

TSD-PI: ESTABLISHING CUT-OFF SCORES

Personality as a Predictor of Performance in the Canadian Armed Forces:

Establishing Cut-Off Scores

by

Carrie N. Topping

A thesis submitted to
Saint Mary's University, Halifax, Nova Scotia
in Partial Fulfillment of the Requirements for
the Degree of Masters of Science in Applied Psychology
with Specialization in Industrial/Organizational Psychology

December, 2017, Halifax, Nova Scotia

© Carrie Topping, 2017

Approved: Dr. Vic Catano
Faculty Supervisor

Approved: Dr. Wendy Darr
Committee Member

Approved: Dr. Kevin Kelloway
Committee Member

Approved: Dr. Stephanie Gilbert
External Examiner

Date: December 11th, 2017

Table of Contents

Table of Contents	ii
List of Tables	iv
List of Figures	vi
Abstract	vii
Introduction	1
Purpose	1
The Five Factor Model of Personality	2
The Five Factor Model and Performance Outcome Criteria	3
Trait Activation Theory in Support of Personality as a Selection Tool	4
Personality as a Predictor in the Military Environment	8
The Trait Self Descriptive Personality Inventory (TSD-PI)	9
Psychometric Properties of the TSD-PI	10
Trait Activation Theory: Justifying the use of Conscientiousness and Emotional Stability	13
Need for Cut-Off Scores	14
Methods of Establishing Cut-Off Scores	17
Current Research	20
Method	21
Participants	21
Measures	23
Trait Self Descriptive Personality Inventory (TSD-PI)	23
Training Outcomes	24
Job Performance	25
Procedure	27
Results	31

TSD-PI: ESTABLISHING CUT-OFF SCORES	iii
Missing Values Analysis and Data Preparation	31
Assumptions	32
Initial Analyses.....	32
Validity Across NCM Versus Officer Groups.....	35
Validity Across Occupational Families	40
Establishing Cut-Off Scores.....	46
Forward Regression Analyses	46
Recommending Criterion-Related Cut-Off Scores.....	49
Recommending Norm-Referenced Cut-Off Scores.....	53
Summary of Recommended Cut-Off Scores	54
Discussion.....	56
Limitations	60
Measures	60
Samples.....	64
Implications.....	66
Future Research Directions	68
Conclusion	71
References.....	73

List of Tables

Table 1. Job Demands at the Task, Social and Organizational Levels for Conscientiousness and Emotional Stability.....	14
Table 2. Dataset Demographic Information.	22
Table 2b. Cleansed Dataset Demographic Information – English Language Participants Only.....	23
Table 3. Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Training Outcome Variables	34
Table 4. Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Performance Outcome Variables	35
Table 5. Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Training Outcome Variables for NCMs and Officers	37
Table 6. Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Performance Outcome Variables for NCMs and Officers	38
Table 7. Cohen’s <i>d</i> Mean Comparisons Showing the Differences Between NCM and Officer Groups for Predictor and Outcome Variables	39
Table 8. Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Performance Outcome Variables for Occupational Family Groups (Incumbent Sample)	43
Table 9. Summary of Z-Test Intercorrelation Comparisons for Overall NCM and Occupational Family Groups for Conscientiousness and Emotional Stability as predictors of Job Performance (Incumbent Sample)	44

Table 10. Criterion Variables Included in Subsequent Analyses to Recommend Cut-Off Scores for Officers, NCMs and Occupational Family Groups.....	45
Table 11. Regression Analyses – Conscientiousness and Emotional Stability as Predictors of Job Performance for NCMs and Officers (Incumbent Sample).....	46
Table 12. Regression Analyses – Conscientiousness and Emotional Stability as Predictors of Job Performance for Occupational Families (Incumbent Sample)	48
Table 13. Conscientiousness Cut-Off Scores Based on Regression Analyses of Performance Outcome Criteria (Incumbent Sample).....	51
Table 14. Emotional Stability Cut-Off Scores Based on Regression Analyses of Performance Outcome Criteria (Incumbent Sample).....	52
Table 15. Norm-Referenced Cut-Off Scores Based on a Minimum Level of Acceptable Job Performance (Incumbent Sample)	54
Table 16. Summary of Recommended Cut-Off Scores for the TSD-PI.....	55

List of Figures

Figure 1. Tett and Burnett's (2003) Personality Trait Based Model of Job Performance ..7

Figure 2. Breakdown of Datasets into Samples for Analyses30

Personality as a Predictor of Performance in the Canadian Armed Forces:
Establishing Cut-Off Scores

by

Carrie N. Topping
December 11th, 2017

Abstract

The Trait Self Descriptive Personality Inventory (TSD-PI) has been administered to applicants of the Canadian Armed Forces (CAF) as a selection measure since 2013. The purpose of this study was to present the theoretical and empirical justification for the establishment of cut-off scores for the conscientiousness and emotional stability subscales of the measure, used as predictors of training and job performance outcome criteria. Predictor variable data were collected from 7252 serving members with training results and 1232 with job performance information. Regression results revealed no significant relationships between the personality subscales and training results; however, performance data suggested the value of establishing distinct cut-off scores for officers, non-commissioned members (NCMs) and NCM family groups, categories comprised of occupations sharing common sets of abilities or competencies. Criterion-related and norm-referenced techniques provided recommendations for the establishment of baseline scores. Study limitations, implications and future research directions are discussed.

Keywords: personality, selection, criterion-related, norm-referenced, cut-off score, trait activation theory

Personality as a Predictor of Performance in the Canadian Armed Forces:
Establishing Cut-Off Scores

Purpose

The purpose of this study was to establish cut-off scores for the conscientiousness and emotional stability subscales of the Trait Self Descriptive Personality Inventory (TSD-PI). The TSD-PI is a self-report measure of the Five Factor Model (FFM) of personality that is administered to all applicants to the Canadian Armed Forces (CAF) who meet the basic age, citizenship and education entrance eligibility criteria. The TSD-PI is administered in conjunction with the Canadian Forces Aptitude Test (CFAT), a measure of cognitive ability. Minimum CFAT cut-off scores exist for the officer and non-commissioned member (NCM) populations. More stringent CFAT cut-off scores have been established for certain occupations, which must be met prior to progressing to the interview phase of selection. Scores from the CFAT, TSD-PI, and structured interview are combined in a compensatory model to form an applicant's competitiveness score, with which they are compared to other applicants. The three measures are weighted according to their predictive validity. The CFAT provides 60% of the combined score, the TSD-PI is granted 15% and the interview accounts for 25%.

Currently, TSD-PI results alone do not result in selection decisions to turn down applicants for employment; however, recently, data became available to allow for the establishment of cut-off scores based on desired outcome criteria: training and job performance. This paper will briefly review the FFM as a broad theory of personality and establish the importance of personality as a selection tool using Trait Activation

Theory. Subsequently, focus will shift to research relating to personality as a predictor of performance in the CAF and the requirement for cut-off scores on the TSD-PI.

The Five Factor Model of Personality

Personality, the qualities or characteristics that motivate an individual (Hogan & Roberts, 2001), is a thriving area of psychological research. Of particular interest is its value as a selection tool within organizational settings due to its ability to predict desirable workplace behaviours. For over a century, psychologists have studied personality with the goal of predicting job performance. Early researchers were plagued with issues concerning inconsistent terminology and the examination of an extensive number of facets, which made the comparison of different study findings arduous and, ultimately, resulted in the observed links between personality and job performance being regarded as trivial (Barrick, Mount & Judge, 2001). Little progress was observed in this area until the 1990's, which saw the introduction of the Five Factor Model (FFM) of personality as the predominant theory (Hogan & Roberts, 2001). The FFM successfully focused researchers on a classification system that grouped the numerous facets of personality into five broad categories; conscientiousness, agreeableness, openness to experience, extraversion and emotional stability. Correlations between these five dimensions and job performance were supported, leading psychologists to become optimistic about the value of using personality as a method of personnel selection (Barrick et al., 2001).

In general, despite minor disagreements regarding the assignment of some lower-level components, most research has supported the five main categories identified in the FFM. Following their evaluation of the findings of 15 meta-analyses, Barrick et al.

(2001) summarized the organization of each of the factors. The researchers listed excitement-seeking, sociability, ambition and positive emotionality as characteristics typical of extraversion. Cooperation, trustfulness, and affability were classified under the agreeableness factor. The absence of feeling hostile, anxious, insecure or depressed are descriptors of emotional stability. People high in conscientiousness tend to be dependable, high achievers and planners. Finally, openness to experience was defined by the researchers as unconventionality, creativity and broad-mindedness (Barrick et al., 2001).

The FFM and Performance Outcome Criteria

Since the establishment of the FFM as the leading theory for the classification of personality, research has recognized links to areas of job and training performance. In addition, support for the relationship between personality and performance has been corroborated through numerous meta-analytic evaluations (e.g., Anderson & Viswesvaran, 1998; Barrick & Mount, 1991; Goldberg, 1993; Hough, 1992; Salgado, 1997; Salgado, 1998; Tett, Jackson & Rothstein, 1991).

Where some of the factors relate to overall job performance, irrespective of the task, others have been linked to job areas or specific aspects of overall performance. A large body of research has found conscientiousness and emotional stability to be significant predictors of job performance, regardless of the occupational group (defined by similar job tasks) that is being studied (Barrick et al., 2001; Barrick & Ryan, 2003; Salgado, 2002; Tett et al., 1991). Conscientiousness has also been linked to effectiveness in leadership roles (Drasgow, Stark, Chernyshenko, Nye, Hulin & White, 2012) and there is evidence of a relationship between emotional stability and management performance

(Rothmann & Coetzer, 2003). The remaining three factors have been correlated with specific work-related behaviours. For example, agreeableness is positively associated with customer service jobs that require human interaction, especially positions that necessitate high levels of cooperation or a nurturing role (Barrick et al., 2001; Klang, 2012; Mount, Barrick & Stuart, 1998). Openness to experience has been discovered as a consistent predictor of training performance, irrespective of the desired training objective (Barrick & Mount, 1991; Salgado, 1997; Salgado, 2002). Finally, extraversion has been linked to performance in leadership/managerial positions (Barrick et al., 2001; Mount et al., 1998; Salgado, 2002) and training proficiency (Barrick & Mount, 1991; Hough, 1992). As such, research on the FFM has provided a legally defensible theory that contributes personality as a valuable predictor to strengthen the selection models of organizations who seek high performing employees.

Trait Activation Theory in Support of Personality as a Selection Tool

Assessing personality during the selection process (using measures based on the FFM) allows for the establishment of both performance-based and satisfaction-based person-job fit (Tett, Simonet, Walser & Brown, 2013). Where other measures evaluate an applicant's knowledge, skills and abilities as they relate to the tasks listed in the job analysis, personality measures serve to determine how an individual will respond to those job demands (i.e., perform when faced with situational cues) as well as how closely an individual's personality traits match those requirements and therefore meet the psychological needs of the individual. Fit with respect to meeting situational demands is related to performance and fit regarding the fulfillment of an individual's psychological needs is related to levels of job satisfaction (Tett et al., 2013). This dual relationship,

unique to personality, can be explained using Trait Activation Theory, a person-situation interactionist model of job performance (TAT; Tett & Burnett, 2003).

Tett and Burnett's (2003) model builds upon the work of earlier person-situation interactionist researchers who suggested that the expression of latent personality traits is situation dependent (Bowers, 1973; Ekehammar, 1974; Endler & Magnusson, 1976; Epstein & O'Brien, 1985; Pervin, 1985; Snyder & Ickes, 1985; Weiss & Adler, 1984). This theme was extended to apply to the work setting by Schneider's (1983, 1987) attraction-selection-attrition (ASA) model and Chatman's (1989) model of person-organization fit. Both models suggest that good fit occurs when there is congruence between the values of the individual and the organization. According to the ASA framework, individuals leave when fit is poor. Chatman (1989) suggests that the benefits of good person-organization fit include extra-role behaviours, value change and extended tenure. TAT's contribution is to provide direction on the moderating variables at play on the personality and job performance relationship (Tett & Burnett, 2003).

TAT posits three main principles. First, that personality traits are an individual's underlying tendencies to behave in a specific manner. Second, traits are expressed by how an individual behaves in response to trait-relevant situational cues (e.g., volunteering to assist a colleague in response to a request for help with a work task allows an individual to show high levels of agreeableness). Finally, the individual gains intrinsic satisfaction from the expression of their personality traits (Tett et al., 2013). Situational features allowing for trait expression are found desirable, regardless of extrinsic rewards that may be offered (Tett & Burnett, 2003). Applied to a work setting, TAT (Tett & Burnett, 2003) suggests that the relationship between personality traits and behaviour

(trait expression) is moderated by multiple variables, including situational cues at three different levels (task, social and organizational) and intrinsic and extrinsic rewards. In order for a trait to be activated, resulting in trait expression, situation cues must be thematically connected to the personality trait in question. Those strong in that particular trait will react differently than those at the opposite end of the trait spectrum (e.g., extroverts tend to take charge if presented with a leaderless group whereas introverts are likely to follow an emerging leader). An exception to this is when individual differences in responses are negated by the strength of the situation or when extrinsic rewards are severe (e.g., an individual low in conscientiousness may find it acceptable to show up late to work, however tardiness is not an option when exiting a burning building). In this case, behaviour variability will be minimal and less dependent on individual personality traits. These relationships are presented in Tett and Burnett's (2003) personality trait-based model of job performance, which is displayed in Figure 1.

The first path in the model represents the main effect: the basic assumption that an individual's trait level, signified by their score on a personality measure, is expressed by trait relevant behaviour on the job (Tett & Burnett, 2003). Path 2 is the direct effect of a strong situational cue on an individual's work behaviour, when unique personality traits are less likely to vary responses. Paths 3, 4, and 5 represent the moderating effect of task (i.e., those listed in the job analysis), social (e.g., interpersonal relationships between coworkers) and organizational (e.g., culture and climate of the organization) trait-relevant cues in the workplace. Job performance (path 6) provides a distinction between positive and negative work behaviour, dependent on the context, which is determined by the evaluation (path 7) of work behaviour based on the expectations set by the situational

demands. Paths 8 and 9 represent two motivators of personality expression at work, the intrinsic reward experienced from trait expression and extrinsic rewards conceded as a result of valued work behaviours. The receipt of rewards motivates future behaviours (path 10), which can occur without the influence of unique personality traits if extrinsic rewards are strong enough to overpower individual differences. Finally, path 11 represents the impact that individuals have on their environment (Tett & Burnett, 2003).

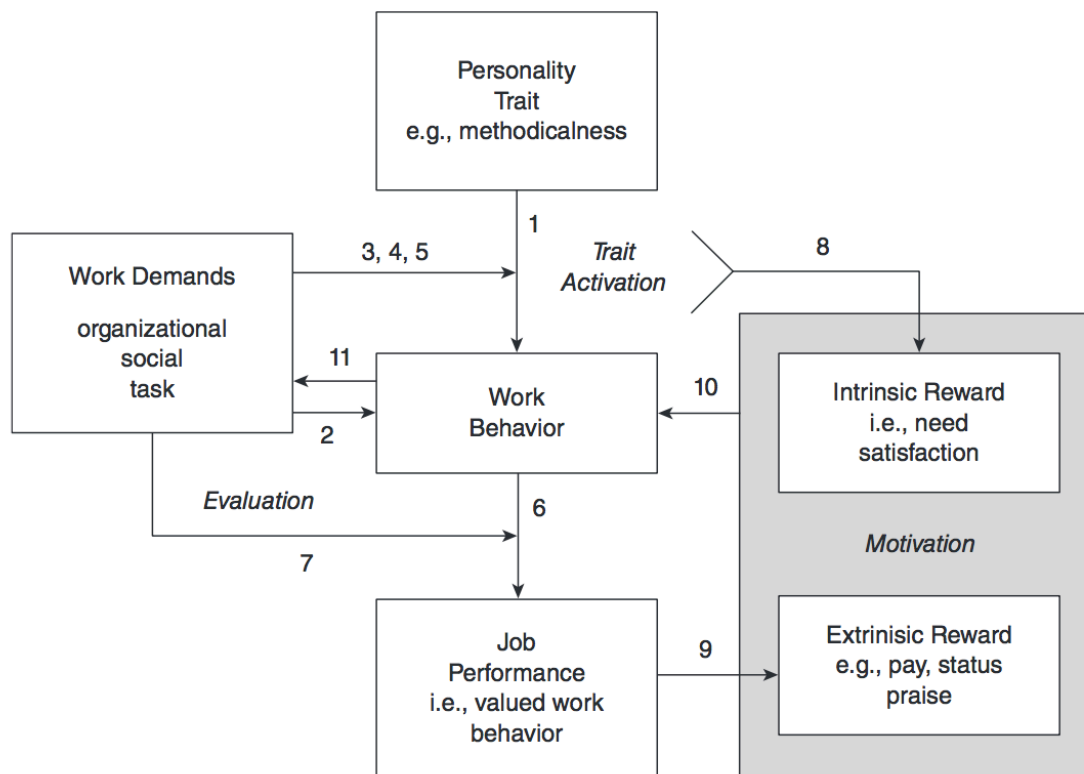


Figure 1 Tett and Burnett's (2003) Personality Trait-Based Model of Job Performance.

TAT suggests that the personality traits measured for the purpose of selection should be those that elicit desired behaviours according to the trait, social and organizational demands of the workplace. As discussed in later sections of this paper, the

CAF's selection of conscientiousness and emotional stability as the factors against which applicants are compared shows value congruence on multiple levels.

Personality as a Predictor in the Military Environment

The CAF is an organization that expends extensive assets to select the most appropriate personnel for employment in positions that are most suited to the individual. Considerable ramifications may result from not doing so (e.g., an emotionally unstable officer responsible for the safety and well-being of 250 sailors at sea, or a nurse who does not work well with others). To date, the vast majority of personality research as it pertains to performance has been conducted using samples from civilian populations. As such, there has been some uncertainty concerning the generalizability of the results to the military population (Tett, Jackson, & Rothstein, 1991). In a meta-analysis that analyzed the personality-outcome relationships of 20 independent military samples, collected from a wide range of occupational groupings, ranks and nationalities and who had been administered variants of the TSD-PI, Darr (2011) sought to address this concern. She acknowledged that, theoretically, the structured environment provided by the military setting could mitigate the effects of personality on an individual's behaviours due to strong organizational norms, regulations and consequences for undesired actions, as depicted in paths 2 and 10 in Tett and Burnett's (2003) Personality Trait-Based Model of Job Performance (Figure 1). However, Darr's (2011) findings were contrary to the hypothesis, in that the majority of the measured personality-outcome effects were either equally robust, or even stronger than those observed in the civilian samples. These results have important implications as selection testing of civilian applicants is conducted prior

to exposure to the military environment. If the results were not generalizable, it would have invalidated the use of personality as a predictor of performance in the CAF.

The Trait Self Descriptive Personality Inventory

The Self Descriptive Inventory (SDI), developed by the United States (U.S.) Air Force, was the first measurement tool to assess the FFM within the military setting (Christal, 1994). The SDI consists of a 163 item inventory that, through The Technical Cooperation Program, was made available to several military organizations and subsequently adapted for research purposes. Canada's version of the SDI, the Trait Self Descriptive Personality Inventory (TSD-PI), saw the reduction of the original 163 items to 75 (Boyes, 2006). The scale was also changed to a seven point Likert-type scale, where individuals are asked to identify the extent to which an adjective or statement is characteristic of them; one being extremely uncharacteristic and seven extremely characteristic (Boyes, 2005). Finally, the TSD-PI was translated into the French Canadian language and psychometric properties were improved through some item adaptations (Darr & Bergevin, 2009). The revised personality measurement was introduced into the CAF recruiting process in 2009 for research purposes (Sutherland, 2012) and was successfully implemented into the CAF selection system in January of 2013 (Darr, 2013). Currently, only the conscientiousness and emotional stability subscales of the measures are being considered in combination with other constructs (i.e. cognitive ability and person-job fit) to guide selection decisions, although all five subscales of the TSD-PI (including extraversion, openness to experience and agreeableness) are administered to applicants (Darr, 2016). Raw scores are converted into percentiles using normative data for the appropriate population (i.e., English or

French, depending on the language that the inventory was written in). The normative data were updated based on the most recent applicant data in 2015 and the percentile rank norm tables were made available in the TSD-PI Manual (Darr, 2016).

Psychometric Properties of the TSD-PI

The psychometric properties of the TSD-PI have been studied for over a decade (Darr, 2016). This summary is provided to substantiate the usefulness of the the inventory for selection purposes and focusses on the two factors currently being employed within the CAF, conscientiousness and emotional stability.

Reliability of the inventory is supported through measurements of internal consistency and test-retest reliability. Coefficient alpha values for conscientiousness and emotional stability were consistent between two samples at $\alpha = .90$ for both the French and English language inventories (Darr & Kemp, 2011; Darr, 2016). These values are considered excellent (US Department of Labor and O*NET, 2000). Examinations of test-retest reliability are reported in the TSD-PI Manual (Darr, 2016). Correlations between the first and second writing sessions (averaging 13.5 months between attempts) were .62 for both the conscientiousness and emotional stability subscales in the English language sample. In the French language sample (averaging 9.2 months between attempts), correlations were .76 for emotional stability and .77 for conscientiousness (Darr, 2016). These figures are consistent with the ranges reported for other commercial personality measures, such as the NEO Personality Inventory, (.63 -.81; Costa & McCrae, 1994) and the Sixteen Personality Factor Questionnaire (.53-.87; Cattell & Mead, 2008).

The Standard Error of Measurement (SEM) takes into consideration the variability in test scores that is accounted for by potential error (i.e., caused by testing conditions,

motivations of the test taker or factors relating to the measure itself; Harvill, 1991). SEM values for the TSD-PI as reported in the TSD-PI Manual account for errors that stem from the measure's structure and time of administration (Darr, 2016). SEM values for the conscientiousness subscale are 4.80 for the English language sample and 3.74 for the French language sample. The emotional stability values are 5.66 for the English language and 4.43 for the French language measures (Darr, 2016). Although not currently used to aid in interpretation of individual scores, SEM values should be applied to recommended cut-off scores, therefore they are relevant to this study.

Examinations of construct validity through factor analyses (e.g., Boyes, 2005; O'Keefe, 1998) have supported the FFM structure that the TSD-PI is intended to measure (Darr, 2016). Boyes' (2005), study of TSD-PI convergent validity found strong correlations with Costa and McCrae's (1994) NEO-PI-R (ranging from .64 to .82) and Lee and Ashton's (2004) HEXACO (.43 to .79). In tandem, the stable internal structure and convergent validity demonstrated by the TSD-PI provide substantial support for the construct validity of the measure (Darr, 2016)

Criterion validity of the TSD-PI, was established by Darr's (2011) meta-analyses of the inventory and its variants, including many CAF samples. Her findings for conscientiousness and neuroticism (the reverse of emotional stability) were as strong or stronger than the estimates reported in other meta-analyses (e.g., Barrick, Mount, & Judge, 2001). For neuroticism, corrected estimates (r - corrected for criterion and predictor unreliability) were -.22 for predicting job performance and -.13 for training performance. The corrected conscientiousness subscale estimates were .35 for job performance and .18 for training performance (Darr, 2011).

The TSD-PI was also examined for equivalence across groups. Early studies (Darr & Bergevin, 2009; Kemp, 2010) resulted in the revision of items (Darr & Kemp, 2011) to improve equivalence between the English and French language versions of the measure. Subsequent differential item-functioning analyses found that the largest differential effect (an openness to experience item) caused a variance of only .36 points out of 105 possible points on the measure when scaled back into raw scores (Darr & Kemp, 2011). Group comparisons for gender, visible minority and Aboriginal groups as well as officer and NCM populations found only small differences (Darr & Kemp, 2011). Officer applicants tended to score slightly higher on conscientiousness than NCM applicants (English: -.14; French: -.12). Empirical evidence links this factor to educational criteria (i.e., ability to learn, grade point average and academic achievement; Chamorro-Premuzic & Furnham, 2003; Paunonen & Ashton, 2001; Major, Turner, & Fletcher, 2006). Therefore, the slightly higher officer scores may be a function of the CAF requirement that all officers possess a university degree, which is not a requirement for the NCM population (Darr, 2016).

Overall, the extensive research into the psychometric properties of the TSD-PI has provided evidence for a reliable and valid measure that is consistent with other well-known measures of personality. Group comparisons indicate strong linguistic equivalence and only small differences across evaluated designated groups (i.e. gender, visible minorities and Aboriginal persons).

Trait Activation Theory: Justifying the use of Conscientiousness and Emotional Stability

TAT explains the conclusion that conscientiousness is a universal predictor of job performance (Barrick and Mount, 1991) because all jobs provide situational cues for achievement and dependability. However, the unique environment of the military provides additional justification due to the potential life or death implications if trait specific behaviours are not displayed in operational settings. Task level traits such as precision, high quality task completion and rule-following are essential requirements for the day to day tasks of all occupations while deployed. At the social level, military personnel depend on one another for their safety and rely upon precise and explicit communications for the passage of orders. Finally, the organizational level demands success and loyalty.

Sundstrom (1999) classified military units as action or performing work groups due to the highly specialized roles of their personnel and the fact that the circumstances that they face are rapidly changing and must be reacted to promptly. Tett and Burnett (2003) suggested that emotional stability would elicit favorable performance on the job in this type of team. Further to this conclusion, even when not deployed, military personnel tend to work in groups, manage risks when making decisions and live in an atmosphere of uncertainty with frequently changing policies and procedures, postings and little notice for deployment. As such, high levels of emotional stability are essential.

The job demands for both conscientiousness and emotional stability as outlined by Tett and Burnett (2003) are detailed in Table 1. The cues listed under these two personality factors align at all levels with the requirements of the various military

occupations working together in teams within their highly structured, yet ever changing, organization.

Table 1
Job Demands at the Task, Social and Organizational Levels for Conscientiousness and Emotional Stability

Situational Feature	Source/Level		
	Task	Social	Organizational
Conscientiousness Job Demands	Detail, precision, rule-following, deadlines, high quality task completion	Precise and explicit communications, responsibility, dependability	Success, competition, loyalty
Emotional Stability Job Demands	Responsibility with no control over outcomes, high risk management	Handling aggressive coworkers	Atmosphere of uncertainty

Note: Adapted from Tett and Burnett (2003).

Need for Cut-Off Scores

The CAF's current selection model requires that applicants pass a multi-hurdle approach which includes measurements of personality, cognitive ability and the results of a structured interview. If the baseline cognitive ability cut-off scores are met, the scores of all three measures are combined using a compensatory model to determine the applicant's potential to succeed in the CAF and their competitiveness compared to other applicants. These scores form the basis for selection decisions. The selection methods that are presently being applied for the conscientiousness and emotional stability subscales of the TSD-PI during this process are essentially implying that the entire range of possible scores are considered acceptable for entrance into the organization. The primary purpose of a cut-off score is to make better predictions on future performance

that will distinguish applicants who are more likely to succeed at training and on the job from those who are not (Kehoe & Olson, 2005). In accordance with TAT (Tett & Burnett, 2003), ensuring high levels of conscientiousness and emotional stability in the applicants that make it through selection, increases the likelihood of eliciting desired work behaviours that are intrinsically motivated, rather than through the expectation of extrinsic rewards. This should result in a workforce that will perform even when not under supervision. Introducing a minimum acceptable level for each trait on the TSD-PI is one way to establish a highly conscientious and emotionally stable workforce. Cut-off scores also provide additional benefits to the organization, such as the potential reduction in selection costs and time spent conducting interviews for applicants scoring below the cut-off, and the ability to adjust the cut-off based on the needs of the organization (e.g. to allow more or less people through depending on the number of positions available).

The use of cut-off scores in selection is a legally defensible method of narrowing down the applicant pool (Kehoe & Olson, 2005). Prospective employees to the CAF are already required to meet established cut-offs on the CFAT, which is administered upon application to the CAF in conjunction with the TSD-PI. The most common practice for cut-off scores on personality measures are those used in clinical settings for the diagnosis of various conditions, including, but not limited to, antisocial, borderline and avoidant personality disorders (Neal, Fox, Carroll, Holden & Barnes, 1997). However, cut-score selection systems using personality measures are widespread in the private sector, usually requiring applicants to pass a personality assessment as part of a multi-hurdle approach, in conjunction with other selection criteria and methods (Berry & Sackett, 2009).

Although an extensive search resulted in no published studies relating to the

establishment of cut-off scores on personality measures for selection purposes, recent articles by Berry and Sackett (2009) and Bott, O'Connell, Ramakrishnan and Doverspike (2007) explore potential issues relating to the establishment of personality cut-off scores using incumbent data. This suggests that establishing cut-off scores on personality measures for selection is becoming an accepted practice, not only in the private sector, but by researchers as well.

Despite being used as an integral part of the CAF selection model since 2013, cut-off scores for the TSD-PI have yet to be established. The creation of these scores requires the collection of a large amount of criterion-related data. This has proven to be a time consuming undertaking as occupational training data is unavailable until applicants have reached their occupational functional point (OFP), which, depending on the occupation, may take between 2 to 6 years from the time of enrolment. Information on job performance in the form of Performance Evaluation Reports (PERs) may be available earlier than training data, but timing is also a function of the occupation.

Since the commencement of the administration of the TSD-PI to CAF applicants, the accumulation of sufficient CAF predictor and criterion data allow for the examination of relationships between the TSD-PI and desired outcome criteria. Of particular interest to the Canadian Forces Recruiting Group (CFRG) is the creation of a separate set of cut-off scores for officers and NCMs on the conscientiousness and emotional stability subscales of the TSD-PI should there be a theoretical and empirical standpoint for doing so (DGMPPRA, 2016). Relationships between educational criteria and the conscientiousness factor (Chamorro-Premuzic & Furnham, 2003; Paunonen & Ashton, 2001; Major et al., 2006) provide theoretical justification for distinct cut-off scores due to

the higher education selection criteria expected of officer applicants over NCMs. Moreover, evidence linking conscientiousness and emotional stability to leadership (Drasgow, Stark, Chernyshenko, Nye, Hulin, & White, 2012) and managerial performance (Rothmann & Coetzer, 2013), which are typical expectations of entry-level officer positions, presents additional theoretical grounds for separate cut-off scores between the officer and NCM populations.

In addition to differential cut-off scores for officers and NCMs, it is possible to examine the feasibility for distinct cut-off scores for occupational families that have been identified for the NCM population. Extensive research has resulted in the grouping of the majority of the NCM occupations under categories based on common sets of abilities or competencies (Catano & Ibel, 1995). Catano & Ibel (1995) identified five occupational families: administrative, general military, mechanical, operator and technical. These groupings have been used to facilitate the establishment of cut-off scores for the CFAT rather than conducting independent validation studies for each occupation. As such, this research intends to explore the possibility of establishing cut-off scores for occupational families based on job performance outcome criteria (training data cannot be used due to limitations within the dataset, which will be discussed below).

Methods of Establishing Cut-Off Scores

There are several accepted methods for establishing cut-off scores. Often the technique selected depends on the method used to collect the data. Two types of information are relevant, test information and work information or desired outcome criteria (Kehoe & Olson, 2005). The source of each of the two types of information determines the techniques that are most appropriate. The source can either be non-

existent, meaning that the information is unavailable, or it can be collected empirically or through expert judgements.

In the case where both the work information and test information are provided by experts, techniques such as those established by Angoff (1971), Ebel (1972), Nedelsky (1954) or Jaeger (1982) may be appropriate. These techniques require job experts to make two independent judgements and identify incumbents who perform at a desired threshold and then estimate the test scores that this group would achieve on the selection measure (Kehoe & Olson, 2005). These methods are most frequently used for multiple choice threshold knowledge exams (Berry & Sackett 2009). They are inappropriate with a pre-employment measure of personality because it is unreasonable to expect job experts to accurately predict how scores on a personality assessment will relate to an acceptable level of job performance.

Where empirical data is available for the test information, but not the work information, techniques such as Gulliksen's (1987) contrasting groups method or Livingston & Zieky's (1982) borderline group method are appropriate. These methods require experts to identify either high and low performing groups or to divide performers into successful and unsuccessful subgroups based on a desired outcome criterion. These two groups are then compared using empirical data collected on the pre-test measure to determine a cut-off score that differentiates them. These methods are not often used in making employment decisions (Kehoe & Olson, 2005).

Where no work information is available, but empirical test information exists, norm-referenced methods may be used (Cascio, Alexander & Barrett, 1988). These methods are favoured when the concerns are related to administrative issues such as cost

control or minimizing adverse impact to certain groups (Kehoe & Olson, 2005).

Normative information is required and no consideration is given to desired work outcomes.

The final group of methods is employed when empirical data is available for both work outcome information and test information (Kehoe & Olson, 2005). Four very different methods may be employed when both test and criterion data are available. Criterion banding methods use the relationship between test scores and the outcome of interest to establish test score bandwidths, rather than single cut-off scores (Kehoe & Tenopyr, 1994). Decision-theoretic methods are used when multiple competing organizational values need to be optimized (Martin & Raju, 1991). Regression methods, either forward (Chuang, Chen & Novick, 1981) or reverse (Cascio et al., 1988), identify a cut-score that is statistically related to a criterion score that defines minimal competence on a desired outcome. Finally, expectancy chart methods (Taylor & Russell, 1939) represent a graphical version of the regression techniques, the chart is observed and the cut-off is chosen that meets the score associated with the level of work criteria desired by the organization (Kehoe & Olson, 2005).

For this study, empirical data is available for both the TSD-PI entrance test and work outcome information; therefore, regression methods will be used to establish the personality cut-off scores. The difference between forward and reverse regression is the order in which the variables are entered into a regression model. In forward regression the criterion scores are regressed on the test scores, and in reverse regression the test scores are regressed on the criterion scores (Kehoe & Olson, 2005). Jeanneret and Stelly (2003) compared the two methods and found that reverse regression produces cut-off

scores that are closer to the mean performance on the predictor measure as opposed to those produced by forward regression. In contrast, forward regression has the ability to produce extreme cut-scores when the performance threshold is set either very low or very high. The choice of which regression method to use depends on the needs of the organization. Where the desire is to predict the likelihood of success on the outcome criteria, forward regression is recommended. However, if the concern is to mitigate adverse impact or maximize the overall pass rate, then reverse regression is preferred (Kehoe & Olson, 2005). As adverse impact does not appear to be a concern with the TSD-PI (Darr, 2016), the forward regression methods will be used to recommend cut-off scores.

Current Research

The aim of this study was to establish cut-off scores for the TSD-PI scales of conscientiousness and emotional stability using training performance and performance on the job as the criterion variables. Forward regression models were used to ascertain the validity of these two subscales in predicting scores on the desired outcome variables prior to recommending cut-off scores for the selected sub-groups. It was expected that the results found for the samples in this study would be consistent with previous personality performance research and produce small but positive effects between the predictor and criterion variables. Where the goal was to differentiate between various populations, no hypotheses were made. Rather, two research questions were examined:

- (1) Is it feasible to recommend distinct cut-off scores for the NCM and officer populations on the conscientiousness and emotional stability subscales of the TSD-PI using both training results and job performance as outcome criteria?

(2) Is it feasible to recommend distinct cut-off scores for the conscientiousness and emotional stability subscales of the TSD-PI for the established NCM occupational families using job performance as the outcome criteria?

Method

Participants

The two datasets used in this study, one containing training data and the other with performance outcomes, provide an accurate representation of the CAF population (Darr 2016). There is a substantially higher number of males (83% between the two samples) than females, and more NCMs (71% between the two samples) than officers in the samples. The dataset containing performance data is comprised of only English language evaluations; however, the training data sample represents the population with more English language speaking participants (74.7%) than French. Information on age was not available; however, due to the minimum age requirements of the CAF and known demographics regarding the population, the age of participants at the time of application likely ranged from 15 to 58 years with a mean of 24 (Darr, 2016). Participants will have aged during the collection of data between 2-6 years depending on their occupation; however, research supports the stability of personality factors within adults over time (Costa & McCrae, 1997). A breakdown of the available demographic data is presented in Table 2. Sample sizes, before data cleaning, were 7252 for the first dataset containing training outcomes and 1232 for the dataset with job performance information.

Table 2

Dataset Demographic Information.

Demographic	Training Outcomes Dataset <i>N</i> = 7252	Job Performance Outcomes Dataset <i>N</i> = 1232
Gender		
Male	4893 (67.47%)	1060 (86.04%)
Female	1049 (14.46%)	169 (13.72%)
Missing	1310 (18.06%)	3 (0.24%)
First Official Language		
French	1515 (20.89%)	0
English	5415 (74.67%)	1232 (100.00%)
Missing	322 (4.44%)	0
Rank		
Officer	1910 (26.34%)	181 (14.69%)
NCM	4149 (57.21%)	1051 (85.31%)
Missing	993 (13.69)	0

Note: First Official Language refers to the language that participants identified as their primary language, which may be different from the language they wrote their selection measures in.

As separate TSD-PI measures and normative data exist for the French and English language applicant populations, the training outcomes dataset was divided into the two language groups for analyses. This process resulted in sample sizes for the French language population that were too small to be subdivided into the desired groups for meaningful analyses. In conjunction with job performance outcomes not being available for this group, the decision was made to focus this study's examination on the English language population only, for which sufficient sample sizes and data for both outcome criteria had been compiled. In order to create English language only samples, all participants who had identified their first official language as French, or had written the selection measures in French, were removed from the training dataset. Demographics

after this process and following data cleaning (as described in detail in the results section) are shown in Table 2b.

Table 2b

Cleansed Dataset Demographic Information – English Language Participants Only.

Demographic	Training Outcomes Dataset <i>N</i> = 3530	Job Performance Outcomes Dataset <i>N</i> = 1232
Gender		
Male	2843 (80.54%)	1060 (86.04%)
Female	682 (19.32%)	169 (13.72%)
Missing	5 (0.14%)	3 (0.24%)
Rank		
Officer	1178 (33.37%)	181 (14.69%)
NCM	2352 (66.63%)	1051 (85.31%)

Note: Participants retained for this study include those who both selected their first official language as English and completed the English versions of the selection measures.

Measures

Trait Self Descriptive Personality Inventory (TSD-PI). The TSD-PI (Darr, 2016) contains a total of 75 items, 15 items to measure each of the five factors of the FFM; conscientiousness, agreeableness, extraversion, openness to experience and emotional stability. The items of the five subscales are interspersed throughout the inventory and separated into two parts. Part 1 is made up of 26 adjectives (“talkative”) and Part 2 contains 49 statements (“In meetings, I tend to let others do most of the talking”). Respondents are required to rate the degree to which each statement or adjective is characteristic of them using a 7-point Likert-type scale where 1 = extremely uncharacteristic and 7 = extremely characteristic. The TSD-PI may be administered in a paper-pencil format or through one of two computerized formats (electronic, eTSD or internet, iTSD). The TSD-PI is administered in the respondent’s language of choice,

either English or French. Although there is no time limit for completion, the computerized versions time out after 40 minutes as a method of test control; the majority of respondents complete the questionnaire within 15 minutes. Subscale totals are calculated and converted into percentiles using normative data for ease of interpretation. For the purpose of this study, analyses were completed using the raw data rather than the percentile conversions. Only the conscientiousness and emotional stability scale totals were analyzed although applicants completed the full TSD-PI. Internal consistency reliabilities for the samples used in this study ranged from $\alpha = 0.89 - 0.91$ for the conscientiousness subscale and $\alpha = 0.86 - 0.91$ for emotional stability.

Training Outcomes. For the purpose of this study, training information was split into two categories, basic training and occupation-related training. Basic training included the Basic Military Qualification (BMQ) for NCM participants and the Basic Military Officer Qualification (BMOQ) for officers. These courses are the same for all NCMs or officers entering the CAF, regardless of occupation. Occupation-related training included only the participant's first job-related course following basic training, as subsequent courses suffered from range restriction due to applicants being weeded out at previous stages of occupational training. In many occupations, the completion of the first occupational course also marks the point at which the members reach their occupational functional point (OFP). OFP is the point at which all qualifications required for their first employment in the military occupation have been met. This occurs following completion of their first developmental period (DP). Subsequent courses are considered more specialized training. For example, a Vehicle Technician's first occupational course would be Vehicle Technician DP1, outcomes from this training have been retained, whereas

Vehicle Technician DP2 have been excluded from this study. Course information is recorded as pass or fail. Although percentage and grade information is available for some of the dataset, using this criteria, rather than the pass or fail information would substantially reduce the size of the sample. Successful completion of the training outcomes requires the expression of conscientiousness and emotional stability trait level cues. Military training is also very team oriented, requiring responses to social level cues. Training outcomes provide a form of evaluation in line with path 7 of Tett & Burnett's (2003) Personality Trait-Based Model of Job Performance (Figure 1) which allowed for the determination of whether or not the participants displayed desired work behaviours (path 6).

Job Performance. Information on job performance is collected through Performance Evaluation Reports (PERs) which are completed on an annual basis. PERs are separated into multiple sections. The first assessment section evaluates the employee's performance based on 16 criteria (i.e., supervising, evaluation and development of subordinates, team building, leading change, working with others, problem solving, decision making, effectiveness, initiative, verbal communication, written communication, applying job knowledge/skills, resource management, accountability, reliability and ethics and values). The supervisor rates the employee using a 6-point Likert-type scale (i.e., 1= unacceptable, 2=needs improvement, 3=developing, 4=skilled, 5=exceeded standard, 6=mastered). For five of the criteria (i.e., supervising, evaluation and development of subordinates, team building, leading change and written communication), supervisors have the option of selecting "not observed". A final criterion finishes off this section and evaluates the employee's conduct on/off duty (either

“acceptable” or “unacceptable”). The second evaluation section of the PER rates the employee’s potential for performance at the next rank level. It is comprised of six criteria (i.e., leadership, professional development, communication skills, planning and organizational skills, administration and dedication). Ratings are based on a 4-point Likert-type scale (i.e., 1=low, 2=normal, 3= above average, 4=outstanding). Employees are also given a ranking within their unit compared to others in the same occupation and a promotion recommendation (either “no”, “developing”, “ready” or “immediate”). For this analysis, only the first evaluation section (i.e., performance) is used. The reason for this is that the “potential” section is a subjective determination of how the person will perform at the next rank level, rather than an evaluation of behaviours which constitute current performance. Correlations of the 16 performance criteria listed above (excluding conduct) ranged from .58 to .89, suggesting a strong relationship between the individual performance criteria. Exploratory factor analysis was conducted using principal component analysis of the items. One factor, which explained 74.37% of the variance was extracted. All 16 items loaded strongly onto the single performance factor (factor loading values ranged from .81 to .90). Rotation could not be completed due to the single factor extraction. As such, a composite score for the performance outcome variable was created by averaging participants’ scores on the 16 criteria. For participants whose supervisors indicated “not-observed” for one or more of the criteria, these items were omitted in the calculation of their composite score on this measure. Internal consistency reliabilities for the composite scale ranged from $\alpha = 0.96 - 0.98$ with the study samples. During analyses, mean scores of 4 or above (corresponding to the “skilled” descriptor) were considered successful performance. Participants who scored less than a mean of 4

were considered to be unsuccessful (corresponding performance descriptors are: “unacceptable,” “needs improvement,” or “developing”). The PER is a form of evaluation congruent with path 7 of Tett and Burnett’s (2003) Personality Trait-Based Model of Job Performance (Figure 1), which evaluates individual response to situational cues at all three levels of work demands. This informed path 6 of the model in allowing for the determination of whether or not the behaviours assessed were desired. Although not addressed in this thesis due to the absence of promotion information, successful job performance as rated by the PER also serves as a form of extrinsic reward (path 9 in Figure 1) because it is used to inform promotion decisions and ranking within the unit.

Procedure

The data for this study were previously collected by the CAF for the purpose of selection or performance management. The study involves two samples: the first contains TSD-PI scores and training outcome data, and the second contains TSD-PI scores job performance data. Both data sets include demographic information. The data were provided for analyses under a collaborative research agreement between the CAF’s Director General Military Personnel Research and Analysis (DGMPRA), this study’s author and Saint Mary’s University (DGMPRA, 2016). Personal identifiers were removed from the datasets prior to the researcher having access. The demographic information includes: gender, date of TSD-PI administration and member’s occupation, as well as whether the applicant was enrolled as an officer or NCM. The TSD-PI scores were collected at the time of application for CAF entry, as early as January of 2010, when the inventory was first administered to applicants for research purposes. TSD-PI scores were obtained from a Treasury Board database. The training outcome data and job

performance information was added later, when it became available, and matched to the applicable participant's data. Despite inclusion in the datasets, gender was not considered a variable of interest beyond establishing generalizability. Differences between male and female responses to the TSD-PI have been considered small and comparable to those found on other well-known measures such as the NEO-PI-R (Darr & Kemp, 2011).

The two datasets were not completely independent as 150 of the 1232 participants from the job performance dataset were also part of the training outcome dataset ($N = 7252$). The extent of overlap was likely reduced during data cleaning, due to the removal of cases from the training dataset that did not meet the classification criteria of basic or initial occupational training. Cases were also removed on account of missing TSD-PI data. Although not completely independent, the outcome variables assessed in each dataset were different (training and performance).

The two datasets represent three populations:

- (1) Applicants ($N = 1361$): a sample containing information regarding training outcomes on basic training for those who had been admitted to the CAF;
- (2) Prospective Incumbents ($N = 2169$): those members who had successfully completed basic training and the results of their first occupation-related training course. This sample had yet to be employed in their desired occupation. Employment is contingent on successful training results at this stage; and
- (3) Incumbents ($N = 1232$): a sample with information on their job performance, following successful completion of occupation-related training.

Each of these samples were divided into groups based on commission (officer or NCM)

and by NCM occupational families for evaluation using forward regression methods, as shown in Figure 2. The occupational family groups listed under the Applicant population in the figure were not analyzed because basic training is the same for all NCMs regardless of the family that their occupation is categorized into. The remainder of the samples that have been crossed out in the figure could not be analyzed due to issues with the data as described in the results section.

TSD-PI: ESTABLISHING CUT-OFF SCORES

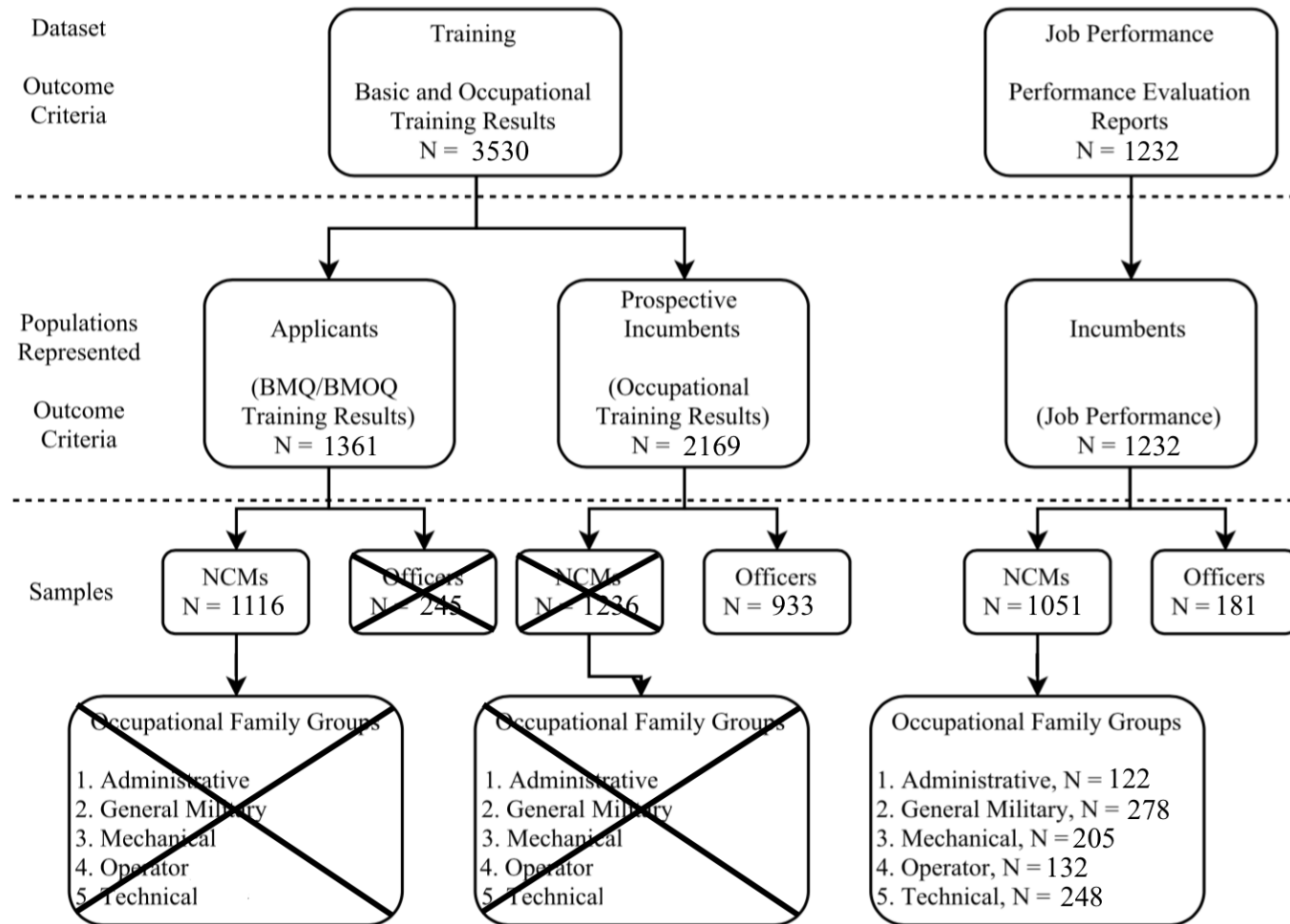


Figure 2. Breakdown of Datasets into Samples for Analyses. Crossed out samples were not analyzed due to various data issues (described in the results section).

Results

Missing Value Analysis and Data Preparation

Prior to analyses, the datasets were prepared by assessing missing data and identifying duplicates, outliers and influential cases. Creation of the English only sample from the original training outcome dataset saw the removal of 100 individuals who wrote their selection measures in their second official language, 322 who were missing the language variable and 1515 who identified their first official language as French. Missing Value Analyses revealed that the dataset contained a significant number of cases without TSD data. As the subtests of the TSD were administered at the same time, those missing data in one of the variables also tended to lack data in the other predictor variable. Therefore, the cases with missing data were most likely candidates selected to join the CAF prior to the TSD being administered as part of the selection process. Pairwise deletion decreased the sample to 3530 and allowed for the retention of 1426 cases with conscientiousness scores and 1413 with emotional stability scores in addition to the training outcome variable.

The second dataset had a minimal quantity of missing data, with only 5 cases without conscientiousness scores and 2 missing emotional stability scores. Pairwise deletion allowed for the retention of 1227 cases with conscientiousness data and 1230 with emotional stability scores in addition to the job performance outcome variable. No duplicates existed in either dataset and Cook's distance (Cook, 1977) values confirmed no influential outliers impacted the datasets (highest values were .10 in the training outcome dataset and .02 in the performance dataset). The remaining cases in both datasets left adequate sample sizes for the subgroups identified for analyses in Figure 2.

Assumptions

Several assumptions were assessed to ensure that conclusions regarding the CAF applicant population could be drawn from the selected samples. The predictor variables, both of which are continuous, were evaluated by calculating zero order correlations, which indicated that multicollinearity was not an issue. Correlations along with means and standard deviations of all variables in the study can be viewed in Tables 3 and 4. The highest correlation between the conscientiousness and emotional stability variables for both datasets was .41, $p < .001$. Normal probability plots were reviewed to evaluate the means of the residuals and revealed random, normally distributed variables with means of zero. The graphs revealed no concerns with respect to linearity or homoscedasticity. Skewness and kurtosis statistics fell between the accepted range of ± 2 to support normality (Field, 2009); additional confirmation was achieved through visual assessment. The residuals of each predictor were symmetrical at all levels of the variable and the mean values of the outcome variables for each of the predictors approximated a straight line, showing a linear relationship. Range restriction on the predictor and performance outcome scales was noted as a potential issue and is discussed as a limitation following the results section of this thesis.

Initial Analyses

The two datasets were divided into the three samples described previously, and descriptive statistics were compared to include means and standard deviations. Correlations were run between the independent and dependent variables to determine the direction and strength of relationships that existed for each sample and the outcome criteria being measured. A summary of intercorrelations, means and standard deviations

for predictor and the respective training and performance outcome variables are presented in Tables 3 and 4.

The relationship between the independent variables (IVs, conscientiousness and emotional stability) and each dependent variable (DV, basic and occupation-related training outcomes and job performance criteria) were expected to be small but positive (Barrick et al., 2001). The first two samples produced near-zero correlations. The relationship between conscientiousness and emotional stability with basic training results for the applicant sample was $.05, p > .05$ for conscientiousness and $.02, p > .05$ for emotional stability. The correlations for the prospective incumbent sample were small and negative, with emotional stability having a non-significant correlation with the occupational training outcome variable of $-.03, p > .05$. The relationship between conscientiousness and occupational training results gave a significant correlation of $-.10, p < .05$. Although the final relationship was significant, the small correlations suggest that conscientiousness and emotional stability may be unrelated to training outcomes within the given samples. The correlations between the IVs and the job performance scores (incumbent sample) showed more promising results for the likelihood of establishing cut-off scores for the subscales of the TSD. The correlation between conscientiousness and job performance was $.17, p < .01$, and emotional stability was found to be related to job performance with a correlation of $.22, p < .01$.

Table 3
 Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Training Outcome Variables

Measure	Applicant Sample (Basic Training)			Prospective Incumbent Sample (Occupation Training)		
	1	2	3	1	2	3
1. Conscientiousness	-			-		
2. Emotional Stability	.41***	-		.40***	-	
3. Pass/Fail Training Outcome	.05	.02	-	-.10*	-.03	-
<u>All Participants</u>						
<i>N</i>	877	876	1361	549	547	2169
<i>M (SD)</i>	83.38 (10.08)	83.20 (12.33)	.82 (.39)	85.85 (10.20)	84.46 (12.55)	.89 (.32)
<u>Successful (Pass)</u>						
<i>N</i>	685	685	1113	491	489	1921
<i>M (SD)</i>	83.65 (9.86)	83.32 (11.97)	-	85.51 (10.26)	84.31 (12.47)	-
<u>Unsuccessful (Fail)</u>						
<i>N</i>	192	191	248	58	58	248
<i>M (SD)</i>	82.42 (10.81)	82.75 (13.56)	-	88.74 (9.34)	85.71 (13.24)	-

Note: *M* = Mean; *SD* = Standard Deviation. Missing data were pairwise deleted, therefore the *N* value is listed for each comparison pair. * $p < .05$; *** $p < .001$ (two-tailed).

Table 4
Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Performance Outcome Variables

Measure	Incumbent Sample (Job Performance)		
	1	2	3
1. Conscientiousness	.90		
2. Emotional Stability	.36**	.89	
3. Performance Outcome	.17**	.22**	.97
<u>All Participants</u>			
<i>N</i>	1227	1230	1232
<i>M (SD)</i>	85.72 (10.02)	84.17 (11.83)	4.78 (.88)
<u>Successful Performance</u>			
<i>N</i>	970	972	974
<i>M (SD)</i>	86.20 (9.91)	85.15 (11.40)	5.13 (.61)
<u>Unsuccessful Performance</u>			
<i>N</i>	257	258	258
<i>M (SD)</i>	83.91(10.24)	80.52 (12.72)	3.46 (.34)

Note: *M* = Mean; *SD* = Standard Deviation. Missing data were pairwise deleted, therefore the *N* value is listed for each comparison pair. Scale reliabilities (α) appear in bold font. ** $p < .01$ (two-tailed). Successful performance was determined by a mean score of 4 or greater, which corresponds to the “skilled” descriptor. Unsuccessful performance is a mean score less than 4.

Validity Across NCM Versus Officer Groups. The three samples were subsequently divided further into officer and NCM groupings to determine if group membership impacted the relationship between the variables. Due to limitations with the sample sizes after division, the officer group was not analyzed on the basic training outcome variable as the sample size for those who were unsuccessful was extremely small ($N=9$, only 3.7%, compared to $N=236$ for those who were successful). As such, the officers’ training outcome criteria included only the occupation-related training results.

In contrast, the NCM group had a much higher failure rate on basic training (21.4%) than occupation-related training (2.4%). As such, the number of NCM participants who were unsuccessful on occupation-related training was too small for analysis ($N=30$ out of a total of 1206 NCMs) and the training outcome criterion for this group that was included for analysis encompassed only basic training results. Correlations were re-run to see if the strength and direction of the relationships between the TSD-PI subtests, and training and performance outcomes differed between officers and NCMs. Although correlations remained small and non-significant for the training outcome variables for both NCMs and officers, as shown in Table 5, differences existed for the performance outcome criteria (Table 6). Where conscientiousness and emotional stability were found to have small, non-significant correlations with job performance for officers (.00 and .01 respectively, $p > .05$), significant relationships existed for both IVs and the performance outcome variable for NCMs (conscientiousness: .20, $p < .01$ and emotional stability: .24, $p < .01$).

Table 5

Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Training Outcome Variables for NCMs and Officers

Measure	NCMs			Officers		
	Applicant Sample			Prospective Incumbent Sample		
	(Basic Training)			(Occupation Training)		
	1	2	3	1	2	3
1. Conscientiousness	-			-		
2. Emotional Stability	.40***	-		.49***	-	
3. Pass/Fail Training Outcome	.05	.03	-	-.15	-.01	-
<u>All Participants</u>						
<i>N</i>	734	733	1116	171	171	933
<i>M (SD)</i>	82.88(10.09)	83.01 (12.42)	.79 (.41)	86.30 (9.81)	85.76 (12.68)	.77 (.42)
<u>Successful (Pass)</u>						
<i>N</i>	549	549	877	123	123	715
<i>M (SD)</i>	83.17 (9.76)	83.21 (12.00)	-	85.41 (9.88)	85.67 (12.71)	-
<u>Unsuccessful (Fail)</u>						
<i>N</i>	185	184	239	48	48	218
<i>M (SD)</i>	82.01 (10.68)	82.42 (13.60)	-	88.58 (9.35)	85.98 (12.73)	-

Note: *M* = Mean; *SD* = Standard Deviation. Missing data were pairwise deleted, therefore the *N* value is listed for each comparison pair. *** $p < .001$ (two-tailed).

Table 6

Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Job Performance Outcome Variables for NCMs and Officers

Measure	NCM			Officer		
	Incumbent Sample			Incumbent Sample		
	(Performance Evaluation Reports)			(Performance Evaluation Reports)		
	1	2	3	1	2	3
1. Conscientiousness	.90			.91		
2. Emotional Stability	.37**	.89		.32**	.86	
3. Performance	.20**	.24**	.98	.00	.01	.96
<u>All Participants</u>						
<i>N</i>	1047	1046	1051	180	180	181
<i>M (SD)</i>	85.85 (9.92)	84.05 (12.03)	4.73 (.90)	84.99 (10.61)	84.92 (10.64)	5.12 (.70)
<u>Successful Performance</u>						
<i>N</i>	803	805	806	167	167	168
<i>M (SD)</i>	86.47 (9.68)	85.16 (11.54)	5.11 (.62)	84.92 (10.90)	85.10 (10.69)	5.23 (.56)
<u>Unsuccessful Performance</u>						
<i>N</i>	244	245	245	13	13	13
<i>M (SD)</i>	83.80 (10.42)	80.41 (12.85)	3.46 (.35)	85.92 (5.99)	82.62 (9.97)	3.51 (.26)

Note: *M* = Mean; *SD* = Standard Deviation. Missing data were pairwise deleted, the *N* value is listed for each comparison pair. Bold values on the diagonal indicate scale reliabilities (α). ** $p < .01$ (two-tailed). Successful performance was determined by a mean score of 4 or greater, which corresponds to the “skilled” descriptor. Unsuccessful performance is a mean score less than 4.

Table 7
Cohen's *d* Mean Comparisons Showing the Differences Between NCM and Officer Groups for Predictor and Outcome Variables

	Conscientiousness	Emotional Stability	Outcome Variable
<u>Training Outcome Dataset</u>			
Overall Sample	.34	.22	.05
Successful (Pass)	.23	.20	0
Unsuccessful (Fail)	.65	.27	0
<u>Performance Outcome Dataset</u>			
Overall Sample	.08	.08	.48
Successful (Mean \geq 4)	.15	.01	.20
Unsuccessful (Mean $<$ 4)	.25	.19	.16

Note: Effect sizes of $d = .2$ are considered small, $d = .5$ are medium effects and $d = .8$ are large effects (Cohen, 1992)

To further justify splitting the datasets based on commission (i.e. NCM or Officers), means were compared using Cohen's *d* (Cohen, 1992) in order to determine if sufficient discrepancies existed between the groups within each sample to warrant separate cut-off scores. The results of this analysis are presented in Table 7. The most notable differences occurred between officers and NCMs on the conscientiousness measure for those who failed their training courses ($d = .65$, approaching a large effect) and the overall job performance measure ($d = .48$, a medium effect), with officers scoring higher on both variables. Values approaching medium effects were discovered for the differences between officers and NCMs on the conscientiousness scale for the overall sample ($d = .34$) and the emotional stability measure for unsuccessful participants ($d = .27$) in the training dataset. Many other comparisons resulted in at least a small effect. As a result of the findings suggesting differences between officers and NCMs on various

measures, different cut-off scores were justified for the two populations and the data remained split for the remainder of the analyses.

Interpretations of the analyses thus far have resulted in the elimination of training data as criterion variables for the recommendation of cut-off scores for the TSD-PI. This decision was made based on the near-zero correlations for both predictor variables with the combined (NCM and officer) applicant sample's basic training results and their negative (however small) correlations with the occupational training outcome criteria. These findings suggest that the NCM and officer comparisons from this dataset are showing differences in the independent variables' ability to predict the distinct outcome criteria (basic training versus occupational training) rather than the differences between the variables' ability to predict NCM and officer training success. Observing each group independently, near-zero correlations suggest that conscientiousness may be unrelated to NCM success on basic training and officer success on occupational training. The same interpretation can be made for emotional stability and its ability to predict basic training results in the NCM population. The small negative correlations found between conscientiousness and occupational training results suggested that establishing cut-off scores may eliminate candidates within this sample who would succeed on training and select those who would potentially be unsuccessful. Due to these findings, training outcomes were removed as potential criteria for establishing cut-off scores for the conscientiousness and emotional stability subscales of the TSD-PI. All further analyses focused on job performance as the outcome criteria of interest.

Validity Across Occupational Families. Finally, initial analyses continued to determine if there was value in recommending cut-off scores for the family groups

present within the NCM incumbent population. In addition to the near zero correlations for NCMs discussed above, sufficient data were not available to divide the applicant and prospective incumbent samples by family for the training outcome variable due to the low NCM failure rate on occupation-related training, and the fact that basic training is the same for all NCMs, regardless of their occupation. Therefore, only analyses regarding the TSD-PI subscales' relationships with performance outcome criteria for the incumbent sample could be conducted. It is important to note that the subdivided samples are not independent of the overall NCM sample. The possibility of confounded results will be discussed in the limitations section.

Correlations and descriptive statistics were observed for the variables to determine the feasibility of differing cut-off scores for each of the five occupational family groups. Table 8 provides a summary of intercorrelations, means, and standard deviations for predictor and performance outcome variables for each group. Both conscientiousness and emotional stability were significantly related to performance in the general military family with conscientiousness accounting for 3.61% of the relationship ($r = .19$, $r^2 = .036$, $p < .01$) and emotional stability 1.96% ($r = .14$, $r^2 = .02$, $p < .05$). Effect sizes were larger for the mechanical family with conscientiousness accounting for 11.56% ($r = .34$, $r^2 = .116$, $p < .01$) and emotional stability 7.29% ($r = .27$, $r^2 = .073$, $p > .01$) and the technical family; conscientiousness, 4.41% ($r = .21$, $r^2 = .044$, $p > .01$) and emotional stability, 12.96% ($r = .36$, $r^2 = .13$, $p > .01$). However, only emotional stability (accounting for 4.89%) was correlated to performance scores for the administrative family (conscientiousness: $r = .13$, $r^2 = .017$, $p > .05$, emotional stability: $r = .22$, $r^2 = .049$, $p < .05$) and neither predictor variable resulted in a significant relationship with the

outcome measure for the operator group, each variable accounting for 2.56% of the relationship (conscientiousness: $r = .16$, $r^2 = .026$, $p > .05$, emotional stability: $r = .16$, $r^2 = .026$, $p > .05$). Non-significant values may be a function of smaller sample sizes for the administrator and operator groups. There are almost 100 fewer data points compared to the other occupational groups. Although relationships between the predictor variables and scores on the performance measurement varied depending on group membership, z-test results (Table 9) showed that only four comparisons proved to be statistically different. For conscientiousness, this included the administrative versus mechanical groups ($z = 1.93$, $p = .05$) and the NCM versus mechanical group comparison ($z = 1.97$, $p < .05$). For emotional stability, it was the technical versus general military ($z = 2.69$, $p < .01$) and technical versus operator ($z = 1.98$, $p < .05$) group comparisons that were different. Further analyses were undertaken to explore the possibility of a requirement for NCM applicants to meet differing cut-off scores depending on the family group that their desired occupation belongs to.

Prior to describing the additional analyses, Table 10 summarizes how the training criterion variables were eliminated during the preceding steps. Subsequent stages involved only the job performance outcome criteria from the officer, NCM and occupational family incumbent samples.

Table 8
 Summary of Intercorrelations, Means, and Standard Deviations for Predictor and Job Performance Outcome Variables for Occupational Family Groups (Incumbent Sample)

Measure	Administrative			General Military			Mechanical			Operator			Technical		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
1. Conscientiousness	.90			.89			.90			.91			.91		
2. Emotional Stability	.35**	.91		.36**	.88		.43**	.89		.42**	.88		.32**	.90	
3. Performance Outcome	.13	.22*	.98	.19**	.14*	.97	.34**	.27**	.98	.16	.16	.96	.21**	.36**	.98
<u>All Participants</u>															
<i>N</i>	122	122	122	277	277	278	204	205	205	132	132	132	248	249	249
<i>M (SD)</i>	88.97 (9.50)	83.24 (12.80)	4.99 (0.87)	85.19 (9.75)	85.81 (11.13)	5.12 (.77)	85.23 (9.63)	83.34 (11.83)	4.38 (.95)	85.82 (9.95)	84.20 (11.17)	4.61 (.79)	85.52 (10.38)	83.11 (12.72)	4.56 (.91)
<u>Successful Performance</u>															
<i>N</i>	103	103	103	252	252	253	130	131	131	97	97	97	171	171	171
<i>M (SD)</i>	88.88 (9.87)	83.94 (12.54)	5.27 (.62)	85.37 (9.64)	85.91 (11.04)	5.27 (.57)	87.15 (8.50)	84.99 (11.52)	4.97 (.62)	86.37 (9.98)	84.37 (10.74)	4.97 (.56)	86.43 (9.99)	85.56 (11.47)	5.04 (.64)
<u>Unsuccessful Performance</u>															
<i>N</i>	19	19	19	25	25	25	74	74	74	35	35	35	77	78	78
<i>M (SD)</i>	89.42 (7.39)	79.42 (13.85)	3.49 (.33)	83.40 (10.82)	84.76 (12.17)	3.45 (.50)	81.85 (10.60)	80.40 (11.88)	3.34 (.32)	84.29 (9.84)	83.71 (12.47)	3.60 (.23)	83.48 (11.01)	77.74 (13.72)	3.5 (.34)

Note: *M* = Mean; *SD* = Standard Deviation. Missing data were pairwise deleted, therefore the *N* value is listed for each comparison pair. Bold values on the diagonal indicate scale reliabilities (α). * $p < .05$, ** $p < .01$ (two-tailed). Successful performance was determined by a mean score of 4 or greater, which corresponds to the “skilled” descriptor. Unsuccessful performance is a mean score less than 4.

Table 9

Summary of Z-Test Intercorrelation Comparisons for Overall NCM and Occupational Family Groups for
Conscientiousness and Emotional Stability as predictors of Job Performance (Incumbent Sample)

	Administrative	General Military	Mechanical	Operator	Technical
Conscientiousness					
Administrative					
General Military	0.56				
Mechanical	1.93*	1.74			
Operator	0.24	0.29	1.71		
Technical	0.74	0.24	1.48	0.48	
Overall NCM	0.74	0.15	1.97*	0.44	0.15
Emotional Stability					
Administrative					
General Military	0.75				
Mechanical	0.46	1.47			
Operator	0.49	0.19	1.02		
Technical	1.37	2.69**	1.05	1.98*	
Overall NCM	0.22	1.53	0.42	0.89	1.86

Note: * $p < .05$, ** $p < .01$ (two-tailed).

Table 10

Criterion Variables Included in Subsequent Analyses to Recommend Cut-Off Scores for Officers, NCMs and Occupational Family Groups

	Applicant Sample (Basic Training)	Prospective Incumbent Sample (Occupation-Related Training)	Incumbent Sample (Job Performance)
Officers	Eliminated due to insufficient data.	Eliminated due to near zero and small negative correlations.	Included in subsequent analyses.
NCMs	Eliminated due to near zero correlations.	Eliminated due to insufficient data.	Included in subsequent analyses.
Occupational Family Groups	Not applicable (because the training is the same for all groups)	Eliminated due to insufficient data.	Included in subsequent analyses.

Establishing Cut-Off Scores

Forward Regression Analyses. Multiple forward linear regression analyses were used to further evaluate the relationships between conscientiousness and emotional stability, and the job performance criteria for each of the samples identified (officers, NCMs and occupational families). The subscales are independent; however, the variables were entered as simultaneous predictors into the regression models as they are used together during CAF selection. This practice is consistent with that used in other CAF validity studies for the purpose of establishing cut-off scores for the CFAT, which also consists of multiple subscales (e.g., Piasentin, 2010; Scholtz, 2004). Based on empirical evidence supporting conscientiousness and emotional stability as predictors of overall job performance (Barrick et al., 2001), small to moderate effects were expected for each of the samples. A summary of the regression results for conscientiousness and emotional stability as predictors of job performance for the NCMs and officers are presented in Table 11.

Table 11
Regression Analyses – Conscientiousness and Emotional Stability as Predictors of Job Performance for NCMs and Officers (Incumbent Sample)

	NCMs			Officers		
	B	SE B	β	B	SE B	β
Conscientiousness	.012	.003	.132***	.000	.005	-.004
Emotional Stability	.014	.002	.189***	.001	.005	.015
R ²		.072			.000	
N - Conscientiousness		1047			180	
N - Emotional Stability		1050			180	

Note: *** $p < .001$.

Consistent with previous research, although the effect sizes were small, conscientiousness and emotional stability were significant predictors of on-the-job performance for NCMs (conscientiousness: $B = .012$, $SE B = .003$, $\beta = .132$, $p < .001$; emotional stability: $B = .014$, $SE B = .002$, $\beta = .189$, $p < .001$), showing that as scores on the TSD-PI subtests increase, so do NCMs' performance scores. The regression results for the officer group did not correspond to findings from earlier studies. Neither conscientiousness nor emotional stability were significant predictors of job performance (conscientiousness: $B = .000$, $SE B = .005$, $\beta = -.004$, $p > .05$; emotional stability: $B = .001$, $SE B = .005$, $\beta = .015$, $p > .05$). The results, however, were consistent with the initial analyses conducted on the variables, which showed no relationship between the predictor and outcome variables for our sample of the officer population.

Additional regression analyses determined the ability of conscientiousness and emotional stability to predict job performance within the family groups. The results of these analyses are presented in Table 12. Conscientiousness was a significant predictor of performance for general military ($B = .012$, $SE B = .005$, $\beta = .159$, $p < .05$) and mechanical ($B = .027$, $SE B = .007$, $\beta = .271$, $p < .001$) occupations. Emotional stability significantly predicted performance for administrative ($B = .013$, $SE B = .006$, $\beta = .194$, $p < .05$), mechanical ($B = .013$, $SE B = .006$, $\beta = .156$, $p < .05$) and technical ($B = .023$, $SE B = .004$, $\beta = .321$, $p < .001$) occupational families. Neither predictor criterion was a significant predictor of job performance for the operator family of occupations. Where results were significant, the standardized validities (β) discovered in the regressions were used in the next step to recommend cut-off scores for the various populations.

Table 12

Regression Analyses – Conscientiousness and Emotional Stability as Predictors of Job Performance for Occupational Families (Incumbent Sample)

	Administrative			General Military			Mechanical			Operator			Technical		
	B	SE B	β	B	SE B	β	B	SE B	β	B	SE B	β	B	SE B	β
Conscientiousness	.006	.009	.063	.012	.005	.159*	.027	.007	.271***	.009	.007	.116	.009	.005	.107
Emotional Stability	.013	.006	.194*	.006	.004	.085	.013	.006	.156*	.008	.007	.110	.023	.004	.321***
R ²		.050			.042			.135			.036			.137	
N - Conscientiousness		122			277			204			132			248	
N - Emotional Stability		122			277			205			132			249	

Note: * $p < .05$. *** $p < .001$.

Recommending Criterion-Related Cut-Off Scores. Where found to be significant, the regression results from the job performance outcome criteria were used to recommend cut-off scores for the conscientiousness and emotional stability subscales of the TSD-PI. In accordance with Tables 12 and 13, this included both predictors for the NCM sample and the mechanical family sample. Regression results were also used for the conscientiousness variable for the general military family and the emotional stability variable for the administrative and technical families. The test scores were converted to z-scores to simplify the regression weights for each variable so that the intercept was zero and the slope was equal to the validity coefficient discovered in the earlier run regressions. Various levels of work thresholds were observed in order to determine how the resulting cut-off impacted the applicant population, prior to recommending a cut-score for each group. The performance thresholds that were examined, began with average performance (i.e., 50%) and various thresholds below average (i.e., 40%, 35%, 30%, 25%, and 20%). These thresholds correspond to criterion z-scores of 0.0, -0.25, -0.39, -0.52, -0.67, and -0.84. On the job performance outcome, incumbents considered to be successful have an average composite score of 4 or above, which corresponds to the descriptor “skilled” on the performance evaluation. The above mentioned criterion z scores were input back into the linear regression formula (Kehoe & Olson, 2005), along with the validities discovered in the forward regressions from the previous step. Considering the following formula (Chuang, Chen & Norvick, 1981), α and β are the intercept (zero when converted to z scores) and slope (validity) and y is the performance ability (the criterion z score):

$$y = \alpha + \beta (\textit{cut - off score})$$

such that:

$$\textit{cut - off score} = \frac{y}{\beta}$$

The cut-off scores, were then converted back to test scores and the SEMs calculated in the TSD-PI Manual (Darr, 2016) were applied to account for any measurement error that the structure of the subscales or timing of test administration may have caused. Finally, the resulting test scores were compared against the TSD-PI scores achieved by the applicant sample and the established normative data (Darr, 2016) to determine the impact that the recommended cut-off scores would have on the applicant population. The resulting cut-off scores for populations with significant results are summarized in Table 13 for the conscientiousness subscale and Table 14 for emotional stability. Although scores from all of the examined thresholds are presented in the table, only cut-scores corresponding to the average (50%) and slightly below average (40%) thresholds are recommended because the remainder resulted in minimal impact to the applicant population. These two options allow for adjustments according to the organization's recruitment requirements.

Table 13
 Conscientiousness Cut-Off Scores Based on Regression Analyses of Performance Outcome Criteria (Incumbent Sample)

Population	Threshold	y	β	Standardized Cut-Off (y/β)	Mean	SD	Test Score Cut-Off	Recommended Cut-Off (SEM Applied)	Comparison to Applicant Dataset (# Cut / 877)	Comparison to Applicant Dataset (%ile)	Comparison to Normative Data (%ile)
Officers	Non-Significant Results				84.99	10.61					
	20%	-0.84		-6.36			22.72	17.92	0	0.00	NA
	25%	-0.67		-5.08			35.50	30.70	0	0.00	NA
NCMs	30%	-0.52	0.132	-3.94	85.85	9.92	46.77	41.97	1	0.11	NA
	35%	-0.39		-2.95			56.54	51.74	3	0.34	1
	40%	-0.25		-1.89			67.06	62.26	27	3.08	5
	50%	0.00		0.00			85.85	81.05	352	40.14	40
Families:											
Administrative	Non-Significant Results				88.97	9.5					
	20%	-0.84		-5.28			33.68	28.88	0	0.00	NA
	25%	-0.67		-4.21			44.11	39.31	1	0.11	NA
General	30%	-0.52	0.159	-3.27	85.19	9.75	53.30	48.50	3	0.34	NA
	Military	35%	-0.39		-2.45		61.27	56.47	10	1.14	2
	40%	-0.25		-1.57			69.86	65.06	49	5.59	7
	50%	0.00		0.00			85.19	80.39	322	36.72	37
Mechanical	20%	-0.84		-3.10			55.38	50.58	3	0.34	1
	25%	-0.67		-2.47			61.42	56.62	10	1.14	2
	30%	-0.52	0.271	-1.92	85.23	9.63	66.75	61.95	21	2.39	5
	35%	-0.39		-1.44			71.37	66.57	55	6.27	10
	40%	-0.25		-0.92			76.35	71.55	95	10.83	17
	50%	0.00		0.00			85.23	80.43	322	36.72	37
Operator	Non-Significant Results				85.82	9.95					
Technical	Non-Significant Results				85.52	10.38					

Note: y = criterion z score, Test Score Cut-Off = converted from standardized cut-off ($SD * Std\ Cut + Mean$), $SEM = 4.8$, NA = Recommended Cut-Off is below scores achieved by the normative population, %ile = Percentile, Normative Data = English Applicant Normative Population.

Table 14
Emotional Stability Cut-Off Scores Based on Regression Analyses of Performance Outcome Criteria (Incumbent Sample)

Population	Threshold	y	β	Standardized Cut-Off (y/ β)	Mean	SD	Test Score Cut-Off	Recommended Cut-Off (SEM Applied)	Comparison to Applicant Dataset (# Cut / 877)	Comparison to Applicant Dataset (%ile)	Comparison to English Normative Data (%ile)	
Officers	Non-Significant Results				84.92	10.64						
	20%	-0.84		-4.44			30.58	24.92	11.5	0.11	NA	
	25%	-0.67		-3.54			41.40	35.74	1	0.11	NA	
	30%	-0.52	0.189	-2.75			50.95	45.29	6	0.68	1	
	35%	-0.39		-2.06	84.05	12.03	59.23	53.57	14	1.60	4	
NCMs	40%	-0.25		-1.32			68.14	62.48	49	5.59	10	
	50%	0.00		0.00			84.05	78.39	275	31.36	38	
	Families:											
	Administrative	20%	-0.84		-4.33			27.82	22.16	1	0.11	NA
		25%	-0.67		-3.45			39.03	33.37	1	0.11	NA
30%		-0.52	0.194	-2.68	83.24	12.8	48.93	43.27	4	0.46	1	
35%		-0.39		-2.01			57.51	51.85	12	1.37	2	
40%		-0.25		-1.29			66.75	61.09	44	5.02	9	
General Military	50%	0.00		0.00			83.24	77.58	250	28.51	36	
	Non-Significant Results				85.81	11.13						
	20%	-0.84		-5.38			19.64	13.98	1	0.11	NA	
	25%	-0.67		-4.29			32.53	26.87	1	0.11	NA	
	Mechanical	30%	-0.52	0.156	-3.33	83.34	11.83	43.91	38.25	2	0.23	NA
35%		-0.39		-2.50			53.77	48.11	9	1.03	2	
40%		-0.25		-1.60			64.38	58.72	30	3.42	7	
Operator		50%	0.00		0.00			83.34	77.68	250	28.51	36
		Non-Significant Results				84.2	11.17					
	20%	-0.84		-2.62			49.83	44.17	4	0.46	1	
	25%	-0.67		-2.09			56.57	50.91	10	1.14	3	
	Technical	30%	-0.52	0.321	-1.62	83.12	12.72	62.51	56.85	22	2.51	6
35%		-0.39		-1.21			67.67	62.01	49	5.59	10	
40%		-0.25		-0.78			73.21	67.55	96	10.95	18	
Technical		50%	0.00		0.00			83.12	77.46	250	28.51	36

Note: y = criterion z score, Test Score Cut-Off = converted from standardized cut-off (SD * Std Cut + Mean), SEM = 5.66, NA = Recommended Cut-Off is below scores achieved by the normative population, %ile = Percentile, Normative Data = English Applicant Normative Population.

Recommending Norm-Referenced Cut-Off Scores. Where scores could not be recommended based on forward regression analyses of the samples within the provided datasets, an alternate method of recommending cut-off scores was employed. Cascio, Alexander and Barrett's (1988) norm-referenced methods allow for the creation of cut-off scores when empirical test score data exists in addition to well established normative data. Cascio et al. (1988) suggest basing norm-referenced cut-off scores on base-rate information. This involves establishing a minimum level of acceptable job performance from the present workforce and applying it to the applicant normative data to establish a cut-off score. The performance measure used in this study establishes a minimum level of acceptable job performance at a mean score of 4 ("skilled"), excluding employees whose performance scores are considered "unacceptable", "needs improvement" or "developing". As the performance measure is a composite created from the mean score on 16 performance criteria, mean scores between 3.5 and 4.49 have been considered to have met the minimum mean of 4. In order to determine cut-off scores from this minimum level of acceptable job performance, the predictor variable test scores of those achieving this established base-rate are observed. Therefore, the mean conscientiousness and emotional stability scores for those achieving the minimum level of acceptable performance were calculated. These scores subsequently had the SEM applied to them and were compared to the normative applicant population. This process was used, for each of the populations whose cut-off scores could not be set through regression analyses. This includes conscientiousness scores for the officer population and the administrative, operator and technical family groups as well as emotional stability cut-off

scores for the officer population and the general military and operator occupational family groups. The results of this process are summarized in Table 15.

Table 15
Norm-Referenced Cut-Off Scores Based on a Minimum Level of Acceptable Job Performance (Incumbent Sample)

Population	TSD-PI Subtest	Mean	Recommended Cut-Off (Mean with SEM applied)	Percentile (Normative Population)
Officers (<i>N</i> = 31)	Conscientiousness	85.97	81.17	40
	Emotional Stability	85.16	79.50	41
Administrative Family (<i>N</i> = 28)	Conscientiousness	84.43	79.63	34
General Military Family (<i>N</i> = 51)	Emotional Stability	82.12	76.46	33
Operator Family (<i>N</i> = 51)	Conscientiousness	84.14	79.34	34
	Emotional Stability	83.67	78.01	38
Technical (<i>N</i> = 92)	Conscientiousness	82.92	78.12	31

Note: Mean corresponds to the mean TSD-PI subscale scores achieved by those who just meet the minimum acceptable level of performance on the job, conscientiousness SEM = 4.8, emotional stability SEM = 5.66, all recommended cut-offs were rounded down for comparison to the normative population.

Summary of Recommended Cut-Off Scores. Table 16 provides a summary of all recommended cut-off scores, the method in which they were set and their percentile comparison to the English normative applicant population.

Table 16
Summary of Recommended Cut-Off Scores for the TSD-PI

Population	TSD-PI Subscale	Method Set	Recommended Cut-Off	Percentile (Normative Population)
Officers	Conscientiousness	Norm-referenced	81	40
	Emotional Stability	Norm-referenced	79	41
NCMs	Conscientiousness	Criterion-related	50% - 81 40% - 62	40 5
	Emotional Stability	Criterion-related	50% - 78 40% - 62	38 10
Occupational Families				
Administrative	Conscientiousness	Norm-referenced	79	34
	Emotional Stability	Criterion-related	50% - 77 40% - 61	36 9
General Military	Conscientiousness	Criterion-related	50% - 80 40% - 65	37 7
	Emotional Stability	Norm-referenced	76	33
Mechanical	Conscientiousness	Criterion-related	50% - 80 40% - 71	37 17
	Emotional Stability	Criterion-related	50% - 77 40% - 58	36 7
Operator	Conscientiousness	Norm-referenced	79	34
	Emotional Stability	Norm-referenced	78	38
Technical	Conscientiousness	Norm-referenced	78	31
	Emotional Stability	Criterion-related	50% - 77 40% - 67	36 18

Note: Recommended cut-off scores have been rounded down to the nearest whole number as test scores on the TSD-PI are measured in whole numbers. Cut-off scores for the NCM population would be used for applicants to occupations that do not belong to one of the occupational family groups. Where the criterion-related method was used, cut-scores corresponding to the average (50%) and slightly below average (40%) thresholds are recommended.

Discussion

The purpose of this research was to establish cut-off scores on the conscientiousness and emotional stability subscales of the TSD-PI for the CAF applicant population. Two job-related outcome criteria were provided in the datasets. Unfortunately, the training data contained in the samples were not predicted by either conscientiousness or emotional stability. The negligible relationships realized between the predictor and training outcome variables are believed to have resulted from issues concerning the data itself. The dichotomous nature of the criterion variable, and the fact that comparatively few trainees were assessed as not having met the standard in each of the samples, produced validities that were inconsistent with previous research (Darr, 2011) and resulted in small sample sizes. As a result, the entire dataset was discarded causing the loss of training data as outcome criteria with which to establish cut-off scores. Therefore, all recommended cut-off scores are solely based on the subscales' ability to predict job performance.

With job performance as the focal outcome variable, standardized regression values were found to be significant within some of the samples observed, which allowed for criterion-related cut-off scores to be recommended for the 50% and 40% thresholds for selection purposes. This included both conscientiousness and emotional stability cut-off scores for the NCM group and mechanical family group, as well as a conscientiousness cut-off for the general military family and emotional stability cut-off scores for the administrative and technical family groups. Neither conscientiousness nor emotional stability were significant predictors of the operator or officer groups. Where non-significant relationships existed within the samples provided for this study, norm-

referenced techniques were employed in order to recommend cut-off scores (see Table 16).

As with the training outcome criteria, the officer sample containing job performance data produced unexpected results. In fact, the personality measures accounted for 0% of the variance for this group. This finding is contrary to previous research, which suggested that a stronger relationship would exist between the predictor variables and officer job performance. Research supports links between educational criteria and the conscientiousness factor of personality (Chamorro-Premuzic & Furnham, 2003; Paunonen & Ashton, 2001; Major et al., 2006). Evidence also exists for the predictive validity of both subscales with respect to leadership (Dragow, Stark, Chernyshenko, Nye, Hulin, & White, 2012) and managerial performance (Rothmann & Coetzer, 2013). Despite the higher education selection criteria required of officer applicants and the leadership abilities expected of junior officers, the relationships between the personality factors and officer performance did not exist in this sample. Similar to the samples containing training outcome information, these findings may also be attributed to limitations with the data. A comparatively small sample was available for the officer group (contrary to the numbers provided for the NCM sample). Of the 181 officers whose data were observed, only 13 were assessed as not meeting the job performance standard expected. Unsuccessful incumbents comprised only 7.2% of the entire sample. Where this may point to an issue with range restriction on the performance measure, as discussed in the limitations section of this paper, it could also suggest that the sample used in this study is not representative of the overall officer applicant population. Further analyses of the sample were conducted revealing that it was comprised of

members from 25 different officer occupations however, 34% of the participants were employed in one of three jobs; artillery officer, infantry officer or pilot, all of which are highly operational in nature.

Although the nature of the data is likely a significant factor for those results found to be non-significant in this study, theoretical explanations could be provided by TAT. First, only paths 1 and 6 (see Figure 1) of Tett and Burnett's (2003) model of job performance were observed within this research. Sufficient data was not available to assess the specific situational cues (job demands) or the internal and external motivators that moderated the relationships between personality and performance. In addition, Tett and Burnett's (2003) model argues for the presence of additional situational features, further to the job demands discussed earlier in this work, that impact whether personality is expressed as valued on-the-job behaviour. These features include distracters, constraints, releasers and facilitators that operate at the task, social and organizational levels to determine whether a person will react as desired and, in accordance with their relevant personality traits in a given work situation. Distractors, if responded to, tend to interfere with performance. For example, an officer who is high on the conscientiousness scale may be distracted by the requirement to have a detailed plan, to the point that they are unable to make quick and decisive decisions when reacting to changing situations on the job. On the other hand, a constraint negates the effect that a personality trait will have on a desired behaviour (e.g., an attack by enemy forces prompts quick action preventing a junior officer from establishing a plan that considers all aspects of the situation). Releasers are work-related events that counteract constraints (e.g., the weather caused the retreat of enemy forces, allowing the officer time to come up with a detailed plan of

attack). Finally, facilitators increase how noticeable existing trait relevant cues are in a given situation. An example applicable to our highly conscientious junior officer could be his former training and the emphasis that was placed on proper operational planning prior to field assessments leading to his deployment. This would increase the officer's awareness of the requirement to appropriately plan, thus bolstering the need to be highly conscientious. Tett and Burnett (2003) explain that these situational features may result in what is considered a positive trait (e.g., high levels of conscientiousness) turning into behaviour that may be evaluated as undesired in certain work situations. This could be one explanation for the negative or negligible relationships that existed for some of the samples in this study between the predictor and the outcome criteria. Unfortunately, these situational features could not be measured within the constraints of this study.

Despite the practical constraints evident in this applied research, it offers an opportunity to lessen the divide that exists between research and practice in personnel selection (Anderson, 2005; Anderson, Herriot & Hodgkinson, 2001; Dunnette, 1990, Hodgkinson, Herriot & Anderson, 2001). It is an attempt at what Anderson (2005) has labeled Pragmatic Science, practically relevant applied research. It is not an additional example of a theory-oriented investigation to answer an academic question of interest, which Briner and Rousseau (2011) state are increasingly common in the field. That being said, Briner and Rousseau (2011) advocate for evidence-based industrial-organizational psychology, which implores practitioners to make decisions based on four sources of information. These include practitioner expertise and judgement, evidence from the local context, critical evaluation of the best available research and perspectives of those who may be affected by the decision (Briner & Rousseau, 2011). From a researcher's

perspective, this intensifies the importance of acknowledging the limitations of our knowledge where evidence is mixed, causing grounds for uncertainty (Briner & Rouseau, 2011). Applied to the CAF's decision to implement TSD-PI cut-off scores for the purpose of selection, the role of this work may serve to suggest that additional data must be collected for research purposes. This would allow for a more complete picture of how personality may or may not be predictive of training and performance outcome criteria in the population from a research perspective. The organization then must take into consideration the extensive knowledge and experience of selection practitioners in the CAF, perspectives of the selection officers and military career counsellors whose jobs will be affected and evidence from the local context (i.e., how it will impact the normative population), prior to establishing TSD-PI cut-off scores. In other words, this research is just one small piece of information which should inform the overall decision with respect to the implementation of future changes to the selection system.

Limitations

As with all research, there are limitations that have resulted from the design of this study, which must be highlighted for the CAF for the purpose of evidence-based decision making. The measures used to establish a participant's level of performance on the job and during training limited the type of analyses that could be conducted. In addition, as discussed, the samples provided by the organization impacted generalizability and may have been confounded due to overlapping analyses. Each of these aspects are elucidated in more detail below.

Measures. When deliberating on the value of the outcome criteria in the establishment of cut-off scores, the limitations of each measure must be taken into

consideration. With respect to the training outcome data, the most significant limitation existed with the pass/fail information provided. If grade or percentage information were available for a larger number of participants, more variability in the range of scores may have allowed for significance to be found between the predictor variables and outcome criteria. Unfortunately, the pass/fail data did not permit the observation of relationships between the variables, if they existed. Percentage information may have allowed for a better determination of the predictive validity of the conscientiousness and emotional stability subscales with respect to training outcomes. Regrettably, results from the analysis of this training data could not be used to support the use of conscientiousness and emotional stability as predictors of training success, let alone form the basis for the establishment of cut-off scores for the subscales. The pass/fail data are only one possible explanation for the insignificant relationships that existed with the conscientiousness and emotional stability subscales. Another conceivable rationalization, which is consistent with the majority of personality research, is that neither subscale is a reliable predictor of training success (Darr & Kelloway, 2016). As mentioned earlier in the discussion section, this inconsistency in personality research could be related to unaccounted for situational features (Tett & Burnett, 2003).

The job performance information provided through the PER process also has its limitations. Although Canadian Forces Personnel Appraisal System (CFPAS) guidelines are provided for completing the reports, there is likely a large amount of inconsistency in the assignment of scores between the raters. This is simply because no inter-rater reliability training is delivered to supervisors who are obliged to complete evaluations of their subordinates. This leaves it up to the supervisor to interpret the criteria and assign a

score that they feel reflects their subordinate's performance over the past year. In addition, each element of the PER has a different standard that must be met to achieve a certain score, depending on the rank level of the subordinate. The standard for higher ranking subordinates is set above that expected for a lower ranking employee. For example, the "mastered" score for the supervision criterion may require a subordinate who holds the rank of Chief Warrant Officer to effectively supervise an entire battalion, whereas a "mastered" rating can be achieved by a Master Corporal responsible for a section of five of his peers or less. Therefore, the ratings are not only subjective depending on the supervisor's interpretation of the guidelines provided, but are relative to the rank and position held by the subordinate. The data on this variable also suggested that, although each item is meant to measure a specific performance criterion, the PER only assesses one overarching performance factor. This indicates that supervisors may not be rating the individual items as the measure was intended and may even be rating a construct unrelated to performance (such as seniority or popularity). The PER is the single most influential consideration used to make promotion decisions in the CAF. The data were restricted in range and the mean was shifted to the higher end of the scale, with substantially more members' performance being assessed as "mastered" than "unacceptable". This could be a consequence of supervisors' inflation of scores, a problem that is inherent in the PER system, as right-aligned assessments ("mastered" on all items) are required for promotion in many occupations. Not assessing an employee deserving of promotion as "mastered" is likely to adversely impact them in promotion boards, where appraisals are made in contrast to their peers for promotion decisions. Therefore, the PER may be a measure of who is most likely to be promoted (regardless of

supervisors' motivation), rather than an accurate measure of members' performance.

This range restriction resulting from the extremely high number of outstanding evaluations was particularly evident in the officer samples. The inflation of performance scores is not an issue that is isolated to the PER. Jawahar and Williams (1998) found that when performance appraisals were obtained for administrative purposes, such as promotions or pay raises, supervisors are more lenient with the ratings compared to assessments collected for the purpose of research, employee development or feedback (i.e., one third of a standard deviation higher). This bias may occur in attempt to avoid providing negative feedback (Fisher, 1989), to obtain positive consequences (e.g., a promotion), avoid negative consequences resulting from harsh but truthful ratings (e.g., no promotion) or to motivate an employee who is not performing up to the standard expected (Murphy & Cleveland, 1991). Whereas, ratings collected only for the purpose of research may encourage accurate performance evaluations due to promises of confidentiality (Jawahar & Williams, 1998). It is for this reason that in their revised principles for the validation and use of personnel selection procedures, the Society for Industrial and Organizational Psychology (SIOP; under review) recommends using performance ratings that are collected for research purposes rather than administrative functions.

The TSD-PI subscales also suffered from range restriction. Numerous applicants scored the highest possible score (105) on the subscales, however scores of 58 on the emotional stability subscale and 51 on the emotional stability subscale were considered outliers. The sample data were compared to the normative applicant population and it was observed that although both sets of data were shifted to the higher end of possible

scores, the sample data appeared to be chopped off on the right side of the normal distribution curve, whereas the normative population decreases to form a tail as expected. In addition, the sample data lacked an extended tail on the lower end of the scales as well. This difference between the study data and normative population can be explained by the participants within. Where the applicant normative population contained TSD-PI results of all applicants to the CAF, including those who were selected out, the sample data only contained participants who made it through the selection system, as their job and training outcome data were a requirement for inclusion in this study. The people found not suitable or competitive for employment in the CAF may have had their overall selection scores impacted by low TSD-PI scores on the conscientiousness and emotional stability subscales. This may have had an impact on the study results because it is unknown if those selected out would have successfully performed in the occupations that they applied for. It may also have shifted the mean scores of this sample to the right compared to the applicant normative population.

Samples. In addition to the previously discussed differences between the study data and applicant normative population on the TSD-PI scores, there were other limitations with the sample that may have impacted this study. All of the samples contained a comparatively small number of training failures, or people who were assessed as not meeting the minimum acceptable level of performance. As such, the NCM and officer samples could only be assessed on one of the two training outcomes and neither resulted in a contribution to the establishment of cut-off scores. Of the samples that could be used, the small number of unsuccessful participants may have influenced the findings, particularly for the officer group. A comparison of the number of participants in the

‘unsuccessful performance category’ for the NCM and officer incumbent samples shows that there are only 23% and 7%, respectively. This suggests a constrained or restricted range of scores. The low number of failures likely contributed to non-significant findings for these variables. Validities for the performance outcome may also have been stronger if there had been a larger number of job incumbents whose performance was considered unacceptable. It is possible that with sufficient numbers, criterion-related procedures could have been used for the populations that relied on norm-referenced techniques for the establishment of cut-off scores.

With respect to generalizability, this study only contained performance data from the Anglophone CAF population, which formed the basis for the recommended cut-off scores. Despite language equivalence between the English and French versions of the TSD-PI (Darr & Kemp, 2011), separate normative data exists for the Anglophone and Francophone populations. Therefore, the results of this study should only be applied to the Anglophone applicant population. Although there was a substantial gender imbalance in this study due to the large population of male applicants, it is not seen as a limitation to generalizability because it is representative of the serving CAF population, and the applicant population for which this study was conducted. In addition, it is expected that the age distribution was negatively skewed, with most candidates applying prior to their 30th year. This is also representative of the typical population of applicants to the CAF and not considered a limitation of this study.

Finally, results relating to the occupational family groups may have been confounded with the overall NCM sample because the family group samples were formed from the same dataset. Therefore, all of the NCMs included in the family groups

also formed a part of the overall NCM sample. Although there are additional occupations included in the overall NCM sample that do not fall within one of the occupational families, it is possible that this overlap may have resulted in the findings from the family groups being a reflection of the overall NCM sample from which they came. It is recommended that this study be repeated with independent samples to rule out the potential of confounded results.

Implications

This research met the requirement of the Canadian Forces Recruiting Group (CFRG; DGMPRA, 2016) by recommending a separate set of cut-off scores for officers and NCMs on the subscales of the TSD-PI, while presenting both the theoretical and empirical justification for doing so. If the organization makes the evidence-based decision to adopt cut-off scores for the conscientiousness and emotional stability subscales, it will require applicants to meet the minimum raw score associated with the appropriate applicant group, in addition to meeting the already established cut-off scores for the CFAT. The scores of both measures would then be combined with the results of the structured interview using the existing compensatory model to determine the applicant's potential to succeed in the CAF and their competitiveness, as compared to other applicants. Selection decisions would be made based on these scores, in accordance with status quo. With only one additional step (verification that the cut-off scores on the TSD-PI have been met), additional validity is added to the selection model to allow for better predictions of applicants' future performance. Despite insignificant results for the training outcomes in this study, the cut-off scores are also likely to save the organization

training costs and resources associated with occupational reassignment for those members who would make it through training but be unsuccessful on the job.

In addition to the enhanced predictive validity of the selection system, the cut-off scores provide a legally defensible method of narrowing down the applicant pool (Kehoe & Olson, 2005) prior to the structured interview. This has the potential to reduce the organizational resources required, both financial and human, in the selection of applicants to the CAF. In reducing the workload of the Military Career Counsellor, by decreasing the number of interviews required, applicants also benefit through a reduction in the amount of time between initial application to enrolment. This addresses one of the Canadian Forces Recruiting Group's current priorities. This study also presented alternate cut-off scores (the 40% threshold) for many of the applicant populations which gives the CAF the ability to adjust the minimum required score based on the needs of the organization (e.g. to allow more or less people through to subsequent phases of selection depending on the number of positions available).

Further to the conservation of selection resources, implementation of the cut-off scores would allow for the elimination of the group of applicants who is most likely to engage in faking, the intentional inflation test scores. Darr and Kemp (2011) compared the results of the conscientiousness and emotional stability subscales for trainees who were administered the TSD-PI for research purposes during their basic training (Darr & Bergevin, 2009) to an applicant sample. For the conscientiousness subscale, applicant scores were higher in the group that placed below the 50th percentile on the normative data compared to the trainee sample. For emotional stability, score differences were greater in the applicant sample for the group scoring in the 25th percentile. The

researchers concluded that applicants truly higher on the subscales tended to respond to the self-report scale honestly, whereas those at the lower end of the continuums were more liable to attempt to increase their chances of being selected. Therefore, the use of the recommended cut-off scores, would exclude those who may have engaged in faking.

An unexpected implication that resulted from this study stemmed from the insignificant regression analyses effects for the officer sample. These results suggest issues relating to the use of the personality subscales as selection criteria to predict future officer performance, or problems with the PER as a performance measure for this group. Areas for future research regarding this finding are discussed below. These suggestions could benefit both the selection system and performance measurement for the provision of feedback and informing of promotion decisions.

The differential findings that resulted in distinct cut-off scores for NCM family groups further substantiates the categorization and supports their continued use for validation studies and the establishment of cut-off scores for selection purposes. The organization of the five family groupings originally identified by Catano and Ibel (1995) have been altered slightly over time due to changes to the individual job descriptions (e.g. some general military occupations have become more technical in nature), however the main headings still exist. They provide an acceptable alternative when it may not be feasible to analyze each occupation individually due to the amount of time required to collect data and the resources necessary to conduct the research.

Future Research Directions

Prior to implementing cut-off scores, the most imperative research must result in the recommendation of TSD-PI cut-off scores for the French language applicant

population. Larger datasets will be required before this can be completed. Due to the already established language equivalencies between the French and English language versions of the TSD-PI (Darr & Kemp, 2011), it is expected that the recommended raw cut-off scores for the French population will be very similar to those found within this study. The normative data for the conscientiousness subscale has the 50th percentile corresponding to a raw score of 84 for both the English and French language normative populations. Therefore, it is expected that equal raw cut-off scores will also have a similar impact to the French and English language applicant populations. The emotional stability score corresponding to the 50th percentile for the English language normative data is 82, whereas the French language population is 84. Therefore, equivalent raw score cut-offs for the two language groups, would result in the elimination of a slightly higher percentage of English language applicants than French language applicants.

It is recommended that this study be repeated with alternate measures of performance, as suggested by SIOP (under review). As discussed earlier, because the PER is also used for promotion decisions, the system has resulted in inflated scores which may no longer provide accurate assessments of individuals' performance on the job. A combination of performance evaluations collected for research purposes may serve as a better measurement of this outcome criteria. Self-assessments can be combined with peer rated measures and supervisor evaluations that do not have career implications (i.e. they are used for research only). It is also recommended that different aspects of performance be measured. Factor analysis showed that the PER measured only one overall performance factor in this study's sample, however research suggests that there are distinct aspects of performance. Campbell, McCloy, Oppler, and Sager's (1993) model of

job performance includes eight factors: job-specific task proficiency, non-job-specific task proficiency, written and oral communication task proficiency, demonstrating effort, maintaining personal discipline, facilitating peer and team performance, supervision– leadership, and management– administration. Many of these aspects are evaluated using single items on the PER, however they may have been confounded for this sample by how the measure is employed. A combination of misinterpreted intentions on the proper use of the measure and supervisors’ desire to do well by their subordinates could mean that, as it is currently being used, the PER is not a realistic measure of actual job performance, it is inaccurate or alone may be of limited value for research purposes. It is possible that, if these factors can be extricated, different aspects may be more significant for various CAF populations. Of particular interest, is whether the TSD-PI subscales are predictive of these performance aspects for the officer population. An additional benefit of collecting performance data through alternate measures is that it would allow for the validation of the PER as it is currently being used. This is advantageous to the organization beyond selection as the performance evaluation tool would need to be reviewed for utility in its intended role of providing feedback and informing promotion decisions.

Future studies should use samples that include a larger number of training failures and members whose performance is evaluated as unsatisfactory. The few numbers incorporated in this study may have resulted in lower validity coefficients. Independent samples should be created for the NCMs and family groups to avoid potential confounding. It is also suggested that the study be repeated to control for variation in performance that is predicted by other hurdles in the selection system. Specifically,

CFAT data should be added as a first step in the regression analyses to find validity coefficients for the TSD subscales after controlling for cognitive ability. These new validities may make it possible for significance to be found with the training outcome variables and could alter the figures used to determine cut-off scores in this study with the performance outcome. Where data permits, research should be completed at the occupation level for jobs that fall outside the occupational families and those contained within to establish more precise baseline scores as valid predictors of performance and occupational training outcomes. These individual occupational analyses would also be better suited to take into consideration the unique situational features that may be missed when lumping various occupations into large samples for analyses. Finally, the potential for the existence of a curvilinear relationship with outcome criteria should be explored for both predictor variables, which may help to explain the non-significant results and small, but negative, relationships.

Conclusion

The establishment of cut-off scores for the conscientiousness and emotional stability subscales of the TSD-PI has the potential to reduce the amount of time and resources it takes for a candidate to progress from initial application to enrolment into the CAF. In addition, there exists the possibility to save in future training expenditures that would be dedicated to candidates that make it through the selection process and who, subsequently, do not succeed in their occupations. Cut-off scores have been recommended within this study for the general officer core, the general NCM core and the five NCM occupational families. It is recommended that the results provided in this study be used to inform the evidence-based decision-making process in determining if

these cut-off scores are adopted as a complementary aspect of the CAF selection procedure. The goal of every organization is to select the most appropriate people for the available positions while maintaining operational and resource efficiency and effectiveness; the potential application of cut-off scores to the subscales of the described TSD-PI measure is one mechanism to continue to strive towards this objective.

References

- Anderson, N. (2005). Relationships between practice and research in personnel selection: Does the left hand know what the right is doing? In A. Evers, N. Anderson., & O. Voskujl. (Eds.), *Handbook of Personnel Selection* (pp. 1- 24). Boston, MA: Boston Publishing.
- Anderson, N., Herriot, P., & Hodgkinson, G.P. (2001). The practitioner-researcher divide in Industrial, Work and Organizational (IWO) psychology: Where are we now, and where do we go from here? *Journal of Occupational and Organizational Psychology*, 74, 391-411.
- Anderson, G. & Viswesvaran, C. (1998). An update of the validity of personality scales in personnel selection: A meta-analysis of studies published after 1992. Paper presented at the 13th Annual Conference of the Society of Industrial and Organizational Psychology, Dallas.
- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Barrick, M. R. & Mount, M. K., (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-24.
- 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.
- Barrick, M. R., & Ryan, A. M. (2003). *Personality and work: Reconsidering the role of personality in organizations*. San Francisco: Jossey-Bass.

- Berry, C. M., & Sackett, P. R. (2009). Faking in personnel selection: Tradeoffs in performance versus fairness resulting from two cut-score strategies. *Personnel Psychology, 62*(4), 833-863.
- Bott, J. P., O'Connell, M. S., Ramakrishnan, M., & Doverspike, D. (2007). Practical limitations in making decisions regarding the distribution of applicant personality test scores based on incumbent data. *Journal of Business and Psychology, 22*(2), 123-134.
- Bowers, K. S. (1973). Situationism in psychology: An analysis and a critique. *Psychological Review, 80*, 307-336.
- Boyes, F. A. J. (2005). Personality as a predictor of military performance and counterproductive behaviour. Unpublished master's thesis. Saint Mary's University, Halifax, Nova Scotia, Canada.
- Boyes, F. A. J. (2006). The Trait Self Descriptive Inventory: An examination of the psychometric properties and exploration of an abbreviated version [Director Human Resources Research and Evaluation Technical Note 2006-02]. Defence Research and Development Canada, Ottawa, ON.
- Briner, R. B., & Rousseau, D. M. (2011). Evidence-based I-O psychology: Not there yet. *Industrial and Organizational Psychology, 4*, 3-22.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance: In N. Schmitt & WC Borman (Eds.), *Personnel Selection in Organizations* (pp. 35-70).
- Cascio, W. F., Alexander, R. A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology, 41*(1),

1-24.

- Catano, V.M. & Ibel, R.H. (1995). *Clustering Canadian Forces military occupations using statistical analyses of ability profiles*. [Canadian Forces Personnel Applied Research Unit Technical Report 6/95]. Defence Research and Development Canada, Ottawa, ON.
- Cattell, E. E. P. & Mead, A. D. (2008). The Sixteen Personality Factor Questionnaire (16PF). In G. J. Boyle, G. Mathews, & D. H. Saklofske (Eds.), *The SAGE Handbook of Personality Theory and Assessment* (pp. 135–159). Retrieved from <http://www.central.rcs.k12.tn.us/Teachers/cowartg/documents/16PFCattell.pdf>
- Chamorro-Premuzic, T. & Furnham, A. (2003). Personality traits and academic examination performance. *European Journal of Personality*, 17, 237–250.
- Chatman, J. A. (1989). Improving interactional organizational research: A model of person–organization fit. *Academy of Management Review*, 14, 333–349.
- Christal, R (1994). *The air force self descriptive inventory*. Final R&D Status Report: United States Air Force.
- Chuang, D. T., Chen, J. J., & Novick, M. R. (1981). Theory and practice for the use of cut-scores for personnel decisions. *Journal of Educational Statistics*, 6(2), 129-152.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155-159.
- Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 19(1), 15-18.
- Costa, P. T. & McCrae, R. R. (1994). Revised NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI): Professional Manual. Florida: Psychological Assessment Resources, Inc.

- Costa, P.T. Jr & McCrae, R.R. (1997). Longitudinal study of adult personality. In R. Hogan, J. Johnson, S. Briggs (Eds.), *Handbook of Personality Psychology*. San Diego: Academic.
- Darr, W. A. (2011). Military Personality Research: A Meta-Analysis of the Self Description Inventory. *Military Psychology*, 23(3), 272–296. doi: 10.1080/08995605.2011.570583
- Darr, W. (2013). The Trait Self Descriptive Personality Inventory [Director General Military Personnel Research and Analysis Briefing Note]. Defence Research and Development Canada, Ottawa, ON.
- Darr, W. (2016). Trait Self Descriptive Personality Inventory (TSD-PI) Manual: Revised 2015 [Director General Military Personnel Research and Analysis Scientific Report 2016-R064]. Defence Research and Development Canada, Ottawa, ON.
- Darr, W. A. & Bergevin, J. (2009). A comparison of the French and English versions of the Trait Self Descriptive (TSD) Personality Inventory [Director General Military Personnel Research and Analysis Technical Memorandum 2009-014]. Defence Research and Development Canada, Ottawa, ON.
- Darr, W., & Kelloway, E. K. (2016). Sifting the Big Five: examining the criterion-related validity of facets. *Journal of Organizational Effectiveness: People and Performance*, 3(1), 2-22.
- Darr, W. A. & Kemp, C. (2011). A psychometric examination of the e-version of the Trait Self Descriptive Personality Inventory (TSD-PI) [Director General Military Personnel Research and Analysis Technical Memorandum 2011-010]. Defence Research and Development Canada, Ottawa, ON.

Director General Military Personnel Research and Analysis. (2016). *Task/Project and Project Work Plan Agreement: Establishing cut-off scores for the Trait Self Descriptive Personality Inventory (TSD-PI)*. [Collaborative Research Agreement 2014003, Task 2016002]. Defence Research and Development Canada, Ottawa, ON.

Dragow, F., Stark, S., Chernyshenko, O. S., Nye, C. D., Hulin, C. L., & White, L. A. (2012). Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support army selection and classification decisions (Technical Report 1311). *Ft. Belvoir, VA: U.S. Army Research Institute*.

Dunnette, M. D. (1990). Blending the science and practice of industrial and organizational psychology: Where are we now and where are we going? In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed., Vol. 1, pp. 1-37). Palo Alto, CA: Consulting Psychologists Press.

Ebel, R. L. (1972). *Essentials of educational measurement* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Ekehammar, B. (1974). Interactionism in personality from a historical perspective. *Psychological Bulletin, 81*, 1026–1048.

Endler, N. S., & Magnusson, D. (1976). Toward an interactional psychology of personality. *Psychological Bulletin, 83*, 956–974.

Epstein, S., & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin, 98*, 513–537.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: SAGE.

- Fisher CD. (1989). Current and recurrent challenges in HRM. *Journal of Management*, 15, 157- 180.
- Gulliksen, H. (1987). *Theory of mental tests*. Mahwah, NJ: Erlbaum. (Original work published 1950).
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, 48, 26-34.
- Harvill, L. M. (1991). Standard error of measurement. *Educational Measurement: issues and practice*, 10(2), 33-41.
- Hodgkinson, G. P., Herriot, P., & Anderson, N. (2001). Re-aligning the stakeholders in management research: Lessons from Industrial, Work and Organizational Psychology. *British Journal of Management*, 12, 41-48.
- Hogan, R. T. & Roberts, B. W. (2001). Introduction: Personality and Industrial and Organizational Psychology. In B. Roberts, & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 3-16). Washington: American Psychological Association.
- Hough, L. M (1992). The 'Big Five' personality variables – construct confusion: Description versus prediction. *Human Performance*, 5, 193-155.
- Jaeger, R. M. (1982). An iterative structured judgment process for establishing standards on competency tests: Theory and application. *Educational Evaluation and Policy Analysis*, 4(4), 461-475.
- Jawahar, J., & Williams, C. R. (1998). Where all the children are above average: A meta analysis of the performance appraisal purpose affect. *Scholarship and professional*

work from the College of Business, 2.

- Jeanneret, P. R., & Stelly, D. J. (2003). Setting cutting scores using regression: Would you do it backwards? Master Tutorial presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Kehoe, J. F., & Olson, A. (2005). Cut scores and employment discrimination litigation. In F.J. Landy (Ed.) *Employment discrimination litigation: Behavioral, quantitative, and legal perspectives* (pp. 410–449). San Francisco, CA: Jossey-Bass. I.
- Kehoe, J. F., & Tenopyr, M. L. (1994). Adjustment in assessment scores and their usage: A taxonomy and evaluation methods. *Psychological Assessment, 6*(4), 291-303.
- Kemp, C. (2010). A Rasch model analysis of the Trait Self Descriptive Personality Inventory. An investigation of scale functioning, reliability, dimensionality, and differential item functioning [Director General Military Personnel Research and Analysis Technical Report 2010-003]. Defence Research and Development Canada, Ottawa, ON.
- Klang, A. (2012). The Relationship between Personality and Job Performance in Sales: A Replication of Past Research and an Extension to a Swedish Context. *Doctoral dissertation*, Stockholm.
- Lance, C. E., Butts, M. M., & Michels, L. C. (2006). The sources of four commonly reported cutoff criteria: What did they really say? *Organizational Research Methods, 9*, 202–220.
- Lee, K., & Ashton, M. C. (2004). The HEXACO Personality Inventory: A new measure of the major dimensions of personality. *Multivariate Behavioral Research, 39*, 329–358.

- Livingston, S. A., & Zieky, M. J. (1982). *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests*. Princeton NJ: Educational Testing Service.
- Major, D. A., Turner, J. E., & Fletcher, T. D. (2006). Linking proactive personality and the Big Five to motivation to learn and development activity. *Journal of Applied Psychology, 91*, 927–935.
- Martin, S.L., & Raju, N. (1991). *Determining optimal cutoff scores in criterion-related validity studies*. Paper presented at the conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.
- Mount, M.K., Barrick, M.R., & Stewart, G.L. (1998). Five-factor model of personality and performance in jobs involving interpersonal interactions. *Human Performance, 11*, 145-165.
- Murphy KR, Cleveland JN. (1991). *Performance appraisal: An organizational perspective*. Boston: Allyn and Bacon.
- Neal, L., Fox, C., Carroll, N., Holden, M., & Barnes, P. (1997). Development and validation of a computerized screening test for personality disorders in DSM-III-R. *Acta Psychiatrica Scandinavica, 95*(4), 351-356.
- Nedelsky, L. (1954). Ability to Avoid Gross Error as a Measure of Achievement. *Educational and Psychological Measurement, 14*(3), 459-472.
- O’Keefe, D. F. (1998). Investigating the use of occupational personality measures in the Canadian Forces selection system [Director Human Resources Research and Evaluation Technical Note 98-14]. Defence Research and Development Canada,

Ottawa, ON.

Paunonen, S. V. & Ashton, M. C. (2001). Big Five factors and facets and the prediction of behavior. *Journal of Personality and Social Psychology*, 81, 524–539.

Pervin, L. A. (1985). Personality: Current controversies, issues, and directions. *Annual Review of Psychology*, 36, 83–114.

Piasentin, K. (2010). Establishing a Canadian Forces Aptitude Test (CFAT) Cut-off Score for the Army Communications and Information Systems Specialist (ACISS) Occupation [Director General Military Personnel Research and Analysis Technical Note 2010-001]. Defence Research and Development Canada, Ottawa, ON.

Rothmann, S., & Coetzer, E. (2003). The big five personality dimensions and job performance. 29(1), 68-74.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, 82(1), 30-43.

Salgado, J. F. (1998). Big five personality dimensions and job performance in army and civil operations: A European perspective. *Human Performance*, 11(2/3), 271-288.

Salgado, J. F. (2002). The Big Five personality dimensions and counterproductive behaviors. *International Journal of Selection and Assessment*, 10, 117–125.

Schneider, B. (1983). Interactional psychology and organizational behavior. In L. L. Cummings & B. M. Staw (Eds.), *Research in organizational behavior* (Vol. 5, pp. 1–31). Greenwich, CT: JAI Press.

Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40, 437–453.

Scholtz, D. C. (2004). Validation of the CFAT and Establishment of Cutoff Scores for

- Steward Selection [Director Human Resources Research and Evaluation Technical Note 2004-01]. Defence Research and Development Canada, Ottawa, ON.
- Shepard, L. (1980). Standard setting issues and methods. *Applied Psychological Measurement*, 4(4), 447-467.
- Snyder, M., & Ickes, W. (1985). Personality and social behavior. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 2, pp. 883–947). New York: Random House.
- Society for Industrial and Organizational Psychology. (Under Review). *Principles for the validation and use of personnel selection procedures* (5th ed.).
- Sundstrom, E. (1999). The challenges of supporting work team effectiveness. In E. Sundstrom & Associates. *Supporting work team effectiveness: Best management practices for fostering high team performance* (pp. 3–23). San Francisco: Jossey-Bass.
- Sutherland, S. M. (2012). *Trait Self Descriptive Inventory: The search for validity and structure*. Unpublished master's thesis Saint Mary's University, Halifax, Nova Scotia, Canada.
- Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of applied psychology*, 23(5), 565-578.
- Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88(3), 500-517.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703-742.

Tett, R. P., Simonet, D. V., Walser, B., & Brown, C. (2013). Trait activation theory:

Applications, developments, and implications for person-workplace fit. In N. D.

Christiansen & R. P. Tett (Eds.), *Handbook of personality at work* (pp. 71–100).

New York: Routledge.

U.S. Department of Labor and O*NET (2000). *Testing and Assessment: An Employer's*

Guide to Good Practices. Retrieved from <http://www.onetcenter.org/guides.html>

Weiss, H. M., & Adler, S. (1984). Personality and organizational behavior. In B. M. Staw

& L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. VI, pp. 1–50).

Greenwich, CT: JAI Press.