dimensions are directly linked to the MPO occupation. These

dimensions may not explicitly reflect a specific personality trait or be named as such, but

the dimensions can be linked to things such as the Big Five traits (Lievens et al., 2009)

and cognitive ability.

Lievens and De Fruyt (2001) analyzed notes taken by 403 assessors and found

that the use of five factor model (FFM) trait descriptors varied across exercises. For

example, the presentation exercise, which presumably exerts pressure on candidates,

generated specific descriptions that related to the trait of Emotional Stability. Based on

past research, we expect that the MPOAC dimensions have links to established

personality traits (e.g. Lievens et al., 2009).

Cognitive ability is also expected to influence some dimensions and exercises

more than others (Meriac et al., 2008; Goldstein, Yusko, Braerman, Smith & Chung,

1998). Spector and colleagues (2000) found a stronger association between cognitive

ability and AC exercises that involved problem solving, such as the in-basket exercise.

Goldstein and colleagues (1998) showed that of seven AC exercises, the in-basket

exercise had the largest association with cognitive ability ($r = .29$). Overall, the authors

concluded that AC exercise ratings vary in their association with cognitive ability. Based

on this past research, we examined links between cognitive ability and dimensions.

Construct validity

Evaluating validity is critical to the AC method because this method continues to be widely used for personnel selection and development decisions (Eurich, Krause, Ciguralov, & Thornton, 2009). In general, validity refers to whether a selection tool provides accurate information about the construct being measured. Validity is established through numerous sources (i.e. construct, criterion-related). Having numerous supporting sources provide stronger evidence for a selection tool's validity (Binning & Barrett, 1989).

Although ACs have strong criterion-related and incremental validity, establishing construct validity has been a topic of debate for the past three decades. In earlier literature, the lack of construct validity is viewed as a weakness, often referred to as the *exercise effect* (i.e. Schneider, & Schmitt, 1992). Dimensions within the same exercise tend to be more strongly related to each other than to similar dimensions that are measured across different exercises (Lance, 2008; Sackett, & Dreher, 1982). The traditional approach used to establish AC construct validity is to examine MTMM matrices, a construct validation procedure first recommended by Campbell and Fiske (1959). This analysis is accompanied by factor analysis (FA; e.g. Lance, Lambert, Gewin, Lievens, & Conway, 2004). In this approach, dimensions are viewed as traits and exercises are viewed as methods. The exercise effect is present when there is a pattern of strong correlations between different dimensions measured in the same exercises. High construct validity is present when stronger correlations are observed in dimensions across different exercises. In an FA, items clustering according to the exercises rather than by dimensions indicates low construct validity. AC research has repeatedly found that

correlations tend to cluster by exercise versus dimension (e.g. Lance et al., 2004, Lievens, 2001a).

Significant efforts in the AC research field have been dedicated to understanding the exercise effect. Researchers have proposed possible explanations and mitigating AC design characteristics for this ubiquitous effect. Lievens (2002) recommended using psychologists and human resource (HR) professionals over non-HR personnel as assessors in order to increase rating accuracy. Cognitive overload on assessors has also been considered a potential error source (Sackett & Dreher, 1982). A recent shift in thinking encourages researchers to reconsider how the exercise effect is viewed. Jackson and colleagues (2007) propose that these findings are indicative of situationally driven performance and not of error. Consider a common AC exercise used to elicit and assess problem-solving behaviors, the leaderless group dynamics exercise. Aspects of problem solving captured during the leaderless group dynamics may vary when compared to problem-solving during a situationally different exercise such as a case-analysis. Further, Lievens (2001b) recommended that AC research move beyond the traditional MTMM conceptualizations.

Prior to considering more recent methods of establishing construct validity, the first goal of the current study was to conduct an initial exploration of the AC's construct validity using the methods that have been used historically. It is expected that an MTMM conceptualization of the MPOAC will demonstrate a lack of clear evidence supporting the AC's construct validity, which aligns with past research (e.g. Sackett & Dreher, 1982). Therefore, the following Hypothesis was proposed:

Hypothesis 1: The military police officer assessment center ratings will predominantly reflect exercise effects.

The Role of Trait Activation Theory for Examining AC Ratings

Trait Activation Theory

Over the past decade, trait activation theory (TAT) has gained popularity in the industrial-organizational psychology field (i.e. Lievens, Chasteen, Day and Christiansen, 2006). "Trait activation is the process by which individuals express their traits when presented with trait-relevant situational cues" (Tett & Burnett, 2003, p. 502). Using a person-situation interaction perspective, TAT proposes that situations differ in their ability to elicit behaviors related to specific traits (Wirz, Melchers, Schultheiss, & Kleinmann, 2014). Thus, when presented with situations that have similar trait-relevance and activation potential, behavioral consistency can be expected (Wirz et al., 2014). Applied to ACs, TAT can provide "a framework for understanding how AC participant behavior can be affected by situational demands" (Oliver, Hausdorf, Lievens & Conlon, 2016, p. 1997).

Candidate behavior is neither determined solely by dispositional factors (e.g. personality characteristics) or situational factors (e.g. AC exercises). Candidate performance is the result of the interaction of the person and situation (Lievens & Schollaert, 2011). In the case of ACs, candidates participate in various exercises (i.e. situations) that activate relevant competencies (i.e. traits). The strength of the situation also plays a role in how clearly a trait is expressed (Tett & Burnett, 2003), meaning that not all situations will elicit the same degree of trait expression (Lievens, de Koster, & Schollaert, 2012). For example, if a candidate is confronted with an upset client (i.e. the

cue) in a role-play exercise, this situation should strongly evoke traits such as Emotional

Stability but may also elicit Conscientiousness in varying degrees.

Since AC exercises differ in their ability to elicit certain traits it is unlikely that

researchers will find consistent behaviors for dimensions across exercises (Lievens et al.,

2012). Exercises that are similar in relevance and trait activation will result in higher

rating convergence. Lievens et al. (2012) provide explanation for this position:

> For example, consider ratings on the dimension of Interpersonal Influence, which
>
> are based on behaviors that are expressions of the Big Five Trait of Extraversion.
>
> As a leaderless group dynamics and role-play exercise can be both expected to
>
> provide cues relevant to this trait, convergence between ratings should be
>
> expected (p. 224 – 225).

Research has also examined how personality traits are elicited across varying

exercises. Haaland and Christiansen's (2002) research linked AC dimensions to the Big

Five personality traits. The ratings demonstrated stronger convergence when they were

rated in exercises deemed high in trait activation potential. In comparison, exercises

deemed low in trait activation potential showed lower convergence across dimension

ratings. Research from Lievens and colleagues (2006) supported Haaland and

Christiansen's findings. Convergence was stronger between dimensions deemed higher in

trait relevance that elicited the target dimensions. For example, the competitive leaderless

group dynamics, oral presentation and tolerance for stress dimensions were classified as

high in trait activation potential for the FFM trait Emotional Stability. These three

exercises had a strong correlation with Emotional Stability ($r = .45$) whereas the

remaining exercises classified as low in trait activation potential had a weaker correlation

($r = .34$). Moreover, results indicated that discrimination between ratings within exercises was stronger for AC dimensions that did not elicit the same underlying traits. Specifically, discrimination was higher for exercises that involved more complex interaction between the candidate and others such as assessors and peers (e.g. role-play exercise).

The typical approach to examining AC construct validity through analyzing all of the correlations among dimensions was discussed earlier. According to TAT, convergence should not be expected equally across dimensions evaluated in different exercises because each exercise presents differing trait activation potential and relevance (Lievens et al., 2012). Thus, convergence should only be expected between exercises that provide similar trait activating opportunities (Lievens et al., 2006). Lievens and colleagues (2006) found support for this application of TAT noting that theoretically, trait activation provides a framework for better understanding construct validity in ACs.

Research supports the application of TAT and has previously shown promising results. Therefore, the trait activation potential of exercises should be taken into consideration when examining AC performance. To investigate the influence of personality on AC performance, the dimensions and exercises that comprise the MPOAC were examined for relevance according to the TAT. Based on the results of this examination, hypotheses were formulated.

Personality

The five factor model (FFM) of personality has been one of the most widely used approaches to the trait theory of personality. The FFM model is made up of five personality factors: Conscientiousness, Neuroticism, Extraversion, Agreeableness, and

Openness (Costa, & McCrea, 1992). The following trait descriptions are from Barrick

and Mount's (1991) research. Conscientious individuals are dependable, hardworking,

achievement-oriented and persevering. Neuroticism (or Emotional Stability) includes

factors such as anxiety, depression, anger, emotionality and insecurity. Extraversion is

typically associated with being sociable, gregarious, assertive, talkative and active.

Agreeable individuals tend to be flexible, trusting, good-natured, forgiving, and tolerant.

Lastly, Openness includes traits such as being imaginative, curious, original, broad-

minded and intelligent.

Meta-analyses conducted over the past two decades using the FFM framework

have demonstrated that personality traits are valid predictors of job performance (e.g.

Barrick & Mount, 1991; Tett, Jackson, & Rothstein, 1991; Oh, Wang, & Mount, 2011).

Salgado (2003) found that in estimating performance ratings, the operational validity for

Conscientiousness and Neuroticism was higher in FFM models (.28 and .16, respectively)

in comparison to non-FFM models (.18 and .05, respectively). Oh, Wang, and Mount

(2011) conducted a meta-analysis of observer ratings of the FFM traits and overall job

performance. The results indicated that the estimated mean true validities were highest

when the average number of raters per trait was higher (i.e. 1 vs. 3): Conscientiousness (r

=.31), Agreeableness (r = .23), Emotional Stability (r = 17), Extraversion (r = .21) and

Openness (r = .20).

The relationship between personality traits and AC ratings is complex, evidenced

by mixed associations found in the literature (Meriac et al., 2014). In a recent study, only

Extraversion and to a lesser extent, Openness, correlated with AC dimensions (Meriac et

al., 2014). However, an earlier study found that the OAR could be predicted from

Agreeableness, Emotional Stability, Extraversion, and Openness; with Extraversion having the strongest association (Collins, Schmidt, Sanchez–Ku, Thomas, McDaniel, & Le, 2003). Relative to findings examining cognitive ability as a predictor of performance, these correlations are low. However, stronger relationships have been found between personality traits and exercises such as role-plays that tend to be more interpersonally oriented (Spector et al., 2000).

Conscientiousness. Exercise D and Exercise E are distinct exercises designed to elicit multiple dimensions. Exercise D requires candidates to complete a complex, administrative task that requires attention to detail and the ability to process information in a limited timeframe. Exercise E requires that candidates organize and interpret information and present potential solutions to a problem. Conscientiousness captures personal characteristics that are key for accomplishing tasks, such as those during Exercise D and Exercise E. Individuals who are more conscientious are likely to perform better during these exercises because of their hard working, achievement-oriented, and orderly disposition.

Compared to other exercises, the situations presented in the Exercise D and Exercise E should present the relevant cues to elicit Conscientiousness more strongly. Thus, the relationship between AC ratings and Conscientiousness should be stronger in Exercise D and Exercise E because these exercises strongly activate this trait. Therefore, it is hypothesized that:

Hypothesis 2: The relationship between dimension ratings of conscientiousness and TSD-PI Conscientiousness will be stronger for exercises that have high activation potential

(i.e., Exercise D and Exercise E) than for exercises that have low activation potential (i.e., Exercise A, Exercise B, Exercise C, and Exercise F).

Extraversion. Extraversion encompasses traits such as sociability, assertiveness, activity, and talkativeness (Barrick & Mount, 1991). Extraverted individuals are often described as energetic and optimistic (Rothmann & Coetzer, 2003). In Barrick and Mount's (1991) meta-analysis, Extraversion emerged as a valid predictor of job performance for two occupations that involved social interaction. In a study examining personality correlates of AC exercises, Extraversion was the strongest personality correlate of performance ratings in the three interpersonal exercises (Hoffman, Kennedy, Lopilato, Monahan & Lance, 2015). These exercises included a role-play exercise, leaderless group dynamics exercise, and oral presentation.

Extraversion is not measured directly in the MPOAC; however, the personal impact dimension can be directly linked to this trait. Personal impact is designed to measure the degree to which a candidate is confident, commands respect and attention, takes charge and positively handles his or her limitations. Those high in Extraversion may achieve higher ratings in comparison with others when social interaction is at the core of an assessment (e.g. group work).

Compared to other exercises, Exercise A, Exercise F, and Exercise C should present strong and relevant cues to foster extraverted behaviors. Exercise C and both Exercises A and F place candidates in a scenario where they interact with either an assessor or other candidates. Specifically, Exercise C requires that candidates interact with an assessor who is demanding a solution to their issue. During exercises A and F,

candidates work in small groups to develop consensus-driven solutions to the presented problems.

Thus, the relationship between the AC rating of personal impact and Extraversion should be stronger in the Exercise A, Exercise F, and Exercise C because these exercises strongly activate this trait. It is hypothesized that:

Hypothesis 3: The relationship between dimension ratings of personal impact and TSD-PI Extraversion will be stronger for exercises that have high activation potential (i.e., Exercise A, Exercise F, and Exercise C) than for exercises that have low activation potential (i.e., Exercise B, Exercise E, and Exercise G).

Emotional Stability. Emotional Stability is measured during the MPOAC using the dimension labelled performance under stress. Individuals with high Emotional Stability are able to keep their emotions under control when faced with stressful circumstances. When stressed, they are able to effectively manage which prevents stressful stimuli from negatively impacting their performance. Therefore, individuals who have low Emotional Stability possess characteristics that may negatively impact their ability to successfully perform under the circumstances created during the AC (Barrick & Mount, 2001).

Research conducted by Spector and colleagues (2000) found that ratings of interpersonal exercises correlated with Emotional Stability. Although candidates are placed under increased stress due to continuous evaluation throughout the MPOAC, interpersonally challenging exercises may increase stress for candidates. Candidates participate in two interpersonally challenging exercises during the MPOAC. Part of Exercise E requires candidates to deliver a timed presentation and respond to challenging

questions. Exercise C has candidates assume the role of a military police officer and interact with trained assessors who act as a client and subordinate. These stimuli are designed to elicit the performance under stress dimension and evaluate interpersonal skills.

Exercise E and Exercise C should present the relevant cues to strongly elicit performance under stress in comparison to other exercises. Candidates present or act while they are being independently evaluated. The relationship between AC ratings and Emotional Stability should be stronger in Exercise E and Exercise C because these exercises strongly activate this trait. From a TAT perspective, it is hypothesized that:

Hypothesis 4: The relationship between dimension ratings of performance under stress and TSD-PI Emotional Stability will be stronger for exercises that have high activation potential (i.e., Exercise C, Exercise E, and Exercise G) than for exercises that have low activation potential (Exercise A, Exercise B, and Exercise F).

Cognitive Ability

Cognitive ability testing has maintained a prominent position in personnel selection for decades. Although definitions vary, cognitive ability can generally be defined as an individual's ability to correctly grasp and reason with concepts and solve problems (Schmidt & Hunter, 2004). Cognitive ability, or mental ability, is known to be correlated with one's performance on complex, cognitive tasks (Hunter & Schmidt, 1998). Cognitive ability significantly predicts academic and job performance across a range of occupations (Ones, Viswesvaran, & Dilchert, 2005). Meta-analytic results found that cognitive ability has an average validity coefficient of .45 in predicting training performance across range of occupational families (Hunter & Hunter, 1984).

Cognitive ability also increases with the complexity of the job (Schmidt & Hunter, 2004). The MPOAC is designed to evaluate potential military police officers who are charged with providing leadership and professional direction to military police members. Additionally, MPOs manage resources and equipment, enforce discipline, and manage the collection, collation, analysis and dissemination of criminal intelligence. Given the complexity of this military occupation, cognitive ability should be a strong predictor of job success. Cognitive ability (or general mental ability) is also more strongly associated with problem-solving oriented dimensions.

Cognitive ability and TAT. Trait activation theory emphasizes that individual differences are activated in situations where they are cued, resulting in expression of relevant traits (Tett & Burnett, 2003). Cognitive ability can unfold in situations to the same extent as personality traits in response to cognitive ability relevant situational cues (Tett & Burnett, 2003). Therefore, we suggest using the TAT framework to further understand the role of cognitive ability in AC ratings.

Arthur, Day, Mcnelly, & Edens (2003) defined an individual's problem solving as the ability to effectively gather information, analyze data, and generate solutions to problems that are presented. These aptitudes are central to cognitive ability (see Schmidt & Hunter, 2000). The 'problem solving dimension' evaluated in the MPOAC aligns with this definition, which can be associated with cognitive ability. Dilchert & Ones (2009) showed that out of seven AC dimensions, problem solving had the strongest relationship with cognitive ability. For example, candidates with lower cognitive ability are expected to have lower problem-solving abilities which will lead to poorer performance during exercises where problem solving is essential. Specifically, lower problem-solving skills

are expected to have less ability to work through the details of the presented problem, thus not reaching a satisfactory solution. Alternatively, those who possess higher cognitive ability are expected to be better able to apply problem-solving abilities when elicited during an exercise. Therefore, there should be a strong association between problem solving and cognitive ability in those exercises where problem solving is strongly elicited.

In the case of the MPOAC, two exercises present the candidates with situations that demand the use of cognitive abilities. Exercise D requires candidates to review a package of items (e.g. e-mails, organizational charts, employee requests) with varying levels of importance and priorities. Exercise E presents candidates with a management problem for which they must design and deliver possible solutions. These two exercises focus on information processing and problem-solving aspects of one's cognitive ability. For good performance on Exercise D and Exercise E, a candidate would have to demonstrate the requisite cognitive-abilities, such as problem solving. Therefore, it is hypothesized that:

Hypothesis 5: The relationship between CFAT scores of cognitive ability and dimension ratings of problem solving will be stronger for exercises that have strong activation potential (i.e., Exercise D and Exercise E) than for exercises that have low activation potential (i.e., Exercise A, Exercise C, and Exercise F).

Cognitive Ability x Personality Interaction and AC Performance

Lievens, Tett, and Schollaert (2009) argue that a full understanding of AC performance requires the integration of ability and personality. To date, limited research in the AC literature has examined the interaction of cognitive ability and personality on

AC performance (Spector et al., 2000). However, personality-performance studies have examined the potential moderating influence of cognitive ability on job performance.

Conscientiousness and Cognitive Ability

Individuals who are conscientious are often deemed to be self-disciplined and achievement oriented (McCrae, & Costa, 1987), characteristics that are valued for performance. In more extreme cases, highly conscientious individuals may also be viewed as perfectionists who are rigid and inflexible (Le, Oh, Robbins, Ilies, Holland, & Westrick, 2011). These perfectionist tendencies result because these individuals may be predisposed to focus too much on small details, overlooking more important goals when working (Mount, Oh, & Burns, 2008). Therefore, an individual's performance may be curvilinear. Despite the possible negative influence of high levels of Conscientiousness on performance, Conscientiousness generally has a positive association with job performance.

The role of Conscientiousness may be moderated by cognitive ability. Research by Lowery, Beadles, and Krilowicz (2004) found that the interaction between cognitive ability and need for achievement explained incremental variance in employee performance beyond that of these factors individually. Those high in cognitive ability and achievement need performed better in comparison with those who had low cognitive ability and high achievement need. Achievement striving and achievement need are similar and fall under the global trait Conscientiousness (McCrae & Costa, 1987). These results reflected the earlier findings of Wright and colleagues (1995). However, Mount, Barrick, and Strauss (1999) found no support for the interaction of general mental ability

and Conscientiousness. Perry, Hunter, Witt, and Harris (2010) found that only the achievement facet of Conscientiousness interacted with cognitive ability.

This interaction between Conscientiousness and cognitive ability has yet to be tested in the AC performance context. As a population, AC candidates are likely to feel pressure to maximize their performance throughout the entire process. This may be especially salient for ACs that assess entry-level applicants, such as the MPOAC, where a job offer is performance dependent. It is expected that MPOAC candidates will use their cognitive resources to strengthen their performance. Therefore, people with both high Conscientiousness and cognitive ability will outperform those who were highly conscientious and low in cognitive ability, and low in both Conscientiousness and cognitive ability.

Hypothesis 6: Cognitive ability will moderate the relationship between Conscientiousness and AC performance such that the relationship will be stronger for those who are also high in cognitive ability.

Extraversion and Cognitive Ability

Extraversion has been frequently evaluated in the context of performance. Barrick and Mount's (1991) findings suggest that Extraversion is a valid predictor of job performance for jobs considered high in social interaction. Meta-analytic findings from Meriac, Hoffman, Woehr, and Fleisher (2008) showed that when each of the big five personality dimensions was regressed on seven AC dimensions, Extraversion was the strongest personality correlate ($r = .25$). More specifically, earlier research suggested that Extraversion is one of the strongest personality correlates of performance in interpersonal exercises (Spector et al., 2000; Hoffman et al., 2015). Overall, candidates who are high

on Extraversion perform better than those who are low in Extraversion; however, having high Extraversion may not result in better performance in the presence of other attributes, such as general mental ability.

Extraversion and general mental ability were the most important traits affecting AC performance (Collins et al., 2003). The interaction between social skill and general mental ability suggests that "possessing GMA provides those with higher levels of social skill the 'boost' necessary to increase performance and salary" (Ferris, Witt, & Hochwarter, 2001, p. 1081). The attributes of Extraversion that improve performance may be limited by one's cognitive ability. Highly extraverted individuals may be sociable, active, and assertive in situations, but if they have low cognitive ability, they may be ineffective at processing information and generating solutions.

Hypothesis 7: Cognitive ability will moderate the relationship between Extraversion and AC performance such that the relationship will be stronger for those who are also high in cognitive ability.

Potential Adverse Impact

If different groups of individuals score differently on selection tools or methods, adverse impact can occur (Hough, Oswald, & Ployhart, 2001). Specifically, adverse impact occurs when the rate of selection for a specific group of individuals (e.g., female applicants) is lower than the rate for the majority group (Ng & Sears, 2010). Sub-group differences exist on cognitive ability testing. To date, there is an average of one standard deviation difference in scores for Blacks and Whites, with the latter achieving higher scores (Hough et al., 2001; Pyburn, Ployhart, & Kravtiz, 2008). Overall, there are negligible differences for males and females (Hough et al., 2001); however, research has

highlighted differences for specific aptitudes, such as spatial ability. Neisser and

colleagues (1996) highlight research findings indicating a male advantage on spatial-

visual and mental rotation tasks.

There are minimal sub-group differences for the Big Five personality constructs

between ethnic and cultural groups, as well as sex (Hough et al., 2001). In contrast, there

are significant differences between certain personality facets that advantage one sex over

the other. Hough, Oswald and Ployhart (2001) noted that women scored approximately

.40 standard deviations higher than men on the Agreeableness facet. Women also tend to

generally score higher on the dependability facet of Conscientiousness, but differences

disappear on the achievement facet.

AC research demonstrates that this method has less adverse impact on candidates

than do other selection methods, such as cognitive ability testing. Dean, Roth, and Bobko

(2008) conducted the first meta-analytic review examining sub-group differences (*d*) for

ACs. Research revealed an overall Black-White *d* of 0.52, an overall Hispanic-White *d* of

0.28, and an overall male-female *d* of 0.19. Thus, on average, Whites followed by

Hispanics scored higher than Blacks, with female candidates scoring higher than male

candidates. Additionally, a study examining age differences found that older candidates

were rated lower than younger candidates (Clapham & Fulford, 1997). More recent

research conducted with a sample of officer candidates entering into the British Army

revealed that females tended to be rated higher on interpersonally oriented leadership

style, as well as drive and determination (Anderson, Lievens, van Dam, & Born, 2006).

This is not entirely surprising given that females tend to be rated higher than men on

Agreeableness (Hough et al., 2001) which would bolster interpersonal skills. Dean and colleagues (2008) also concluded that ACs tend to be more female friendly.

Racioethnic differences can surface as a function of the degree to which exercises are "cognitively loaded" (Goldstein, Yusko, & Nicolopoulos, 2001). This could present challenges for AC exercises especially if they are cognitively loaded (i.e. in-basket exercise). Therefore, Ployhart and Holtz (2008) concluded that "the extent to which assessment centers reduce racioethnic differences is a function of how much they also measure cognitive ability" (p. 166). In the case of the MPOAC, cognitive loading is likely to have little effect because candidates are subject to a selection bias in that, to attend the AC, candidates must be at or above the 30th percentile of the Canadian Forces Aptitude Test (CFAT).

Taken together, it is important that the fairness of individual operational ACs be evaluated because it cannot be assumed that a particular AC has limited adverse impact based on previous research. Further, adverse impact may not be apparent in one sample but a consideration for another set of circumstances (e.g. non-military developmental AC vs. an entry level selection military AC).

Anderson, Lievens, van Dam and Born's (2006) research revealed that notable differences between male and female candidates in the context of a leadership-role assessment center. Specifically, female candidates were rated higher on interpersonal constructs, such as oral communication and interaction. Dean, Roth and Bobko's meta-analytic review concluded that females outperformed males in ACs ($d = .19$). Given these research findings it is hypothesized that:

Hypothesis 8a: There will be a significant gender difference favoring female candidates on overall assessment center performance.

In addition to the final overall assessment rating, candidates are also classified as either suitable or unsuitable. The final decision to classify a candidate as suitable or unsuitable for employment as an MPO is heavily influenced by the OAR. However, other factors may influence this final decision including clarifications from background checks, changes to medical eligibility, and a candidate's integrity. Candidates who were rated below 3.0 on a 5-point behavioral anchored rating scale (ranging from 1 – 5) for the integrity dimension are classified as unsuitable, regardless of the OAR. Given these external factors and their influence, a candidate's OAR and final classification (suitable vs. not suitable) may differ. To detect possible difference between the OAR and a candidate's final classification, it is hypothesized that:

Hypothesis 8b: There will be a significant gender difference favoring female candidates on the final suitability classification.

Prior to attending the MPOAC, all candidates complete the Canadian Forces Aptitude Test (CFAT). The English and French version of the test are considered separate tests, which required that development of two norms (Boswell & Kuschnereit, 2009). Anglophone and Francophones must meet the same percentile to be considered eligible. Since these percentile cut-offs correspond to differing raw scores, it is possible that Anglophones and Francophones have differences in AC performance. To explore this possibility, the following research question was posed:

Research Question: Is there a difference between Anglophone and Francophone candidates on overall assessment center performance?

Method

Sample

The AC data used were archival, collected during ACs that were organized from 2006 to 2017. Prior to attending the AC, the candidates had already passed several selection hurdles which including meeting minimum medical requirements, a cognitive ability threshold, and s structured interview process. The original sample included 403 cases. However, only 299 cases remained after duplicate cases and cases missing AC data were removed. The sample consisted of 299 candidates who attended the MPOAC for selection purposes. These candidates applied for a Military Police Officer position in the Canadian armed forces (CAF). Of this total, 234 were male (78%) and 65 were female (22%). In comparison, the CAF reported that females comprise 15.3% of the CAF population ("Women in the Canadian Armed Forces", 2010). The MP organization reported that females represented 24.3% of the Military Police Officer population (G. Lawlor, personal communication, July 16, 2018). In total, 244 were Anglophone (82%) and 55 were Francophone (18%). In comparison, Anglophones comprise 68.8% of the Military Police Officer population (G. Lawlor, personal communication, July 16, 2018). Prior to arriving at the MPOAC, all of the participants were screened based on cognitive ability, personality, person-job and person-environment fit.

Measures

Cognitive ability. The Canadian Forces Aptitude Test (CFAT) is a cognitive ability measure. It is used by the CAF to screen applicants by determining one's eligibility for military occupations based on the achieved score. The CFAT has three subscales: verbal ability (15 items), spatial ability (15 items), and problem-solving skills

(30 items). Donohue (2005) reported internal consistency reliabilities for the three subscales (as cited in Boswell & Kuschnereit, 2009). The reported reliabilities were the same for both the English and French version of the test: verbal ability ($\alpha = 0.8$), spatial ability ($\alpha = 0.7$), and problem-solving skills ($\alpha = 0.9$) (as cited in Boswell & Kuschnereit, 2009, p. 6).

The CFAT is a timed test which means that items that are not completed during the allocated time are scored as incorrect. There is an English and French version of the CFAT; each version is considered an individual test (Boswell & Kuschnereit, 2009). Research has shown that the CFAT is a valid predictor for the MP entry-level qualification course (Hodgson, 2005; Ebel-Lam & Carter, 2011) and for recruit basic training (Black, 1999).

Personality. The Trait Self Descriptive Personality Inventory (TSD-PI) is a measure based on the five-factor model of personality. In 2012, the TSD-PI was adapted for use in the CAF and introduced to the CAF selection system. The TSD-PI measures Conscientiousness, Emotional Stability (i.e. Neuroticism), Openness to experience, Extraversion and Agreeableness. Darr and Kempt (2009) reported acceptable Cronbach's alpha coefficients, measures of internal reliability, for the TSD-PI scales: Conscientiousness = .81, Emotional Stability = .90, Openness to Experience = .86, Extraversion = .82 and Agreeableness =.91.

The TSD-PI is a self-report measure comprised of 75 items divided into two parts: 26 adjectives and 49 statements. Each personality trait is assessed using 15 items. Sample items include: "organized", "I try to set a schedule for accomplishing tasks and stick to it", and "I always try to do more than what is expected of me". Participants are asked to

rate the degree to which either an adjective describes them (1 = extremely uncharacteristic; 9 = extremely characteristic). There is no time limit for completion.

MPOAC. The MPOAC is a four-day AC comprised of eight exercises that assessed 10 dimensions. The primary dimensions assessed included: integrity, conscientiousness, performance under stress, professional demeanor, personal impact, problem solving, initiative, flexibility, resource management, and written communication skills (see Table 1 for definitions). Eight exercises were used to measure the dimensions.

One average, one assessor rated two candidates during each exercise. Assessors assigned dimension ratings using a 1 (low) to 5 (high) point rating scale; assigning half scale points (e.g. 1.5) was permitted. For each dimension, every full-scale point (e.g. 1, 5) was anchored to a verbal description, or word picture. The aggregation of dimension ratings made during exercises occurred after candidates completed all exercises. Assessors individually rated each candidate on a dimension which was then shared with the team. A final dimension score was assigned by averaging the individually assigned scores. If more than a 1-point discrepancy among assessors occurred, a consensus-based discussion was held. A senior assessor with significant experience at the MPOAC oversaw the discussion process.

Linking dimensions and traits. Existing resources were used to examine potential links between dimensions and traits. The "MP Branch Assessor Training Manual" and the "Scoring Guide" provide the definitions and behavioral indicators for each MPOAC dimension. Established definitions and indicators of the Five Factor Model (FFM) of personality were used to link dimensions with traits. Specifically, using the definitions

and descriptions provided by Barrick and Mount (1991), each MPOAC dimension was mapped onto at least one personality trait.

The performance under stress dimension, for example, captures one's ability to regulate emotions, self-restraint when faced with stressful working conditions, and ability to effectively manage stress. This dimension closely aligns with Emotional Stability (or Neuroticism). According to Barrick and Mount (1991), "common traits associated with this factor include being anxious, depressed, angry, embarrassed, emotional, worried, and insecure (p. 4)"; these factors would be expected from a candidate low in the performance under stress dimension. Table 3 provides a description of the link between MPOAC dimensions and the five factor model personality traits.

Data Analysis

The means, standard deviations, and zero-order correlations of the study variables are shown in Table 4. Archival data collected over numerous ACs was used in the subsequent analyses. The hypotheses were analysed using a number of statistical techniques using the R statistical environment (R Core Team, 2017).

A significant number of these 299 cases were missing all of the AC performance data (i.e., had overall assessment rating only). Removal of these cases resulted in 136 cases of which a majority were missing information on the suitability classification because they had been deleted. In order to retain the largest N, the decision was made to conduct the regression analyses using the 136 cases and the adverse impact testing (i.e. t-test, chi-square test) using a different subset of cases. The subsequent screening of both is discussed below.

Prior to conducting the regression analyses, a subset of 136 cases were assessed for linearity, normality, outliers, multi-collinearity, and missing data. Listwise deletion was performed for all cases with missing values leaving 98 cases for analysis. Following an examination of standardized scores, two univariate outliers that did not follow their respective distributions were identified and removed from the analysis, leaving 96 cases available for further analysis. No multivariate outliers were identified. Residual and scatter plots were examined and indicated that assumptions of normality, linearity, and homoscedasticity were all satisfied.

Exercise H was excluded from the hierarchical multiple regression and moderated regression analyses because this exercise is only conducted with a small number of candidates that require follow-up from another exercise. This resulted in mostly missing data.

Prior to analysis, suitability, overall assessment rating, sex, and language were examined for missing values and violations of analysis. The variables were examined separately for the 144 male and 44 female candidates, and the 153 Anglophone and 35 Francophone candidates. One case was identified as a univariate outlier in both the male and Anglophone categories. As a result, this case was removed from the dataset leaving 187 cases available for analysis. The requirements for independence of observations, normality, and homogeneity of variance were met.

A multitrait-multimethod matrix was constructed to examine construct validity across the MPOAC. Secondly, a series of two-step hierarchical multiple regressions were conducted using personality variables or cognitive ability as the outcome variables. Exercise ratings that were low in activation potential were entered at step one. In step

two, exercise ratings that were high in activation potential were entered. The exercises ratings were entered in this order to determine whether ratings obtained during exercises that are higher in activation potential account for more variance compared to ratings obtained during exercises that are low in activation potential.

Next, moderated regression analyses were conducted to assess if cognitive ability moderates the relationship between specific personality traits (i.e., conscientiousness and emotional stability) and overall AC performance. To examine for moderation, a multiple linear regression was conducted. The independent variables of the regression were the personality trait, cognitive ability, and the interaction between the personality trait and cognitive ability. The interaction was created by multiplying the personality trait and cognitive ability, the moderator, together after both were mean-centered (i.e., have a mean of 0).

Independent sample t-tests and chi-square analyses were used to examine group differences between male and female candidates, and Anglophone and Francophone candidates on performance. Specifically, Welch's t-tests examined group differences on overall performance across the MPOAC. The chi-square tests were used to examine group differences on the final suitability rating assigned to candidates at the end of the MPOAC.

Despite the nested structure of the data, the data were treated as if it came from the same AC due to a number of limitations. Although dates were provided indicating when a candidate attended an AC, the size of the clusters varied considerably (e.g., $n = 2$, $n = 30$) and the number of candidates entered in the datasheet did not correspond to common practice. Typically, two teams of 12 candidates compose an MPOAC serial.

Table 3

AC Dimensions Linked to Five Factor Model Personality Traits

Dimension	Key words	Trait	Key descriptors [a]
Conscientiousness	Meets/exceed deadlines, strives for excellence, efficient, thorough, hardworking, dependable	Conscientiousness	dependable, achievement-striving, hardworking, persevering, orderly
Performance under stress	Controls emotions, shows emotional restraint, effectively manages stress	Emotional Stability	calm, self-confident, resilient
Professional impact	Confident, command respect and attention, effectively take charge, positively handles own limitations	Extraversion	sociable, talkative, assertive, active
Professional demeanour	Effective interactions with others, respectful, appropriate, sensitive, compassionate, respect beliefs and practices of others	Agreeableness	cooperative, flexible, tolerant, forgiving
Initiative	Self-motivated, self-directed, actively influences events, generates new and imaginative ideas	Extraversion	sociable, talkative, assertive, active

Flexibility	Responsive and adaptable to change, receptive to innovation, embraces change	Agreeableness	cooperative, flexible, tolerant, forgiving

[a] = adapted from Barrick and Mount (1991)

Table 4

Means, Standard Deviations, and Correlations

Variable	*M*	*SD*	1	2	3	4	5	6	7
1. Language[a]	0.18	0.39							
2. Sex[b]	0.22	0.41	.13*						
3. Cognitive Ability	43.53	6.50	.17**	-.12*					
4. Conscientiousness	88.02	13.04	-.10	.10	-.12				
5. Extraversion	73.77	15.51	-.07	-.08	-.13	.50**			
6. Emotional Stability	88.63	14.36	-.03	-.08	-.07	.68**	.64**		
7. OAR	68.87	7.65	.26**	.13*	.17**	.01	.17*	.05	
8. Suitability[c]	0.36	0.48	-.25**	-.13	-.10	-.09	-.21	-.13	-.72**

Note. N ranging from 88 to 299. OAR = overall assessment rating; * indicates *p* < .05; ** indicates *p* < .01; [a]Anglophone = 0, Francophone = 1; [b]male = 0, female = 1; [c]suitable = 0, not suitable = 1. *M* and *SD* are used to represent mean and standard deviation, respectively.

Furthermore, the data provided did not include information about who the assessors are, the type of assessor (i.e., selection officer or police member) or the individual dimension ratings assigned by assessors, and thus could not be modeled. Combined with the large amount of missing data and small overall N (i.e., 96), the decision was made to not analyze the data using multi-level modelling.

Results

Construct-related validity

Hypothesis 1 examined the construct related validity of the MPOAC. A multitrait-multimethod matrix was constructed to examine convergent and discriminant validity. In the matrix, the AC dimensions are treated as traits and the exercises are treated as methods (Campbell & Fiske, 1959).

The monotrait-heteromethod correlations (i.e., same dimension, different method of assessment) were examined to determine the degree of convergent validity. These coefficients ranged from $r = -.01$ to $r = .58$ across the 10 MPOAC dimensions. The mean of these correlations ranged from $r = .19$ to $r = .38$.

The heterotrait-monomethod coefficients (i.e., different dimensions, same method of assessment) were examined to determine the degree of discriminant validity. These coefficients ranged from $r = .23$ to $r = .95$ across the 8 MPOAC exercises. The mean of these correlations ranged from $r = .48$ to $r = .78$. In comparison, the heterotrait-monomethod coefficients were stronger than the monotrait-heteromethod coefficients, providing evidence for the presence of an exercise effect. Therefore, Hypothesis 1 was supported.

Trait Activation Theory

Hypotheses 2 through 5 were tested by estimating the parameters of hierarchical regression models predicting the respective personality trait (e.g. Conscientiousness) or cognitive ability (i.e. Hypothesis 5). In each analysis, the variables entered in the first block were ratings from exercises with low activation potential, followed by the block of variables measuring ratings for exercises high in activation potential. The change in model fit (R^2) resulting from entry of the ratings from exercises with high activation potential was examined in order to determine the strength of their contribution in predicting personality and cognitive ability outcomes.

Hypothesis 2 examined the relationship between ratings of conscientiousness and TSD-PI Conscientiousness between exercises that were low in activation potential (Exercise A, Exercise B, Exercise C, Exercise F) and those that were high in activation potential (Exercise D, Exercise E). Results from a hierarchical multiple regression revealed that none of the exercise ratings that were low in activation potential significantly predicted Conscientiousness in candidates ($R^2 = .04$, $F(4,91) = 1.06$, $p = .38$). Contrary to expectations, the addition of the variables that were high in activation potential did not significantly improve prediction ($R^2 = .05$, $F(6,89) = 0.75$, $p = .61$). Unexpectedly, conscientiousness ratings from Exercise A negatively predicted TSD-PI ratings of Conscientiousness when entered in step one ($B = -3.41$, $p < .05$) and in step 2 ($B = -3.61$, $p < .05$). Table 5 contains the summarized results of the hierarchical multiple regression analysis. Overall, these findings do not provide support for Hypothesis 2.

Hypothesis 3 examined the relationship between ratings of personal impact and TSD-PI Extraversion between exercises that were low in activation potential (Exercise B, Exercise E, Exercise G) and those that were high in activation potential (Exercise A, Exercise C, Exercise F).

Table 5

Summary of Hierarchical Regression Analysis Results for Variables Predicting Conscientiousness

Variable	b	SE	R^2	ΔR^2
Step 1			.04	
Exercise A	-3.42*	1.70		
Exercise B	1.00	1.88		
Exercise C	0.02	2.01		
Exercise F	0.39	1.21		
Step 2			.05	.01
Exercise A	-3.61*	1.75		
Exercise B	0.86	1.94		
Exercise C	0.05	2.03		
Exercise F	0.34	1.31		
Exercise D	0.76	1.32		
Exercise E	0.08	1.29		

Note. $N = 96$; *$p < .05$.

As expected, ratings from exercises that were low in activation potential did not significantly predict a candidate's Extraversion. Together, these three exercise ratings explained 3% of variance ($R^2 = .03$, $F(3,92)=1.00$, $p = .40$). One exercise of the three

assessed as high in activation potential was a significant predictor (Exercise F; $B = 5.41$, $p < .01$). The addition of ratings from exercises high in activation potential resulted in the larger model explaining 14% of the variance $(R^2 = .14, F(6,89) = 2.44, p < .05)$. These results are summarized in Table 6. Therefore, there was only partial support for Hypothesis 3.

Table 6

Summary of Hierarchical Regression Analysis Results for Variables Predicting Extraversion

Variable	b	SE	R^2	ΔR^2
Step 1			.03	
Exercise B	3.81	2.47		
Exercise E	0.38	1.65		
Exercise G	-1.40	2.27		
Step 2			.14*	.11**
Exercise B	3.11	2.62		
Exercise E	-2.56	1.84		
Exercise G	-2.62	2.28		
Exercise A	-1.76	2.07		
Exercise B	3.48	2.36		
Exercise F	5.41**	2.05		

*Note. N = 96; *p < .05, **p < .01.*

To test Hypothesis 4, the relationship between ratings of performance under stress and TSD-PI Emotional Stability between exercises that were low in activation potential (Exercise A, Exercise B, Exercise F) and those that were high in activation potential

(Exercise C, Exercise E, Exercise G) was examined. None of the exercise ratings that were low in activation potential significantly predicted TSD-PI Emotional Stability (R^2 =.05, $F(3.92) = 1.56$, $p = .21$). The addition of the ratings from exercises that were high in activation potential did not significantly improve prediction (R^2 =.05, $F(6,89) = 0.85$, $p = .54$). Therefore, these findings do not provide support for Hypothesis 4. Results are presented in Table 7.

Finally, Hypothesis 5 examined the relationship between ratings of problem solving and cognitive ability between exercises that were low in activation potential (Exercise A, Exercise C, Exercise F) and those that were high in activation potential (Exercise D, Exercise E).

Table 7

Summary of Hierarchical Regression Analysis Results for Variables Predicting Emotional Stability

Variable	b	SE	R^2	ΔR^2
Step 1			.05	
Exercise A	-0.20	1.58		
Exercise B	-2.31	1.67		
Exercise F	2.64	1.39		
Step 2			.05	.00
Exercise A	-0.22	1.65		
Exercise B	-2.48	1.85		
Exercise F	2.27	1.57		
Exercise C	0.69	1.86		
Exercise E	0.75	1.48		

| Exercise G | -0.69 | 1.50 | | |

Note. $N = 96$.

Exercises low in activation potential did not significantly predict a candidate's cognitive ability. These three low activation potential exercise ratings explained 4% of variance ($R^2 = .04$, $F(3,92) = 1.28$, $p < .29$). One of the two exercises high in activation potential was found to significantly predict a candidate's cognitive ability (Exercise D; $B = 1.83$, $p < .01$). Adding the two exercise ratings that were high in activation potential resulted in the larger model explaining 15% of the variance ($R^2 = .15$, $F(5,90) = 3.14$, $p < .01$). Specifically, Exercise D was found to more strongly elicit problem solving because it is high in trait activation potential. Thus, Hypothesis 5 received partial support. The results are summarized in Table 8.

Table 8

Summary of Hierarchical Regression Analysis Results for Variables Predicting Problem Solving

Variable	b	SE	R^2	ΔR^2
Step 1			.04	
Exercise A	-0.47	0.98		
Exercise C	-1.31	0.87		
Exercise F	0.68	0.66		
Step 2			.15**	.11
Exercise A	-.56	0.94		
Exercise C	-1.88*	0.85		
Exercise F	.68	0.68		

Exercise D	1.83**	0.59
Exercise E	0.43	0.58

Note. N = 96; *p* < .05, **p* < .01.

Interaction of Personality and Cognitive Ability

Two moderated regression analyses were carried out to examine moderating role of cognitive ability between personality and AC performance (hypotheses 6 and 7). In these moderated regression models, the control variables (language and sex) and two independent variables were entered in the first step followed by entry of the personality trait x cognitive ability interactions in the second step.

Hypothesis 6 stated that cognitive ability would moderate the relationship between Conscientiousness and AC performance. The positive relationship between Conscientiousness and overall AC performance was predicted to be stronger for candidates who have higher cognitive ability. The Conscientiousness x cognitive ability interaction did not account for unique variance in overall AC performance ($\Delta R^2 = 0.00$). These results indicate that the relationship between Conscientiousness and overall AC performance is not moderated by cognitive ability. Thus, Hypothesis 6 is not supported. Table 9 contains the results of this moderated hierarchical regression analysis.

Finally, Hypothesis 7 stated that cognitive ability would moderate the relationship between Extraversion and AC performance. The positive relationship between Extraversion and overall AC performance was predicted to be stronger for candidates who have higher cognitive ability. The findings showed that the Extraversion x cognitive ability interaction did not account for unique variance in overall AC performance ($\Delta R^2 =$

.01). These results show that the relationship between Extraversion and overall AC

performance is not moderated by cognitive

Table 9

Regression Analyses for Cognitive Ability Moderating the Relationship Between Conscientiousness and Overall Assessment Center Performance

Variable	b	SE	R^2
Step 1			.11
Language	4.13	2.22	
Sex	3.86*	1.89	
Conscientiousness	0.02	0.09	
Cognitive Ability	0.21	0.17	
Step 2			.11
Language	4.07	2.30	
Sex	3.86*	1.90	
Conscientiousness	0.02	0.09	
Cognitive Ability	0.22	0.17	
Conscientiousness*Cognitive Ability	0.00	0.02	

*Note. N = 96; *p < .05*

ability. These findings do not support Hypothesis 7. A summary of these results is

presented in Table 10.

Adverse impact

Hypothesis 8a stated that there would be significant gender differences favoring

female candidates on overall assessment performance. An independent-samples *t*-test and

Cohen's *d* was used to determine the size of any differences between males and females

on overall assessment center performance. There was not a significant difference in the

scores for males ($M = 69.2$, $SD = 7.97$) and females ($M = 70.4$, $SD = 6.90$) conditions,

$t(81.3) = -0.95$, $p = 0.348$. These results suggest that sex does not have an effect on a

candidate's overall assessment center performance rating. Further, Cohen's effect size

value ($d = .15$) reflects a small effect size. These findings do not support Hypothesis 8a.

Table 10

Regression Analyses for Cognitive Ability Moderating the Relationship Between
Extraversion and Overall Assessment Center Performance

Variable	b	SE	R^2
Step 1			.13
Language	4.31	2.18	
Sex	4.42*	1.89	
Extraversion	0.11	0.07	
Cognitive Ability	0.24	0.17	
Step 2			.14
Language	3.80	2.24	
Sex	4.60*	1.88	
Extraversion	0.12	0.07	
Cognitive Ability	0.30	0.17	
Extraversion *Cognitive Ability	-0.01	0.01	

*Note. N = 96; *p < .05*

To further explore a possible effect of sex, Hypothesis 8b stated that significant gender differences favoring females would also be found in the final suitability classification. A chi-square test was calculated to compare the frequency of being assigned a final suitability classification (i.e., suitable or not suitable) for male and female candidates. A significant interaction was not found ($\chi^2(1) =3.10$, $p = .078$) suggesting that women were not more likely then men to be assigned to the suitable classification. Further inspection of the odds ratio (.51) showed that males are approximately half as likely as female candidates to be assigned to the suitable category. Overall, Hypothesis 8b was not supported.

The research question investigated whether any differences between Anglophones and Francophones candidates on overall assessment center performance. An independent samples t-test was performed comparing the mean scores of Anglophones and Francophones on their overall assessment ratings. Results indicated that Francophones ($M = 73.6$, $SD = 5.66$) achieved higher ratings than Anglophones ($M = 68.5$, $SD = 7.84$), $t(67.8)= -4.45$, $p < .001$. The effect size for this analysis ($d = .68$) exceeded Cohen's (1988) criteria for a medium effect ($d = .50$). An exploratory analysis was conducted to compare the frequency of being assigned a final suitability classification (i.e., suitable or not suitable) for Anglophone and Francophone candidates. A significant interaction was found ($\chi^2(1) = 11.40$, $p < .001$, odds ratio = .18) showing that Francophones were almost five times more likely than Anglophones to be assigned to the suitable classification.

Discussion

This study was conducted with several research goals in mind: (a) examining the construct validity of the MPOAC through traditional statistical methods and through the

application of TAT, (b) investigating how the interaction between personality and cognitive ability influence candidate performance, and (c) investigating the presence of adverse impact between candidates based on gender and first official language. The key findings from this study are discussed with respect to each of the research goals together with some practical and theoretical implications, a discussion of limitations, and possible areas for future research.

Traditional Approach to Construct Validity

The traditional approach to construct validity suggested that ratings of the same dimensions rated across exercises will be substantially correlated across the MTMM matrix (Campbell & Fiske, 1959). However, research using MTMM matrices often finds that cross-exercise ratings are not the strongest correlations. Our research using data from the MPOAC showed a pattern of lower correlations among the ratings of the same dimension across exercises and higher correlations among ratings of different dimensions in each exercise. This reflects the presence of an exercise effect. These findings are comparable to existing studies that have examined MTMM matrices in the context of AC construct validity.

Using data from three ACs, Sackett and Dreher (1982) examined the interrelationships among dimension ratings between and within exercises and found that overall, correlations for different traits measured in the same method (heterotrait-monomethod correlations) were higher than correlations for the same trait measured across different methods (monotrait-heteromethod) correlations (e.g., .40 compared to .11). Janesen and Stoop (2001) also found a stronger exercise effect and a weaker

dimension effect. Taken together, these findings highlight the prominence of exercise effects among MTMM studies.

Petrides and colleagues (2010) proposed these findings resulted from a tendency to rate candidates in terms of general performance during each exercise instead of rating candidates' specific behaviors. Halo bias (or error) reflect an assessor's propensity to assign similar scores across distinct dimensions (Murphy & Cleveland, 1995). In the case of the MPOAC, the scoring method used may contribute to the presence of exercise effects. Throughout each exercise, the assessor records the candidate's behavior for all of the relevant dimensions in their notes. After the exercise has concluded, the assessor reviews the scoring guide and assign scores based on their notes. This scoring process creates a situation for halo bias to occur because one assessor rates all dimensions across a given exercise following its completion.

Another explanation for these findings is that different exercises elicit traits differently based on the relevance and strength of the situation. Therefore, one exercise many not as strongly elicit a given dimension as another exercise. This would provide an explanation for the performance inconsistencies reflected across exercises. If this is the case, we would not expect same trait – different method correlations to be uniformly strong across the MTMM matrix. We explored this possibility in the next series of hypotheses.

Trait Activation Theory Approach to Construct Validity

Trait activation theory provides a deeper theoretical explanation for behavioral variations in candidates across exercises (Lievens, 2009). In this study, the FFM framework and exercise descriptions were used to identify exercises that had high and

low trait activation potential for specific traits. This study focused on the FFM traits of Conscientiousness and Extraversion because past research has found that trait activation worked best for these traits (Lievens et al., 2006). However, Emotional Stability and cognitive ability were also included in our exploration. The results from across four hierarchical multiple regression analyses showed partial support for two hypotheses.

As expected, there was some evidence that exercises high in trait activation potential were better predictors of true scores of personality traits than those low in activation potential. This support was partial because only one of three exercises high in activation potential successfully predicted Extraversion. However, when exercises high in activation potential were added to the equation we were able to explain significantly more variance in true scores of personality traits. The exercise rating that predicted candidate Extraversion was obtained during Exercise F.

One contribution of this study is that the results showed partial support for using TAT for characteristics that are not strictly personality traits. Cognitive ability was used as a 'trait' and exercises were coded as being low or high in trait activation potential. Only the problem-solving ratings from Exercise D emerged as significant predictor of overall cognitive ability. This is not surprising because past research has demonstrated that ratings from this exercise type can be predicted from an assesses' cognitive ability (Spector et al., 2000). However, Exercise E was also expected to be high in activation potential and thus a significant predictor. One possible explanation for this finding is that Exercise D primarily required candidates to use cognitive problem-solving skills and did not require candidates to interact with others. Although Exercise E is largely a problem-solving task, candidates are required to deliver a presentation and respond to questions

which introduces an interpersonal aspect to the exercise. Therefore, this exercise may not possess as much activation potential as Exercise D.

No support was found for the distinction between low and high activation potential exercises for Conscientiousness and Emotional Stability. In both cases, the initial model that included exercises low in activation potential explained a small amount of variance in TSD-PI Conscientiousness and Emotional Stability. When added to the regression model, the exercise ratings that were high in activation potential did not significantly improve the overall explained variance. Thus, the overall variance that was explained in the true scores of personality was low in both models.

One possible explanation for these findings is that due to cognitive demands placed on the assessors, cognitive overload may occur and influence the dimension ratings assigned by the assessors. Gaugler and Thornton III (1989) suggested that cognitive overload may be a potential source of error. Effectively, the assessors' abilities to effectively discriminate between each dimension becomes confounded. At the time of this study, the MPOAC used 10 dimensions to assess candidates. Moreover, depending on staffing, assessors may be required to evaluate two candidates during an exercise. Exercises range in the dimensions that assessors rate; however, this can range from a low of five to a high of nine. To limit the cognitive load of assessors, researchers recommend that a maximum of five dimensions be rated per exercise (Gaugler & Thornton III, 1989). Future consideration should be given to the cognitive load placed on assessors during evaluation.

Another possible reason for the low discrimination between exercises with high and low activation may relate to the assessors employed during the ACs. MPOAC serials

employed a mix of personnel selection officers and military police officers or senior

military police members; this core assessor team varied with each serial. Personnel

selection officers have at least an honours degree in psychology and some working

experience. This cannot be said of military police personnel. Researchers have

recommended using psychologists or experienced HR personnel as assessors (Lievens,

2002). Experienced assessors been found to have higher accuracy in distinguishing

between dimensions expressed in candidate performance than inexperienced assessors

(Kolk, Born, Van Der Flier, & Olman, 2002).

Unexpected results were found when the problem solving and conscientiousness

dimensions were examined. Once ratings of problem solving from exercises high in

activation potential were entered, the problem-solving rating from Exercise C became a

significant predictor of cognitive ability, but not in the expected direction. A similar

finding for Conscientiousness occurred with the conscientiousness ratings from Exercise

F. Therefore, the cues contained within these exercises may not be eliciting trait-related

behaviors that are directly related to the target dimensions resulting in poor prediction.

Consideration could be given to redesigning the exercise (e.g., the situation cues) or

dropping the assessment of these dimensions in these exercises.

An alternative explanation is that the MPOAC exercises may not be functioning

as expected. Despite being designed to reflect core job tasks, the exercises may not have

elicited the core dimensions. For example, the conscientiousness dimension evaluated in

Exercise F may be confounded with dimensions that reflect interpersonal skills, such as

Extraversion and Agreeableness, which are more salient to this particular exercise.

Therefore, this exercise is not affording assessors that opportunity to evaluate a candidate's conscientiousness as expected.

In light of these results, we found some limited support for using trait activation theory to increase our understanding of AC construct validity. Few researchers to date have used TAT to examine construct validity. Haaland and Christiansen (2002) found clear evidence for the application of TAT; however, they conducted research using data from a promotional assessment center. The current study uses a military population of candidates who are entry-level job applicants. Thus, further exploration is required to determine the differences between ACs as well as populations.

Cognitive Ability as a Moderator of AC Performance

Next, the influence of cognitive ability and personality on candidate performance was investigated. Specifically, we tested cognitive ability as a moderator for the relationship between personality and overall performance. Although past research on job performance found that the interaction between cognitive ability and personality accounted for additional variance (Wright et al., 1995), our results did not replicate these findings in the AC context. Cognitive ability failed to moderate the relationship for both Conscientiousness and Extraversion. Additionally, there was no main effect of Conscientiousness, Extraversion, and cognitive ability on AC performance. However, significant yet small correlations were found between both Extraversion and cognitive ability, and the performance measure used, the OAR (see Table 3).

This study's operationalization of candidate performance was the overall assessment rating, an aggregate of performance ratings from across the 10 dimensions. This method of capturing performance does not differentiate performance at the

dimension or exercise level. The broad definition of performance may have contributed to the non-significant predictive ability for Conscientiousness and Extraversion. Future research should consider evaluating potential moderators and other relationships at the exercise level (e.g., Spector et al., 2000).

Adverse impact

The analysis of subgroup differences, contrary to expectations, did not demonstrate a significant gender difference favouring female candidates on overall assessment center performance and suitability classification. Therefore, male candidates were not disproportionately rejected in comparison to their female counterparts. Although the results were not statistically significant, the results of the chi-square analysis examining male – female differences approached significance in the hypothesized direction. Female candidates had an increased chance of being classified as suitable (75%) compared to males (60%).

In response to the research question exploring potential differences between Anglophones and Francophones, findings showed that on both the overall assessment rating and suitability classification, Francophones were statistically more likely to achieve higher ratings and to be classified as suitable. One possible explanation for this finding is rooted in how MPOAC serials are conducted. Anglophone and francophone candidates theoretically attend the same AC; however, candidates attend serials in their first official language. The Francophone serials occur less frequently due to the lower number of Francophone applicants, as reflected in this study's sample size. This imbalance may create a higher demand for viable Francophone candidates leading to

more leniency for these candidates during the MPOAC and at the board when final suitability is determined.

Unfortunately, we were not able to investigate subgroup differences for other potentially relevant factors that may have contributed to final rating and decision. For example, research has established that there is on average a one standard deviation difference in intelligence test scores for Blacks and Whites (Pyburn et al., 2008). Given that AC exercises can present a significant demand for mental abilities (e.g., the in-basket exercise), it is possible that subgroup differences exist for different ethnicities. In terms of age, a study that examined an AC that was used for the selection and development of lower-grade managers found that the selection decision was biased by age (Petrides, Weinstein, Chou, Furnham & Swami, 2010). More candidates in the 30 – 39 age brackets were offered more position than those in the 20 – 29 age bracket. Given that the MPOAC is selecting entry level leaders and, to an extent, managers, age may be an advantage because their experience may be highly valued by assessors.

Theoretical and Practical Implications

This research drew on trait activation theory to move beyond MTMM conceptualizations of AC construct validity. Through adopting the view that exercises possess high or low trait activation potential, this study adds to our understanding of how dimensions are expressed through ratings. Despite only finding some partial support, the results demonstrated that for Extraversion and cognitive ability, exercises high in trait activation potential are superior at explaining variances in a candidate's true scores. Also, we were able to demonstrate some support for using TAT for constructs other than personality traits. Overall, these findings highlight the importance of considering the role

of situational strength and trait relevance in eliciting desired behaviors, thus moving past the past belief that dimensions reflect stable traits.

Gender and ethnicity are often considered important factors to considering when investigating the potential for adverse impact. This research demonstrated that, at least in a Canadian context, potential differences in performance and selection ratios for Anglophones and Francophones should be examined. Although further investigation is needed to definitively determine if this finding would be replicated in other ACs, this information can provide valuable guidance for evaluating current MPOAC practices. For example, AC instructions should provide clear directions on the degree that administrative and assessing staff can provide clarification for exercise instructions given to candidates to ensure that candidates in French and English serials are administrated similarly.

The design of the MPOAC should take into consideration. Empirically, a low number of dimensions and behavioral checklists have been shown to increase dimension variance and reduce the cognitive load on assessors (Lievens & Conway, 2001). Currently, the MPOAC uses 10 dimensions. In most exercises the assessors must rate at least six or seven. Consideration should be given to reducing the number of dimensions that are rated by assessors. Alternatively, an assessment process that tests cognitive ability and personality, coupled with a panel interview related to dimensions would be a cheaper and potentially equally valid alternative to an AC. However, this may be more validly accomplished once criterion validity is established, and the predictive validity of the AC's dimensions and exercises is established.

Limitations

Despite the previously discussed results, this research is not without limitations that should be taken into consideration. Specific limitations include the archival data itself, the use of self-report measures, and range restriction.

First, the dataset used was archival and despite a sample size of 402 cases, there was a significant amount of missing data on key variables which resulted in a significantly reduced number of cases for analysis. Multiple imputation was performed and the regression analyses were performed with a larger N. However, similar overall results were obtained and, in some cases, the results were less clear. As a result, the imputed data were not used.

In an attempt to retain a larger sample, two datasets were created; the first was used for the regression analysis whereas the second was used to evaluate the hypotheses testing adverse impact. Therefore, the results from the hypothesis testing the application of TAT and moderation hypotheses cannot be linked directly to the findings addressing adverse impact.

Potential factors that influenced how the data were collected should also be taken into consideration. The data were collected over an unknown number of serials that were conducted over a period of 12 years. During this period, the ACs were conducted across Canada in numerous locations and used a continually changing assessment staff. Importantly, the permanent staff who oversee the administration, execution, data collection, input, and processing changed on several occasions. These changes are likely to have contributed to the large number of missing data and entry inconsistencies. Additionally, the data used in the current study did not allow for a more detailed

examination of inter-rater reliability because only aggregated dimension scores were entered. Further, information on the individual assessors were not recorded making it impossible to determine which individual assessors assigned scores for dimension ratings. Consideration of the effects of inter-rater reliability on dimension ratings could prove fruitful in future research. Accordingly, measures should be adapted to ensure that the amount of missing and unusable data is reduced.

Another limitation was that data on candidate experience and age were not provided. Clapham and Fulford (1997) found a negative correlation between age and AC ratings suggesting that some age bias may exist in AC ratings. This relationship was present despite the researchers having controlled for cognitive ability, education, years of service, and gender. The authors proposed that level of experience may be able to explain this relationship. In the future, information on candidate age and working experience should be collected so that potential differences on these demographic variables can be explored.

The use of self-report personality items is also an important limitation to consider, especially because for this study self-reported personality traits were often used as dependent variables. Self-report personality tests are subject to faking and in general, there is agreement that faking is strong in high-stakes scenarios (e.g., employee selection; Birkeland, Manson, Kisamore, Brannick, & Smith, 2006). All MPOAC applicants completed the TSD-PI in a high stakes situation as they were seeking employment within the CAF. Given that candidates are applying for employment in the military and in the policing field, there may be additional incentive to fake results. At this time, no published research is available addressing faking on the TSD-PI self-report measure.

Range restriction in the cognitive ability test scores is another consideration. CAF applicants are selected and screened on the basis of established scores on the cognitive ability test (i.e., the CFAT). Only applicants who achieved the prescribed cut-off score for the Military Police Officer occupation are made an offer to attend the MPOAC. The sample in this study only contains applicants who were pre-selected to have higher cognitive ability. Our ability to determine the effects of relationships between variables and relationships of interest with cognitive ability may have been influenced as a result.[1]

Scores achieved on the TSD-PI were integrated into the selection process in 2012 and 2013, meaning that the final score for applicants after this period were weighted based on Conscientiousness and Emotional Stability. Taken together, the range and variance of cognitive ability, Conscientiousness, and Emotional Stability in this study's sample would be restricted to varying extents depending on the weight assigned to cognitive ability and personality test scores.

Through determining which dimensions are elicited during individual exercises we may be able to streamline the AC using a more efficient means of assessment. For example, ratings obtained from the group dynamics II exercises were found to be negatively related to true ratings of personality and cognitive ability. Given these results, there may be no added value to using the second group dynamics exercise, which may lessen the workload.

---

[1] The Pearson correlation coefficients between dimension ratings of problem solving (which are theoretically related to cognitive ability) and cognitive ability scores ranged from $r = -.12$ to $r = .28$. Only the correlation between the problem-solving dimension measured in the Exercise D ($r = .28$, $p < .01$) was significant. These weak relationships suggest that ranged restriction likely impacted the results presented in this study.

Future Research

Past research has resulted in considerable discussion about the construct validity of ACs. Generally, the AC method is regarded as possessing criterion (i.e., predictive) validity, which is desirable in the employee selection process. However, investigating and understanding how individual ACs function and what constructs (i.e. dimensions) are being measured is also important. As highlighted by Lievens (2009), there is "little understanding and guidance on the design of exercises that efficiently elicit target behaviors" (p. 141).

Future attempts at improving validity will be facilitated by understanding the MPOAC's construct validity. However, further validation research is needed to determine if the MPOAC in its current form is a reliable and valid method for selecting entry-level military police officers. Examining criterion validity and the incremental validity of AC exercises to predict beyond other valid, less costly measures is important. Unfortunately, we did not have access to a measure of future job performance and therefore cannot provide information on criterion validity. In the future, researchers should strive to fully validate the MPOAC.

In conclusion, poor construct validity may be better explained through examination of the trait relevant situational cues presented during individual exercises. Through understanding how exercises elicit relevant traits, organizations using ACs will be able to better design exercises that activate the desired traits (i.e., dimensions). Overall, this will improve the selection process and potentially streamline the AC method, which is often highly costly for organizations. Finally, our investigation found that although male and female applicants were not rated differently on the overall

performance, Anglophone and Francophones were rated differently. Further research should be undertaken to explore to reasons for this finding, and to determine if these differences are replicated in other assessment centers.

Recommendations

This study examined the psychometric properties of an AC used in the Canadian Armed Forces. The results of this research, possible implications and future directions have been previously discussed. The following section summarizes specific recommendations for future serials of the MPOAC.

Several changes are recommended. First, reducing the number of dimensions (i.e., competencies) that assessors evaluate during each individual exercise should be considered. To improve the accuracy of recording and scoring candidate behavior the use of behavioral indicators in lieu of note taking is recommended. Finally, assessor training should be reexamined to ensure that rater biases are addressed.

Although ACs can be valid, reliable, and defensible assessment methods, other less resource intensive methods should be considered in light of the findings of this research. For example, Exercise Bs targeting the key dimensions (i.e., competencies) and situational judgement tests may be equally valid and reliable alternatives for the current multi-day AC. If the MPOAC continues to be used, the exercises in their current form should be examined for potential areas of improvement. Trait activation theory can be used for guiding changes in exercises to ensure that key traits are being strongly activated.

References

Anderson, N., Lievens, F., Dam, K., & Born, M. (2006). A construct-driven investigation of gender differences in a leadership-role assessment center. *Journal of Applied Psychology, 91*(3), 555-566. http://10.1037/0021-9010.91.3.555

Arthur, W.A., Day, E.A., Mcnelly, T.L., & Edens, P.S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*(1), 125-154. http://10.1111/j.1744-6570.2003.tb00146.x

Arthur, W.A., Day, E.A., & Woehr, D.J. (2008). Mend it, don't end it: An alternative view of assessment centre construct-related validity evidence. *Industrial and Organisational Psychology: Perspectives on Science and Practice, 1*(1), 105–111. http://10.1111/j.1754-9434.2007.00019.x

Aschwanden, D., Martin, M., & Allemand, M. (2017). Cognitive abilities and personality traits in old age across four years: More stability than change. *Journal of Research in Personality, 70*, 202-213. http://10.1016/j.jrp.2017.08.002

Barrick, M.R., & Mount, M.K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*(1), 1 - 26. http://10.1111/j.1744-6570.1991.tb00688.x

Barrick, M.R., Mount, M.K., & Judge, T.A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment*, *9*, 9–30. http://10.1111/1468-2389.00160

Binning, J.F. & Barrett, G.V. (1989). Validity of personnel decisions: A conceptual

analysis of the inferential and evidential bases. *Journal of Applied*

Psychology, 74(3), 478-494. http://10.1037/0021-9010.74.3.478

Birkeland, S.A., Manson, T.M., Kisamore, J.L., Brannick, M.T., & Smith, M.A. (2006).

A meta-analytic investigation of job applicant faking on personality measures.

International Journal of Selection and Assessment, 14, 317-335.

http://10.1111/j.1468-2389.2006.00354.x

Black, M.S. (1999). *The efficacy of personality and interest measures as a supplement to*

cognitive measures in the prediction of military training performance.

Unpublished master's thesis, Saint Mary's University, Halifax, Nova Scotia,

Canada.

Boswell, R., & Kuschnereit, F. (2009). *Canadian Forces Aptitude Test manual*

(DGMPRA CR 2009-002). Ottawa, Ontario, Canada: Director General Military

Personnel Research and Analysis.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the

multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

http://10.1037/h0046016

Christal, R.E. (1994). *Research and design summary report: Non-cognitive research*

involving systems of testing and learning. F33615-91-D-0010. Armstrong

Laboratories, Brooks AFB.

Clapham, M. & Fulford, M. (1997). Age bias in assessment center ratings. *Journal of*

Managerial Issues, 9(3), 373-387.

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

Collins, J., Schmidt, F., Sanchez–Ku, M., Thomas, L., McDaniel, M., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment, 11*(1), 17-29. http://10.111/1468-2389.00223

Costa, P.T., & McCrae, R.R. (1992). Four ways five factors are basic. *Personality and Individual Differences*, *13*, 653–665. http://10.1016/0191-8869(92)90236-I

Catano, V.M., Hackett, R.D., & Wiesner, W.H., (2016). *Recruitment and Selection in Canada* (6th ed.)*.* Toronto, ON: Nelson.

Darr, W. (2009). *Big Five Personality Research in the Military: A Meta-analysis* (TM 2009-023)*.* Ottawa, Ontario, Canada: Director General Military Personnel Research and Analysis.

Dean, M.A., Roth, P.L., & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment center ratings: A meta-analysis. *Journal of Applied Psychology, 93*(3), 685-691. http://10.1037/0021-9010.93.3.685

Dilchert, S., & Ones, D.S. (2009). Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment, 17*(3), 254-270. http://10.111/j.1468-2389.2009.00468.x

Ebel-Lam, A.P., & Carter, J. (2011). Assessing the Predictive Validity of the Canadian Forces Aptitude Test (CFAT) for the Military Police. Director General Military

Personnel Research and Analysis Letter Report 2011-5585, Ottawa, Ontario, Canada.

Eurich, T.L., Krause, D.E., Cigularov, K., & Thornton, G.C. (2009). Assessment centers: current practices in the United States. *Journal of Business and Psychology, 24*(4), 387-407. http://10.1007/s10869-009-9123-3

Ferris, G., Witt, L., & Hochwarter, W. (2001). Interaction of social skill and general mental ability on job performance and salary. *Journal of Applied Psychology, 86*(6), 1075-1082. http://10.1037/0021-9010.86.6.1075

Goldstein, H.W., Yusko, K.P., Braverman, E., Smith, D., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment center exercises. *Personnel Psychology, 51*(2), 357-374. http://10.1111/j.1744-6570.1998.tb00729.x

Goldstein, H.W., Yusko, K.P. & Nicolopoulos, V. (2001). Exploring Black-White subgroup differences of managerial competencies. *Personnel Psychology, 54,* 783-807. http://10.111/j.1744-6570.2001.tb00232.x

Haaland, S., & Christiansen, N. (2002). Implications of trait-activation theory for evaluating the construct validity of assessment center ratings. *Personnel Psychology, 55*(1), 137-163. http://10.1111/j.1744-6570.2002.tb00106.x

Hodgson, K.E. (2005). *Military police selection in Canada: An evaluation of the Canadian Force Military Police Assessment Center (MPAC)*. Saint Mary's University, Halifax, Canada.

Hoffman, B., Kennedy, C., LoPilato, A., Monahan, E., & Lance, C. (2015). A review of the content, criterion-related, and construct-related validity of assessment center

exercises. *Journal of Applied Psychology, 100*(4), 1143-1168.

http://10.1037/a0038707

Hough, L., Oswald, F., & Ployhart, R. (2001). Determinants, detection and amelioration

of adverse impact in personnel selection procedures: Issues, evidence and lessons

learned. *International Journal of Selection and Assessment, 9*(1-2), 152-194.

http://10.1111/1468-2389.00171

Howard, A. (2008). Making assessment centers work the way they're supposed to.

Industrial and Organizational Psychology: Perspectives on Science and Practice,

1, 98–104. http://10.111/j.1754-9434.2007.00018.x

Hunter, J.E., & Hunter, R.F. (1984). Validity and utility of alternative predictors of job

performance. *Psychological Bulletin, 96*(1), 72-98. http://10.1037/0033-

2909.96.1.72

Iles, P. (1992). Centres of excellence? Assessment and development centres, managerial

competencies and human resource strategies. British Journal of Management,

3(2), 79–90. http://dx.doi.org/10.1111/j.1467-8551.1992. tb00037.x

Jackson, D.J., Barney, A., Stillman, J., & Kirkley, W. (2007). When traits are behaviors:

The relationship between behavioral responses and trait-based overall assessment

center ratings. *Human Performance, 20*(4), 415-432.

http://10.1080/08959280701522130

Jackson, D.J., Michaelides, G., Dewberry, C., & Kim, Y. (2016). Everything that you

have ever been told about assessment center ratings is confounded. *Journal of

Applied Psychology, 101*(7), 976-994. http://10.1037/apl0000102

Jansen, P.G., & Stoop, B.A. (2001). The dynamics of assessment center validity: Results

of a 7-year study. *Journal of Applied Psychology, 86*(4), 741-753.

http://1037/0021-9010.86.4.741

Kolk, N.J., Born, M.P., Van Der Flier, H., & Olman, J.M. (2002). Assessment center

procedures: Cognitive load during the observation phase. *International Journal of

Selection and Assessment, 10*(4), 271-278. http://10.111/1468-2389.00217

Lance, C.E. (2008). Why assessment centers do not work the way they're suppose to.

Industrial and Organizational Psychology, 1, 87-100. http://10.111/j.1754-

9434.2007.00017.x

Lance, C.E., Foster, M.R., Gentry, W.A., & Thoresen, J.D. (2004). Assessor cognitive

processes in an operational assessment center. *Journal of Applied

Psychology, 89*(1), 22-35. http://10.1037/0021-9010.89.1.22

Lance, C.E., Lambert, T.A., Gewin, A.G., Lievens, F. & Conway, J.M. (2004). Revised

estimates of dimension and exercise variance components in assessment center

postexercise dimension ratings. *Journal of Applied Psychology, 89*(2), 377-385.

http://10.1037/0021-9010.89.2.377

Le, H., Oh, I., Robbins, S., Ilies, R., Holland, E., & Westrick, P. (2011). Too much of a

good thing: Curvilinear relationships between personality traits and job

performance. *Journal of Applied Psychology, 96*(1), 113-133.

http://10.1037/a0021016

Lievens, F. (2001a). Assessor training strategies and their effects on accuracy, interrater

reliability, and discriminant validity. *Journal of Applied Psychology. 86,* 255-264.

http://10.1037/0021-9010.86.2.255

Lievens, F. (2001b). Assessors and the use of assessment center dimensions: A fresh look at a troubling issue. *Journal of Organizational Behavior, 22,* 203-221. http://10.1002/job.65

Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centers: An examination of assessor and assessee effects. *Journal of Applied Psychology, 87*(4), 675-686. http://10.1037/0021-90210.87.4.675

Lievens, F. (2009). Assessment centers: A tale about dimensions, exercises, and dancing bears. *European Journal of Work and Organizational Psychology, 18*(1), 102-121. http://10.1080/13594320802058997

Lievens, F., Chasteen, C.S., Day, E.A., & Christiansen, N.D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology, 91*(2), 247-258. http://10.1037/0021-9010.91.2.247

Lievens, F. & Conway, J. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait–multimethod studies. Journal of Applied Psychology, 86(6), 1202-1222. http://10.1037/0021-9010.91.3.555

Lievens, F., & De Fruyt, F. (2001). Assessors' use of personality traits in descriptions of assessment centre candidates: A five-factor model perspective. *Journal of Occupational and Organizational Psychology, 74*, 623-636.

Lievens, F., de Koster, L., & Schollaert, E. (2012). Current theory and practice of assessment centers: The importance of trait activation. In S. Cartwright & C.L.

Cooper (Eds.), *The Oxford handbook of personnel psychology* (pp. 215-233).

Oxford; New York: Oxford University Press.

Lievens, F., & Schollaert, E. (2011). Adjusting exercise design in assessment centers:

Theory, practice, and research. In N. Povah, & G.C. Thornton (Eds.), *Assessment*

centres and global talent management (pp. 47–60). Surrey, UK: Gower

Publishing Limited.

Lievens, F., Tett, R.P., & Schleicher, D.J. (2009). Assessment centers at the

crossroads: Toward a reconceptualization of assessment center exercises.

Research in Personnel and Human Resources Management, 28, 99-152.

Lowery, C., Beadles, N., & Krilowicz, T. (2004). Using personality and cognitive ability

to predict job performance: An empirical study. *International Journal of*

Management, 21(3), 300-306.

McCrae, R.R., & Costa, P.T. (1987). Validation of the 5-factor model of personality

across instruments and observers. Journal of Personality and Social Psychology,

52, 81–90. http://10.1037/0022-3514.52.1.81

Meriac, J.P., Hoffman, B.J., Woehr, D.J., & Fleisher, M.S. (2008). Further evidence for

the validity of assessment center dimensions: A meta-analysis of the incremental

criterion-related validity of dimension ratings. *Journal of Applied*

Psychology, 93(5), 1042-1052. http://10.1037/0021-9010.93.5.1042

Meriac, J.P., Hoffman, B.J., & Woehr, D.J. (2014). A conceptual and empirical review of

the structure of assessment center dimensions. *Journal of Management, 40*(5),

1269-1296. http://10.1177/0149206314522299

Mount, M., Barrick, M., & Strauss, J. (1999). The joint relationship of conscientiousness

    and ability with performance: Test of the interaction hypothesis. *Journal of*

    Management, 25(5), 707-721. http://10.1177/014920639902500505

Mount, M.K., Oh, I., & Burns, M. (2008). Incremental validity of perceptual speed and

    accuracy over general mental ability. Personnel Psychology, 61, 113–139.

    http://10.1111/j.1744-6570.2008.00107.x

Murphy, K.R., & Cleveland, J.N. (1995). *Understanding performance appraisal: Social,*

    organizational, and goal-based perspectives. Thousand Oaks, CA: Sage

    Publications.

Neisser, U., Boodoo, G., Bouchard, T.J., Boykin, A. W, Brody, N., Ceci, S.J…Perloff, R.

    (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*(2), 77-

    101.

Ng, E.S., & Sears, G.J. (2010). The effect of adverse impact in selection practices on

    organizational diversity: A field study. *The International Journal of Human*

    Resource Management, 21(9), 1454-1471. http://10.1080/09585192.2010.488448

Oh, I., Wang, G., & Mount, M.K. (2011). Validity of observer ratings of the five-factor

    model of personality traits: A meta-analysis. *Journal of Applied*

    Psychology, 96(4), 762-773. http://10.1037/a0021832

Oliver, T., Hausdorf, P., Lievens, F., & Conlon, P. (2016). Interpersonal dynamics in

    assessment center exercises. *Journal of Management, 42*(7), 1992-2017.

    http://10.1177/0149206314525207

Ones, D.S., Viswesvaran, C., & Dilchert, S. (2005). Cognitive ability in personnel

    selection decisions. In A. Evers, O. Voskuijl, & N. Anderson (Eds.), *Handbook of*

    personnel selection (pp. 143-173). Oxford, UK: Blackwell.

Perry, S., Hunter, E., Witt, L., & Harris, K. (2010). P = f (Conscientiousness × Ability):

    Examining the Facets of Conscientiousness. *Human Performance, 23*(4), 343-

    360. http://10.1080/089285.2010.501045

Petrides, K., Weinstein, Y., Chou, J., Furnham, A., & Swami, V. (2010). An investigation

    into assessment centre validity, fairness, and selection drivers. *Australian Journal*

    of Psychology, 62(4), 227-235. http://10.1080/00049531003667380

Ployhart, R., & Holtz, B. (2008). The diversity-validity dilemma: Strategies for reducing

    racioethnic and sex subgroup differences and adverse impact in selection.

    Personnel Psychology, 61(1), 153-172. http://10.111/j.1744-6570.2008.00109.x

Pyburn, K., Ployhart, R., & Kravitz, D. (2008). The diversity-validity dilemma: Overview

    and legal context. *Personnel Psychology, 61*(1), 143-151. http://10.1111/j.1744-

    6570.2008.00108.x

R Core Team (2017). R: A language and environment for statistical computing. R

    Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-

    project.org/.

Rothmann, S., & Coetzer, E.P. (2003). The big five personality dimensions and job

    performance. *SA Journal of Industrial Psychology, 29*(1), 68-74.

    http://10.4102/sajip.v29i1.88

Rupp, E.D., Hoffman, B.J., Bischof, D., Byham, W., Collins, L., Gibbons, A., ...

    Thornton, G. (2015). Guidelines and Ethical Considerations for Assessment

Center Operations. *Journal of Management, 41*(4), 1244-1273.

http://10.1177/0149206314567780

Sackett, P.R., & Dreher, G.F. (1982). Constructs and assessment center dimensions:

some troubling empirical findings. *Journal of Applied Psychology, 67*, 401–410.

http://10.1037/0021-9010.67.4.401

Salgado, J.F. (2003). Predicting job performance using FFM and non-FFM personality

measures. *Journal of Occupational and Organizational Psychology, 76*(3), 323-

346. http://10.1348/096317903769647201

Schleicher, D.J., Day, D.V., Mayes, B.T., & Riggio, R.E. (2002). A new frame for frame-

of-reference training: Enhancing the construct validity of assessment

centers. *Journal of Applied Psychology, 87*(4), 735-746. http://10.1037/0021-

9010.87.4.735

Schmidt, F.L., & Hunter, J.E. (2004). General mental ability in the world of work:

Occupational attainment and job performance. *Journal of Personality and Social

Psychology, 86,* 162-173. http://10.10.1037/0022-3514.86.1.162

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in

personnel psychology: practical and theoretical implications of 85 years of

research findings. *Psychological Bulletin, 124*(2), 262-274. http://10.1037/0033-

2909.124.2.262

Schneider, J., & Schmitt, N. (1992). An exercise design approach to understanding

assessment center dimension and exercise constructs. *Journal of Applied

Psychology, 77*(1), 32-41. http://10.1037/0021-9010.77.1.32

Shore, T.H., Thornton, G.C., & Shore, L. (1990). Construct validity of two categories of

    assessment center dimension ratings. *Personnel Psychology, 43*(1), 101-116.

    http://10.1111/j.1744-6570.1990.tb02008.x

Society for Industrial and Organizational Psychology. (2003). *Principles for the*

    validation and use of personnel selection procedures (4th Ed.).   Retrieved from

    http://www.siop.org/_principles/principlesdefault.aspx

Spector, P., Schneider, J., Vance, C., & Hezlett, S. (2000). The relation of cognitive

    ability and personality traits to assessment center performance. *Journal of Applied*

    Social Psychology,30(7), 1474-1491. http://1-.1111/j.1559-1816.2000.tb02531.x

Speer, A.B., Christiansen, N.D., Goffin, R.D., & Goff, M. (2014). Situational bandwidth

    and the criterion-related validity of assessment center ratings: Is cross-exercise

    convergence always desirable? Journal of Applied Psychology, 99(2), 282-295.

    http://10.1037/a0035213

Tett, R.P., & Burnett, D.N. (2003). A personality trait-based interactionist model of job

    performance. *Journal of Applied Psychology, 88*, 500-517. http://10.1037/0021-

    9010.88.3.500

Tett, R.P., Jackson, D.N., & Rothstein, M. (1991). Personality measures as predictors of

    job performance: A meta-analytic review. *Personnel Psychology, 44*(4), 703-742.

    http://10.111/j.1744-6570.1991.tb00696.x

Thornton, G.C., & Gibbons, A.M. (2009). Validity of assessment centers for personnel

    selection. *Human Resource Management Review, 19*(3), 169-187. http://

    10.1016/j.hrmr.2009.02.002

van Hooft, E.A. & Born, M.P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, *97*(2), 301-316. http://10/1037/a0025711

Wirz, A., Melchers, K., Schultheiss, S., & Kleinmann, M. (2014). Are improvements in assessment center construct-related validity paralleled by improvements in criterion-related validity? *Journal of Personnel Psychology, 13*(4), 184-193. http://10.1027/1866-5888/a000115

Women in the Canadian Armed Forces. (2018, March). Retrieved from http://www.forces.gc.ca/en/news/article.page?doc=women-in-the-canadian-armed-forces/izkjqzeu

Wright, P., Kacmar, K., McMahan, G., & Deleeuw, K. (1995). P=f(M X A): Cognitive Ability as a Moderator of the Relationship Between Personality and Job Performance. *Journal of Management, 21*(6), 1129-1139. http://10.1177/0149