Modeling Attrition in a Military Selection Context

Aaron Keith Coombs

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science

In

Psychology

Neil M. Hauenstein, Chair
Charles Calderwood
Jorge Ivan Hernandez

October 2, 2020
Blacksburg, Virginia

Keywords: Military, Attrition, Personnel, Selection, Personality

Modeling Attrition in a Military Selection Context

Aaron Keith Coombs

ACADEMIC ABSTRACT

Attrition, employee turnover, self-selection, and withdrawal all refer to an employee's exit from an organization, or from an organization's recruitment or selection process. When individuals with the desired knowledge, skills, abilities, and other qualities (KSAOs) attrit, it represents lost productivity to an organization (Barrick & Zimmerman, 2009). Therefore, organizations should seek a selection program that screens out unwanted characteristics while minimizing the voluntary withdrawal, or quitting, of those who would be a good organizational fit. A military selection context amplifies these two aims because of the limited number of qualified individuals relative to the organization's personnel needs, and because of the high potential cost of a bad hire. However, there are few studies of attrition during a selection process, and even fewer in a military context that combines physical, cognitive, and personality components as relevant performance dimensions.

The purpose of the study was to model attrition from a military special operations selection through training program to determine what combination of physical abilities, cognitive abilities, and personality scales best predicts success. The study examined archival data from 748 candidate records spanning eight different classes during 2019. Secondary purposes of the study included comparing differences in attrition from the first week of the program to the remaining seven weeks, and comparing the predictive validity of a personality trait profile model to a model using personality scales T-scores. In conducting the analysis and modeling, exploratory factor analysis was conducted on the sample Jackson Personality Inventory-Revised (JPI-R) personality scales, finding both similarities and differences with previous study samples (Detwiler & Ramanaiah, 1996; Paunonen & Jackson, 1996).

The result of the study was a logit prediction model with a ROC AUC of .784, and an F_1 score of .69, that incorporated three physical predictors, performance IQ, and three personality variables: JPI-R T-score for sociability, and two composites created from the factor analysis—a Conscientiousness Composite and an Openness Composite (negative relationship with candidate success). Models for week 1 attrition and attrition from weeks 2-8 differed from the 8-week attrition model, and from each other in the significance and the importance of the personality variables and of cognitive abilities. Physical predictors: run score, pushups score, and sit-ups score, were significant and strong predictors of success for each of the time periods. Verbal IQ was not significant for any time period, while performance IQ was significant in predicting 8-week success, and for success during the week 2-8 time period. Personality predictors varied the most by timeframe, although some component of Conscientiousness predicted strongly for each timeframe. Whereas Openness-related facets predicted for 8-week success and success from week 1 with a negative relationship, Openness factors were non-significant in weeks 2-8. In contrast, Anxiety, a related sub-facet of Neuroticism, predicted moderately (negative relationship) for success from weeks 2-8, but was non-significant for week 1 and for the 8-week program.

Unexpected findings included the sample's different factor structure on the JPI-R, the dominance of the physical predictors in all models, and the strength of personality predictors relative to cognitive abilities. Implications for military and similar types of selection contexts, where selection through training includes a significant physical component, are discussed.

Modeling Attrition in a Military Selection Context

Aaron Keith Coombs

GENERAL AUDIENCE ABSTRACT

The study analyzed attrition from a military special operations selection program to determine what combination of individual differences measured before the program best predicted attrition during the program. The individual differences measured prior to the program were physical abilities, cognitive abilities, and personality. Archival data from 748 candidate records spanning eight different classes during 2019 was analyzed. Attrition is the departure of an individual from an organization, or from a hiring process. This study dealt with attrition from a hiring, or personnel selection process, which is less commonly studied than attrition from within an organization. Secondary purposes of the study included how attrition from the first week of the program differed from the remaining seven weeks, and determining if a specific broad personality profile best predicted attrition. The study found additional results that were not anticipated, specifically, that the military sample differed meaningfully on important dimensions of the Jackson Personality Inventory-Revised (JPI-R) personality scales, in comparison with previous study samples (Detwiler & Ramanaiah, 1996; Paunonen & Jackson, 1996).

The practical result of the study was a mathematical prediction model that incorporated a candidate's scores on pushups, sit-ups, 2-mile run, performance IQ, and three personality variables, and calculates a candidates' probability of success. The three personality variables that predicted success were scores for sociability, and two composites—a Conscientiousness Composite and an Openness Composite. Mathematical models for week 1 attrition and attrition from weeks 2-8 differed from the 8-week attrition model, and from each other, suggesting that attrition during different timeframes is due to different reasons. Physical predictors: 2 mile run

score, pushups score, and sit-ups score, were strong predictors of success for each of the time periods. Verbal IQ did not predict for any time period, while performance IQ predicted 8-week success, and success during the week 2-8 time period. Personality predictors varied the most by timeframe, although a component of Conscientiousness predicted strongly for each timeframe. Openness-related personality facets predicted for 8-week success and success from week 1 with a negative relationship. In contrast, Anxiety, a related sub-facet of Neuroticism, predicted moderately (negative relationship) for success only from weeks 2-8.

Unexpected findings included the military sample's different factor structure on the JPI-R, the dominance of the physical predictors in all models, and the strength of personality predictors relative to cognitive abilities. Implications for military and similar types of selection contexts, where selection through training includes a significant physical component, such as police or firefighters, are discussed.

Acknowledgements

I would like to thank my advisor, Dr. Neil Hauenstein, for his mentorship and guidance throughout the process of this study. Further thanks go out to the members of my thesis committee—Charles Calderwood and Ivan Hernandez—whose feedback made this a better study and whose direction made me a better student of I-O Psychology.

TABLE OF CONTENTS

List of Tables

List of Figures

1.0 Introduction

The United States Military's prosecution of World Wars I and II greatly accelerated the implementation of psychological and physical ability assessments to screen, select, and assign Soldiers and to predict Soldiers' performance in a variety of jobs (Harrell & Harrell, 1945; Schmidt & Hunter, 2004). However, over the following several decades, selection research and organizational practice has outpaced the use of predictive selection instruments throughout much of the United States Military. A notable exception during this timeframe has been Special Operations Forces (SOF), which continued to use extensive psychological testing, both cognitive and non-cognitive, to supplement its comprehensive physical ability assessments. Because of a SOF belief that "People are more important than hardware" (Collins, 1987, p.v), special operations recruiting and selection has placed a greater emphasis on individual traits than its broader service components. Additionally, because of the relatively small scale of SOF in comparison with the rest of the military, it can afford to be much more selective.

Given the multitude of knowledge, skills, abilities and other characteristics necessary for success, SOF selection and training programs are a natural fit for the study and analysis of personnel attrition. Although large amounts of individual difference data are available for each SOF candidate, these data are not fully leveraged due to the lack of criterion-related validity research. This study aims to address a scientist-practitioner gap by using available assessment data to predict attrition over an eight-week special operations selection through training program. SOF selection programs subject volunteers to a multiple-hurdle selection process, span several weeks in time, and encompass a mix of physical, cognitive, and emotional challenges. Candidates are free to withdraw from SOF training at any time and be reassigned or return to their previous unit without career penalties, while only candidates successfully completing the training program are eligible to join SOF. As such, special operations selection programs aptly represent the paradigm that “Organizations select individuals for jobs, but individuals also choose where to work” (Ryan et al., 2000).

1.1 Attrition, Withdrawal, Self-Selection, and Voluntary Turnover

Employee turnover, attrition, self-selection, and withdrawal all refer to an employee’s exit from an organization, or from an organization’s recruitment or selection process. Regardless of the nuance of the term, each represents an organizational cost. In cases of voluntary turnover, personnel costs increase due to the need to replace an employee, while lost productivity also taxes the organization (Barrick & Zimmerman, 2005). When prospective employees quit the recruiting or selection process, that phenomenon is often termed self-selection (Ryan et al., 2000), withdrawal (Schmitt & Ryan, 1997), or attrition (Topp & Kardash, 1986). For the sake of consistency, this paper will consider the term attrition to encompass any form of an employees’ departure from either the organization or the recruitment or selection process—a term that encompasses withdrawal, turnover, and self-selection. Attrition from the recruitment and selection process represents a different but still important problem for organizations, since there is an opportunity cost involved in selection processes. Adverse attrition (Ryan et al., 2000) occurs when applicants with otherwise good fit for the organization self-select out of the hiring process. Losing candidates in this type of attrition, via the voluntary withdrawal of those with the knowledge, skills, or abilities likely to help the organization, represents a lost opportunity. It reduces the selection choices available to the organization, while also potentially depriving the organization of viable employees. Alternatively, attrition of the right individuals can help the organization (Schmit & Ryan, 1997), and reduce the cost of additional training or pre-empt errors in the selection process. More generally, the issue of attrition is reflected in Schneider’s

(1987) Attraction-Selection-Attrition (ASA) model in which organizations seek to attract, select, and retain individuals that fit with the organization.

Although there is less research on attrition than job performance, even less common is research dedicated to understanding attrition from recruiting and selection processes (Schmit & Ryan, 1997; Ryan et al., 2000). When it is studied, attrition from selection processes is most commonly studied in the context of police selection and assessment. For example, Topp and Kardash (1986) studied police selection and found that general intelligence predicted performance in police selection, but personality assessment scores better predicted police training completion. Schmit and Ryan (1997) also studied attrition during police selection and analyzed the relationship of many variables on attrition. They found that in self-reported reasons for attrition, perception of lack of job and organization fit was common, followed by applicants' perceptions of their other alternatives. A limitation of their study was its focus on early hurdles in a broader selection context (Schmit and Ryan, 1997), however, and they acknowledged that different factors may influence attrition at different phases of a lengthy selection process. Later, Ryan, et. al., (2000) analyzed attrition of police applicants in a multiple-hurdle process and found completion of training was related to multiple individual differences, including perceptions of the organization, levels of job commitment, expectations of success, perceived employment alternatives, relocation requirements, social influence, and perceptions about the fairness and validity of the selection process. Additionally, Ryan et al., (2000) found differences between those who dropped out of the selection process early versus later, specifically concluding that they had more negative perceptions of P-O fit, and less individual job commitment. Late-stage attrition was unrelated to fit and job commitment. Finally, Hewgley (2013) found that several subscales of the Candidate and Officer personnel Survey (COPS) (Guller & Guller, 2003), specifically biodata scores, personality problems scales, and narrow-band personality scale scores predicted police academy attrition. The personality problems subscale, which measures dimensions of psychopathy similarly to the MMPI, was found to be the predictor with greatest correlation to attrition, with an odds ratio of -.41 for academy attrition (Hewgley, 2013).

A study limitation discussed by Hewgley (2013), but relevant to all of the reviewed studies, is the degree to which its results would generalize. Predictive validities found in the previous studies may not be as strong in other environments where there is greater range restriction, or where multiple hurdle selection processes differ significantly in scope, sequence, or in instrument used. For example, in a military special operations selection context, there is already direct range restriction, and likely indirect range restriction, because all special operations recruits have already passed the hurdle of being selected into the military using similar criterion. Another difference in context is in perceived alternatives, since SOF applicants do not become unemployed if they attrite from SOF training, whereas police trainees may be unemployed if they drop out of training. The differences between previous studies and the proposed study context are significant enough that a study of attrition in a SOF selection program will provide important contributions to the literature. While there is enough existing literature to develop logically sound research questions based on existing predictors, the context is unique.

1.2 Military Special Operations Forces Context

Entry into SOF requires successful completion of a selection through training program in addition to what all servicemembers experience during basic training or boot camp. In this instance, the selection context is an 8-week training program conducted at a U.S. Army based in the southeastern United States. Nine cohorts pass through training each year. Candidates entering training come from two sources. The majority of the candidates come immediately after completing Army basic training and their military occupational specialty schooling. For these candidates, the time in military service ranges from 6 months to 18 months, and the mean candidate age is approximately 22 years. The second category of candidates comes from the Army at-large, and apply to special operations from another unit in which they are already performing their assigned military occupational specialty. These candidates typically spend one to two years beyond their initial entry training in an operational unit, but are still on their initial enlistment in the U.S. Army. This population makes up less than twenty percent of the total candidates, has more realistic military experience, is more likely to be married, and is older on average, with a mean age of approximately 25 years.

Because of fixed resources, the training program has a capacity of 165 candidates per cohort, for a maximum of 1,485 total candidates across the nine cohorts. While awaiting admission to the training program, candidates are received, staged, further prepared, and screened on site. This pre-training phase varies from a week to several months based on the influx of candidates, the time of year and scheduled future courses, the personnel needs of the organization, and the readiness of individual candidates to attempt the training program. All candidates awaiting admission to training are pre-screened on cognitive ability, personality traits, psychological fitness, and physical abilities. This pre-screening is the first hurdle in the training process and is intended to screen out those candidates at high risk of failure in the formal training program. In the past fiscal year, for example, 1479 out of 2684 candidates entered the 8-week course, representing 55% of the total recruited and screened.

1.2.1 Initial Screening. The organization administers the Multi-Dimensional Aptitude Battery- II (MAB-II) (Carless, 2000) cognitive ability assessment during this pre-screening process, and considers any candidates who score lower than one standard deviation below the mean, or two standard deviations above the mean, to be higher training or social risks, warranting a psychological interview prior to acceptance into a class. Although not a hard cutoff, this screening contributes to a population in which cognitive ability scores are rarely below 85 or above 130.

Candidates are also psychologically screened using the Minnesota Multiphasic Personality Inventory (MMPI), and the Jackson Personality Inventory-Revised (JPI-R) instruments. The MMPI is used to screen candidates for psychological fitness and to identify risks. Specifically, the organization psychologist uses the MMPI to measure levels of psychopathy incongruent with service in a high-risk/ high-demand occupation where life and death decisions are common. For instance, the MMPI provides an indicator of levels of neuroticism, extreme cynicism, inefficacy and anxiety that would put a candidate at risk in training or in service to the organization. In addition, the MMPI's validity scales are used as an indirect indicator of a candidates' integrity. Because its intended purpose is clinical, and because its use in SOF screening is solely to reduce risk from those at the extreme ends of psychopathy, the MMPI is not being used in this study. Further, those elements of the MMPI which are likely to be most relevant for predicting attrition are already measured by the JPI-R.

While the JPI-R is not used independently as a selection hurdle, it is administered and considered as part of a candidate's individual profile that may be used in an interview with the psychologist or in future analysis. In this respect the JPI-R is used to develop an understanding of the "whole candidate" in the screening process, and specifically referenced if a candidate is flagged as a fit risk or training risk from another screening method. The JPI-R administration does not specifically contribute to any range restriction of candidates prior to admission to the program.

An Army Physical Fitness Test (APFT) is administered to candidates during the pre-screening process and candidates are admitted to the next class partly based on competitive scoring against other candidates. For physical screening, there is some degree of range restriction created in the physical domain because candidates are considered to be at risk if they score less than 70 points on any individual event, or less than an aggregate APFT score of 230 total points. Because of the competition just to get into the selection through training program, full-scale APFT scores range from 215 on the lowest extreme, to a perfect score of 300, with few candidates admitted who score lower than a total score of 240 on the assessment.

1.2.2 Program Overview. Those candidates passing the initial pre-screen hurdle enter the 8-week training program. The program of assessment and training is designed to test candidates' suitability for the rigors of special operations, to train the candidates on necessary knowledge and skills, and to instill SOF history and unit norms. In this respect, the training program operates as both a selection process, a training program, and a socialization program. The 8-week course progresses to increasingly complex and increasingly difficult tasks. Over the course of training, performance is evaluated at both the individual and team level. An overview of the training program and 2019 success/attrition rates appear in Figure 1.

Because of the pre-screening process and the progressive structure of the selection context, it is likely that candidates drop out for different reasons at different stages. There is support for this phenomenon in previous studies. For example, Ryan et al. (2000) found differences between those who dropped out of police academy training early in the process vs. those who dropped out late in the process, suggesting support for the use of a dual-stage model. In the SOF selection context, early attrition is probably due to differential individual commitment and perceptions of organizational fit, while later attrition is more likely to be due to self-efficacy (Bandura &

Adams, 1977; Bandura, 1986; Chen et al., 2001) and individual performance expectations. As candidates experience more of the course, the increased challenges and the greater quantity of information available in making a stay vs. go decision is likely to change the attrition decision. In this context, 35% of total attrition happens during week one with the remaining attrition unevenly distributed over weeks two through eight.

It is important to highlight that attrition in the SOF selection context differs from other studies of attrition in an important external factor. Candidates who quit do not leave the organization, but rather are reassigned in the same job type within a non-special operations unit. This is an important distinction because perceived alternatives have been demonstrated to influence attrition decisions in other contexts (Schmitt & Ryan, 1997; Ryan et al., 2000; Zimmerman et al., 2016; Hom et al., 2017). In the context of this study, candidates have very little ambiguity with regards to their alternatives.

1.2.3 Importance of Predicting Attrition in SOFs. The special operations human resources structure demands a consistent inflow of successful candidates meeting known standards on critical events. As a result, the selection decision process operates as selection through training—where training proficiency provides the context for assessment. Attrition is based on a completely non-compensatory selection model where the theoretical maximum training success rate for any incoming class is one-hundred percent. Critical event cutoffs are not adjusted to meet needed accessions rates; rather, training success rates vary from cohort to cohort, ranging from a 33% success rate to a 59% success rate in the past fiscal year. Therefore, from a utility perspective, accurate prediction allows for greater consistency in training success rates over time, and potential cost savings by reducing the amount of adverse attrition.

This is an important goal because the organization has an average annual need for 850 new Soldiers, and must operate its selection through training program with a finite number of cadre and amount of time. Increasing the yield from each class could allow the organization to recruit and screen fewer Soldiers, resulting in time and money saved. Or, a reduction in attrition could allow the organization to decrease class size, allowing for more effective cadre-to-candidate ratios and better training. Alternatively, reducing attrition could allow the organization to reduce the number of courses per year, with reinvestment of the resources into other priorities. Finally, reducing adverse attrition could provide more new Soldiers to the organization annually, increasing competition for promotion and advancement within the organization.

Analyzing attrition in a special operations selection context also contributes to existing literature by empirically examining the relative strength of multiple predictors in a multiple-hurdle process that involves not only cognitive and non-cognitive mental traits but also physical ability and risk of physical harm. As a result, the context provides an opportunity to explore the relative importance of individual differences on attrition, rather than the more oft-examined criterion of job performance or training performance. It further allows an examination of the influence of typical predictors such as general mental ability, when so many other dimensions of ability and performance are necessary to successfully complete the entire 8-week course. The need of all successful candidates to perform baseline above-average levels of problem-solving, task mastery, emotion regulation, and physical endurance may moderate the predictive strength of commonly understood constructs.

A final way this study advances the science on attrition is in its use of an uncommon population. Schmit and Ryan (1997) noted that with the exception of limited policy academy studies, the majority of research on attrition from job seeking processes was conducted on college students, implying contextual gaps. A military population demographic may differ significantly from college students or police academy candidates on measures of individual traits and on situational factors affecting attrition. As an example, unlike police academy candidates, a SOF candidate does not face a career change, but rather an internal organizational change. Further, the candidate already has a known alternative for employment. All candidates either volunteer for SOF from other organizations. As a result, the risk-reward decision of SOF attrition is different from many other contexts.

1.3 Current Study Overview

The goal of the current study is to predict overall attrition, as well as attrition during week 1, and separately as a result of the remaining assessment and training program, weeks 2-8. The predictors include cognitive ability, personality scales, and physical abilities. Candidate attrition is greatest during week 1 of training and represents the greatest potential for increasing program utility. The high attrition rate during week 1 is counterintuitive because there are no critical event failures during week 1, which means physical injuries and voluntary quitting are over-represented. As such, it is likely that psychological predictors account for the variance in the week 1 attrition rates. After week 1, attrition rates fluctuate as a function of training and testing difficulty; where more difficult and complex critical events are tested, there is greater attrition from candidates quitting as well as from critical event failures. This surface level observation suggests that personality scales and self-efficacy may predict candidates’ decisions to quit.

2.0 Literature Review

2.1 General

The problem of employee attrition or employee turnover has been studied for over a century, from early published studies (Diemer, 1917) investigating college faculty turnover or explaining reasons or costs of labor turnover (Douglas, 1918; Hom et al. 2017). Over 100 years of scientific effort towards the problem has resulted in several widely accepted conclusions, while also resulting in inconsistent results and debate. From its earliest studies (Diemer, 1917; Douglas, 1918; Slichter, 1921) to recent meta-analysis examining turnover (Heavey et al., 2013), the high cost of voluntary turnover—an employee's "voluntary severance of employment ties" (Hom et. al., 2017; Hom & Griffeth, 1995)—to organizations is widely supported.

Researchers' attempts to better understand attrition have taken various approaches, from examining individual differences (Barrick & Zimmerman, 2005; Griffeth et al., 2000; Maertz & Campion, 2004; Maltarich et al., 2010; Zimmerman, 2008; Zimmerman et al., 2016;), environmental/organizational effects on individuals (Hulin, 1966; Hellriegel & White, 1973; Mobley et al., 1979), broad economic/ labor market correlates (Hulin et al, 1985; Steel & Griffeth, 1989;), and more complex theories incorporating individuals and multiple other levels of analysis such as fit (Lee & Mitchell, 1994; Kristoff, 1996; Maertz & Campion, 2004; Mitchell et al., 2001; Kristoff-Brown et al., 2005; Kozlowsky, 2009; Schneider et al., 1998; Zimmerman et al., 2019). Regardless of the type of model, however, predicting attrition has been characterized by much weaker correlates that job performance. Some of this may be explained by how constructs and dependent variables have been operationalized (Li et al., 2014). How to measure whether an employee's voluntary departure represents an organizational loss or gain also represents an area less well-supported. While voluntary turnover is generally accepted as bad because replacement costs organizations between 90-200-percent of annual salary in lost productivity and hiring costs (Allen, Bryant, & Vardaman, 2010), some turnover is good (Ryan et al, 2000), so long as it is not those with the right knowledge, skills, and abilities quitting. The key, which remains a theoretical, empirical, and practical challenge, is to understand and reduce the amount of voluntary turnover from employees who are benefits to the organization. Based on the context and practical implications of the current study, the most relevant previous studies are those that examine the relationship of individual differences with attrition from a selection through training context. Studies that include cognitive and non-cognitive psychological predictors as well as physical abilities provide insights into the potential interaction of the different domains.

2.2 Models of Voluntary Turnover

Early theoretical models explaining employee attrition are generally attributed to the work of Mobley (1977, 1979; Hom, et. al., 2017), who used rational decision-making theories to explain employee turnover decisions. Although explaining turnover in terms of perceived alternatives, expectancy theory, and rational decision-making still contribute to more comprehensive contemporary models (Zimmerman, et. al., 2016; Zimmerman et al, 2019), the models that best explain attrition from the hiring process—i.e., a selection through training program—are those that combine facets of individual differences and the environment, or what Mischel and Shoda (2010) referred to as "The Situated Person." Therefore, person-organization fit literature is relevant to generating hypotheses about attrition from a selection process. Since the seminal work of Schneider (1987) presented the attraction-selection-attrition (ASA) cycle, theoretical and empirical studies have incorporated its tenets and tested its premises.

Although intended as a way to understand organizational behavior, Schneider’s ASA framework (Schneider, 1987; Schneider et al., 1998) has important implications to individual attrition processes. Specifically, ASA highlights the utility of personality and interest measures in determining potential fit of prospective employees. The theory posits that individuals are initially attracted to those organizations that they perceive have cultures, climates, and characteristics they value and which they share. During the recruiting and selection process, the organization’s culture and organizational personality influence the selection of those individuals who most closely match the organization. Further, individuals’ perceptions of fit within the organization will heavily influence voluntary turnover, or attrition. Over time, this cycle gradually results in an organizational personality towards homogeneity which explains how the organization acts.

Several researchers have advanced the literature on Person-Organization fit (P-O fit), combining theoretical reviews (Kristof, 1996), and empirical studies (Schneider et al, 1998; Kristoff-Brown et al., 2005; Oh et al., 2017). For example, Kristof’s 1996 integrative review qualitatively synthesized two decades of previous findings regarding the complementary or supplementary fit of an individual with the organization. It found that individual personality fit with the organizational climate, and individual value congruence with the organizational culture, both predicted attrition of employees within their first two years of employment. Meanwhile, Schneider et al. (1998) studied the relationship between manager’s personalities and organizational hiring outcomes from over 13,000 applicants in 142 industries and found that managers’ personalities affected hiring decisions within and between industries, providing empirical support to Schneider’s earlier (1987) ASA model.

A separate mid-90's theoretical contribution that contrasted with earlier rational decision-making models of turnover is that of Lee and Mitchell (1994). Integrating concepts from various academic fields such a s social psychology, decision science, and statistics, the authors outlined four distinct decision paths representing the psychological processes involved in voluntary turnover which they termed the unfolding model of turnover. The authors sought to improve on previous theoretical attempts that they considered too simple and instead explain the many ways in which individuals quit. A major contribution of their theory was the concept of shocks—"event[s] that prompts an individual to evaluate his or her current and perhaps other jobs" (Lee & Mitchell, 1994, p.84). The concept of shocks not only explained why attrition tends to occur early—i.e., in the hiring process or organizational socialization process—it was often included in subsequent models of turnover.

Maertz and Campion (2004) further advanced turnover theory and modeling, considering attrition as an interaction of four different decision types and eight types of motivational forces of attachment and withdrawal. They presented the likelihood of quitting by the type of decision-maker an individual is, and the type of motive forces at play. Two relevant contributions from this research include the concepts of pre-planned quitters—those that plan in advance to quit at a certain time in the future—and conditional quitters—those that will quit in some uncertain shock happens. Both of these decision types are likely to influence the attrition from a special operations selection through training program and may explain the high week 1 attrition.

Ryan Zimmerman and Murray Barrick contributed significantly to the understanding of personality as a predictor of employee turnover in multiple studies. In their 2005 study of data on 445 job applicants from two separate organizations they found that biodata, clear-purpose attitudes and intentions, and disguised-purpose dispositions related to retention predicted voluntary turnover. Specifically, measures of self-confidence, decisiveness, desire for the job, and intent to stay all predicted turnover. Collaborating with Amy Kristoff-Brown, and Erin Johnson on a separate but related meta-analysis of individuals' fit at work (Kristoff-Brown, Zimmerman, & Johnson, 2005), researchers analyzed 172 studies of fit, finding support that even during pre-entry (hiring process), fit perceptions strongly influence attrition. While the authors did not include personality traits as antecedents of turnover, their discussion of fit literature and future research highlighted that such research was warranted. In a later study of 354 credit union job applicants, Barrick and Zimmerman found further support that biodata measures, personality measures, and pre-hire attitudes measures predicted voluntary attrition (2009). Specifically, they found that employment motivation, personal confidence, and traits of conscientiousness and emotional stability predicted attrition within the first six months after hire (Barrick & Zimmerman, 2009, p. 200).

In a separate meta-analysis of 86 previous studies, Zimmerman (2008) studied the impact of personality traits on attrition and found that of the five-factor traits, Emotional Stability best predicted employees' intent to quit while Conscientiousness and Agreeableness best predicted actual quitting. An interesting finding was that personality had a direct effect on attrition, independent of commonly modeled mediators such as job satisfaction or job performance. While not focusing specifically on the hiring process, previous conclusions about the stability of personality traits and the meaningfulness of individual-organizational fit during the hiring and selection process suggest that personality traits would similarly predict attrition pre-hire.

The most contemporary work on employee attrition includes theory that incorporates previous advances and an acknowledgement that individual differences, organization factors, and extra-organizational factors can all influence and predict attrition. For example, in an integrative conceptual review of turnover literature, Zimmerman et al. (2016) used the Cognitive-Affective Processing System (CAPS) framework (Mischel & Shoda, 2010) to explain how individual differences such as personality or attitudes operate with respect to environmental influences, resulting in turnover. An area uncovered for future research by the authors was a need to examine the effect of GMA on withdrawal behaviors. While there is rich literature on the impact of GMA on performance (Hunter, 1986; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004, Hunter, 2006), there is much less research suggesting its impact as a predictor of turnover. In their 2000 meta-analysis of antecedents and correlates of turnover, for example, Griffeth et al. found only seven previous studies that measured the correlation of cognitive ability with turnover.

In an empirical study providing an update on the homogeneity hypothesis of Schneider's (1987) ASA framework, Oh et al. (2017) researched the extent to which ASA processes contributed to within-organization homogeneity and between-organization homogeneity. Incorporating large data samples of pharmaceutical, banking, and manufacturing hires, the researchers found that selection processes most heavily contributed to within-organization homogeneity, with new hires' measures of extraversion mostly strongly relating to the organizational norms. Their results suggest that attrition in a hiring process may be heavily influenced not only by person-organizational fit, but by personality trait similarities to the organizational norms.

2.3 Attrition from a Hiring Process

There have been fewer studies of attrition from hiring processes vs attrition/turnover once the employee has been hired, trained, and put into position. This represents a missed opportunity because the costs of recruiting and hiring, while not equivalent, are still considerable. The need to recruit individuals with needed KSAOs, select those with KSAOs and organizational fit, and retain them throughout the process is particularly important where high-skilled workers are required. This is specially a problem if otherwise qualified or competitive minority candidates withdraw, because of its impact on a disparate impact determination. Because of that, much of the existing literature on withdrawal from a hiring process focuses on racial adverse impact, particularly in civil-service/ public safety occupations.

An early example is a study on the differential dropout rates of minority vs. majority candidates by Arvey et al. (1975). The authors found that delays between application and the beginning of selection process had the greatest correlation with attrition, and that attrition was higher for minority candidates than for racial majority candidates. A key contribution of this study was its suggestion that keeping the selection processes as short as possible and minimizing delays would reduce unwanted attrition. A later empirical study of selection length and applicant attrition refuted this long-held belief that longer selection processes increase attrition, however. Using survival analysis in a sample of 222,772 job seekers applying to a variety of jobs in a broad sample of industries, Hardy et al. (2017) found that attrition risk decreases with increased length of the selection process, and that most attrition happens at the very beginning of a selection process. Their recommendation that practitioners prioritize reliability and validity of selection processes over concerns about process length and applicant attrition requires qualification, however. Their study was limited to online application contexts which were measured in minutes and hours, not days and weeks, and therefore may not generalize.

Studies of police hiring have been the predominant context for analysis of applicant attrition during a selection process. For example, Topp and Kardash (1986) studied the relationship between personality traits and police academy performance and attrition in an 11-week selection and training program. Using the 16 Personality Factor Questionnaire (16 PF) to measure traits, the authors found that in a predominantly male, married population, dropouts differed significantly from successful graduates on half of the 16 scales, but not on cognitive ability, which is consistent with the findings of Griffeth et al. (2000). Specifically, they found the greatest differences in measures of extraversion, self-confidence, and emotional stability. The authors also found that personality scales better correlated with attrition than previous years of education, further supporting the role of personality over other measures in predicting attrition.

Schmitt and Ryan (1997) studied applicant withdrawal from a selection process using a police applicant sample of 3,290, in which 618 (18%) withdrew. Their post-hoc model proposed that perceptions of fit with the job and the organization, and the applicant’s assessment of perceived alternatives were two leading causes of attrition. Later, Ryan et al. (2000) examined applicant attrition from a multiple-hurdle hiring process in a population sample of 3,550 police applicants, finding that attrition reasons in early stages of the 8-12-week process differed from those in late-stage attrition. In their study, perceptions of fit and of individual commitment to the job and organization correlated strongly with attrition. Additionally, having other alternatives to being a police officer correlated significantly with attrition. The authors did not measure individual traits, however, and only sought to find correlates of attrition from self-reported questionnaires completed prior to police training. The antecedent variables were attitudinal and behavioral.

Hewgley (2013) specifically sought to analyze the relationships between trait-based predictors and police academy attrition from an 18-week police academy. Using cognitive ability, personality, and biodata scales in a small sample of 117 police officer candidates, the researcher conducted multiple analyses and considered 34 predictive variables, finding evidence that biodata and personality scales predict attrition. The study results suggested that personality is a significant predictor of attrition, with subscales of a COPS instrument related to personality problems and integrity/dishonesty having the greatest predictive validity.

The studies on attrition from selection contexts often focus on individual differences—those traits, attributes, attitudes, or demographic factors that can predict attrition. Collectively, they suggest that length of the selection process is less important than individual factors, so long as process fairness and the validity of the selection measures are maintained. Further, they emphasize the potential utility of a multiple-hurdle process where biodata, personality, and attitudes are measured before more costly selection and training events.

2.4 Military Selection Contexts

Few published studies have explored trait correlates of attrition from a military selection through training process, but the type of prevailing research and limitations on publication have likely prevented more widespread sharing of conclusions. Relevant military studies help form an understanding of the physical domain's importance on attrition outcomes. Because of the physical demands of military service, most previous studies have examined medical indicators and physical ability as predictors of attrition, while fewer have examined the relationship of mental factors on attrition. However, previous research on the role of cognitive ability in training success (Schmidt and Hunter, 1998), and on self-efficacy (Chen et al., 2001) suggests that cognitive ability may better predict attrition in this context than in previous studies (Topp & Kardash, 1986; Griffen et al., 2000).

Pope et al. (1999) examined the predictive validity of several medical, biographical, and physical fitness factors on attrition from Australian Army Basic Training. In a sample of 1317 male recruits over a 12-week training program, they used multivariate survival analysis to find that candidates' fitness on a simple shuttle run test was the best predictor of attrition, followed by medical measures of lower leg bone strength. While their study did not include cognitive or personality predictors, it has relevance on voluntary attrition rates indirectly, as physical demands influence self-confidence, self-efficacy, and test recruits' resolve.

Moran et al. (2011) developed a predictive model for attrition from an eight-month Israeli Defense Force (IDF) combat unit training program, based on both physical and psychological factors. In a combined study of 120 men with an average age of 18.7 years, they developed a model that would predict 69.8 percent of attrition as a function of medical health, entry time physical fitness, and psychological factors. The factors that correlated highest with attrition were perceptions of fit, self-efficacy, and self-confidence of candidates. In a 2012 study of U.S. Army Recruits, Gubata et al. analyzed the AIM (Assessment of Individual Motivation), a noncognitive personality and background test (Knapp et al., 2004). The results of 47,974 recruit records against first-year attrition from the U.S. Army showed that those scoring on the lowest quintile had the highest attrition rates, and that the AIM physical conditioning subscale score had the highest predictive validity. Their results provide insight into how to reduce attrition rates in otherwise at-risk populations, such as young adults without high school diplomas.

Two military population studies published in 2015 took different approaches to studying attrition. In an empirical study of Dutch Infantry training attrition, Binsch et al. (2015) analyzed the predictive relationships of medical, psychological, and organizational indicators on attrition in a 24-week program that had historical attrition rates ranging from 42% to 68%. Although they had a small sample size (N=85), their results suggested that within the physical/ medical dimension, body fat measurements substantially lower than 20 percent predicted attrition, while within the psychological domain, those candidates scoring low on vigilance, self-esteem, and self-efficacy were more likely to dropout. In a qualitative study of attrition within the British Army Infantry, Kiernan et al. (2015), sought to better understand recruits' reasons for quitting its 12- week course. They studied 1000 recruits with exit surveys and had recruits self-report their reasons for attrition. The attrition rate in the study was 36.25%, 47-percent of which was attributed to voluntary attrition (a personal decision to quit). In their post-hoc analysis, researchers concluded that the psychological stress and perceptions of fit were the leading contributors to voluntary attrition. Their conclusions are supported by the temporal distribution of attrition, with a preponderance of attrition from the British Infantry Training being in the first few weeks of the course.

In one of the few published studies of attrition from a Special Operations Forces (SOF) selection and training program, Colosio, Fontana, and Pogliaghi (2016) conducted an observational study of Italian Army Rangers attrition. Among 103 male recruits predominately in their late 20's (M=26 +/-2 yrs), the researchers found that of the 41% attrition rate over the 6-month program, 60% of it (25 subjects) was due to voluntary attrition or quitting for personal reasons, with the first month of training having the highest attrition. Although they did not study psychological predictors, the authors' post-hoc analysis included recommendations to target psychological predictors such as motivation, self-efficacy, and resilience to better predict attrition and screen candidates.

Though previously published studies on attrition from a military selection through training context are sparse, they provide important insights to the antecedents of attrition. First, they suggest that physical fitness is an important indicator of attrition. Second, they suggest that certain psychological constructs are correlated with attrition. Specifically, candidates' perceptions of fit, self-efficacy, and motivation to complete the training and assessment were consistently indicated as important antecedents of attrition. The current study attempts to determine the relative impact of multiple predictors—physical, cognitive, and personality—on attrition from a military selection program, something none of the above-mentioned studies attempted. Because cognitive ability, personality, and physical ability are independent of each other (Zeidner & Matthews, 2000), a model using all three should predict better than models only using one.

2.5 Research Questions

Four research questions are proposed to be answered by the analysis of archival selection program data.

Research Question 1: How does early attrition differ from late attrition during the 8-week program?

Research Question 2: How well does a single model of total attrition compare to dual-stage models of attrition in which attrition from week 1 and remaining attrition from weeks 2-8 are modeled separately?

Research Question 3: How do overall cognitive ability, physical ability, and the five personality clusters interact to predict attrition?

Research Question 4: How well does candidates' personality similarity to an organizational exemplar profile predict attrition in comparison to T-score personality scales prediction?

3.0 Methods

3.1 Participants

The study participants are candidates in a Special Operations selection through training program. Candidates are active-duty U.S. Army members, and thus have already completed medical, psychological, physical, and biodata screening for fitness of general service. Additionally, each of the candidates already has completed U.S. Army Basic Training and Advanced Individual Training (AIT) in their assigned military occupational specialty. As a result, candidates have been in the U.S. Army for at least 6 months, but potentially up to three years. Age ranges from 18 years old to 34 years old, with a mean age of 23 years old and a median age of 21 for the entire sample population. The sample of 748 archival records is predominantly white (greater than 90 percent), and male (greater that 99 percent). The candidates represent 42 military occupational specialties, or job specialties, with the largest proportion (362 / 48%) being infantry. The candidates are all volunteers for the selection through training program, which requires subsequent service in U.S. Army Special Operations.

3.2 Process

While awaiting entry into the program, candidates wait between one week to four months onsite. They are screened for entry, further prepared for training, and provided rehabilitation from existing injuries if needed. During the waiting period, candidates are screened on cognitive ability, psychological risks, and physical ability. Candidates are all exposed to the same cadre of instructors, and complete the same events, in the same sequence, during the eight-week assessment.

The current study models attrition from an eight-week U.S. SOF selection through training program, with a sample of 748 candidates. Modeling is conducted considering a dual-stage approach, and using two types of analysis. The study design includes modeling attrition for the overall program, during week 1 of the program, and again separately from the remainder of the program (weeks 2-8).

3.3 Predictors

3.3.1 Cognitive Ability. Cognitive Ability is measured using the Multi-Dimensional Aptitude Battery – II (MAB-II), which allows assessment of 10 areas of intelligence grouped into three scores: verbal, performance, and composite or full-scale. Created by Dr. Douglas N. Jackson, and licensed by Sigma Assessment Systems, Inc. (sigma systems, 2020), the instrument claims test-retest reliability of the scales to be .95 for verbal, .96 for performance, and .97 for the full-scale scores. An update on the MAB, used widely since 1984, the MAB-II has been independently shown to be comparable to the Weschler Adult Intelligence Scales-Revised, and suitable for a valid measure of verbal, performance, and general intelligence (Carless, 2000). The instrument is widely used in Special Operations Assessment, with published studies showing use in both Air Force Special Operations as well as Army Special Operations (Chappelle et al., 2010). Scores are reported as standard scores of IQ, with population mean = 100 and standard deviation = 15 (Sigma Assessments, 1999). Verbal score ranges from 75 to 139 for candidates, while performance ranges from 71 to 134 and full-scale GMA ranges from 75 to 139 with a mean of 105. Both the verbal and the performance scores will be analyzed as predictors of attrition.

3.3.2 Personality. Personality is measured using the Jackson Personality Inventory-Revised (JPI-R), an instrument developed to measure 15 personality traits using 15 individual scales as well as aggregated composite scales for five cluster scores (Doster et al., 2000). Authored by Dr. Douglas Jackson, and licensed by Sigma Systems, Inc., it was developed to predict a broad range of behaviors in a variety of environments (Sigma Systems, 2020). Each of the 15 scales is scored on a scale from 0 to 100, and are expressed in terms of the total population (T-scores) and broken down by respondent sex. The instrument has been compared and contrasted to the NEO five factor model of personality (Costa & McCrea, 1992), and found to have the similar factor structure for Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness (Detwiler & Ramanaiah, 1996). A later analysis of the JPI-R and its factor structure further replicated the recovery of the big five factor structure (Doster et al., 2000). The 15 scales are: Complexity of Thought, Breadth of Interest, Innovation, Tolerance, Empathy, Anxiety, Cooperativeness, Sociability, Social Confidence, Energy Level, Social Astuteness, Risk Taking, Organization, Traditional Values, and Responsibility. Additionally, JPI-R reports cluster scoring of groupings of scales into 5 clusters which represent composites of the 15 scales: Analytical, Emotional, Extroverted, Opportunistic, and Dependable, also reported in relation to U.S. population norms. Each of the 15 scales will be analyzed as a predictor of attrition.

3.3.3 Physical Ability. Physical Ability is measured using the three-event Army Physical Fitness Test (APFT). The Army has used the APFT for the past 40-plus years to determine baseline physical eligibility for service. In addition, because of its standardization and ease of administration, it has been used widely within the Army as a screening and selection measure for general performance, suitability for career development opportunities such as advanced training, and admission to elite organizations and occupations. The three events consist of push-ups, sit-ups, and a 2- mile run, with each event raw score converted to a 100-point scale, and summed event scores equal to the total APFT score, out of a maximum score of 300 (FM 7-22). The test is designed to test broad spectrum anaerobic and aerobic fitness, muscular strength, and endurance without the need for extensive equipment or time (FM 7-22, 2012; Rapuano et al., 2016). As a result of physical screening and competition to get into each cohort, scores range from 63 to 100 on pushups, from 64 to 100 on sit-ups, from 70 to 100 on 2-mile run, and from 215 to 300 on total APFT score. Each of the three component scores—pushups, sit-ups, and 2-mile run score—will be analyzed as a predictor of attrition.

3.4 Criterion Variables

The dependent variable is candidate success or failure to complete the eight-week selection through training course, regardless of reason. The preponderance of attrition is from voluntary withdrawal, or candidates quitting (217/748 or ~30%). All recorded reasons will be counted as attrition in the current study, and week 1 success, and weeks 2 through 8 success will also be analyzed separately as criteria.

3.5 Control Variables

Three covariates to be controlled include overall class pass rate for the cohort, age of the candidate, and the time candidates spent awaiting admission to the program. Class attrition rates for the available records range from 41% attrition to 67% attrition. The age distribution of candidates ranges from 18 to 34 years old. Time on ground awaiting commencement of a class to start ranges from 1 week to 4 months.

3.6 Sample Size and Power

Sample size and power analysis for three timeframes indicates the 748 archival records are sufficient to find practically meaningful effect sizes (odds ratio of at least 2 or greater) with the 23 variables based on the base rates (Ferguson, 2009). A calculation for sample size requirements in multiple logistic regression (Peduzzi et al., 1996), indicates that for an alpha of .05, and power of .80, given the base rate of .453, and 23 covariates, a sample of 508 is needed for the 8-week timeframe. For the week 1 timeframe, a lower attrition base rate of .194 results in a larger sample of 1,186 required if all 23 potential variables are retained. In order to ensure adequate power, if necessary, I will consider results with an alpha of .10 or greater. Additionally, in developing models, statistically non-significant predictors will be removed, likely lowering the needed sample size required to detect practically meaningful effects. For attrition from weeks 2-8, an attrition rate of .438 results in a needed sample size of 525. Overall, the sample is sufficient for the analysis intended.

3.7 Analysis

The first analysis will be of descriptive statistics representing the candidate sample. Predictors will be checked for assumptions of normality and independence, and means and variance will be measured to compare with wider population norms. Predictor profiles will be calculated and compared for attrition by week, with weeks 5-8 being combined due to sample size and similarity of selection through training events. Point biserial correlations for each predictor and three timeframes—8-week, week 1, and weeks 2-8—will be analyzed.

The focal analyses will be logistic regression models where models will be evaluated using the Receiver Operating Characteristic (ROC) area under the ROC Curve (AUC) and the F_1 score for each of the three models. The ROC AUC method will be used as a general model diagnostic to compare final models, since no a priori costs are known in pre-selection (Flach, 2003, p.109) and this method will give an indication of the models' comparative balance of sensitivity and specificity, or predictive accuracy. However, because it is plausible that there are different costs of type I and type II prediction errors, a confusion matrix for each final model will be developed and comparison of F_1 scores will also be analyzed. Decision cutoffs will be established to balance type I and type II prediction errors. This will allow an analysis of the models' relative balance of recall (sensitivity), and precision, which may have more utility value to the organization (Tharwat, 2018).

Logistic regression analysis will be conducted on the 748 archival data records to determine a logit prediction model using each of the 15 personality predictors, 2 of the cognitive ability predictors, and 3 of the physical predictors, for 20 total predictors, against week 1 attrition.

A subsequent logistic regression analysis will be conducted on combined total attrition from weeks 2 through 8. This analysis will use the sample of the 603 candidates remaining after week 1. The same combination of 20 predictors from personality, cognitive ability, and physical ability will be used to model the odds of attrition vs. successful completion of the remainder of the selection through training program.

A third model will be fit using the entire sample and a logistic regression using the 20 predictors and total attrition from the full 8-week program as the dependent variable.

A final model will analyze the interaction effect of ability with personality in predicting attrition. Using either full scale GMA score for cognitive ability or full APFT score for total physical ability, an interaction effect between ability and personality will be analyzed. The ability variable with the strongest predictive value will be used.

A final analysis to be conducted will be a trait profile analysis. The trait profile analysis will model total attrition from the full eight-week course as a function of how similar a candidate's personality is to organizational norms. The analysis will follow inferential accuracy modeling procedures (Jackson, 1972; Reed and Jackson, 1975; Hauenstein & Alexander, 1991), comparing the correlation and the difference of a candidates' traits with norms existing in the organization's mid-level leaders. A comparison will be made between the trait profile analysis and a T-scores model to see whether candidates are more successful based on degree of similarity (whether positive or negative) from organizational norms versus their T-scores on the personality scales.

4.0 Results

4.1 Descriptive Statistics

Initial descriptive statistical analysis of the 748 data records identified data entry errors. One candidate's age was recorded as 1,000 years, while another candidate's verbal IQ was recorded as zero. Because each of the errors were entered in variables used in analysis as a control or an independent variable, and because power was anticipated to be sufficient without those records, they were excluded, and all subsequent analysis was completed using a total of 746 records.

For each of the included variables, the means, standard deviations, range, and distribution were examined. Each variable had sufficient variance, and with the exception of the three control variables—Class Pass Rate, Candidate Age, and Time on Ground—demonstrated levels of skewness and kurtosis less than |1|, a recommended threshold advocated by West et al. (1996). Class Pass Rate demonstrated kurtosis of -1.31, but little skewness. Candidate Age demonstrated slight skewness of 1.09 towards younger candidates, between 18-20 years of age, which was expected. Time on Ground exhibited both degrees of skewness (1.57) towards less time on ground, and kurtosis (4.44) that exceed recommended thresholds. Performance IQ and Verbal IQ were normally distributed, with mean candidate IQ being 105 for performance IQ and 104 for verbal IQ. Min and Max scores of 71 and 139 bounded the limits of those admitted to the program, indicating little range restriction in cognitive ability screening. Personality scales were generally normally distributed, with mean sample T-scores on the 15 JPI-R personality scales ranging from lows of 42 for scales of both Anxiety and Cooperativeness to a high of 57 for Organization. Notably, in the military sample, the scale means related to the FFM facet of Conscientiousness—Organization, Traditional Values, and Responsibility—were greater than male U.S. population norms. For example, in terms of T-scores, mean scores on those measures were 58, 55, and 55 respectively, indicating a .8 to .5 standard deviation mean difference increase over male U.S. population norms. In contrast, those scales associated with Emotional Stability—Anxiety and Empathy—resulted in lower scores than U.S. population norms. Candidate mean T-scores for Anxiety and Empathy were 41 and 44, representing a .9 and a .6 standard deviation lower than male U.S. population norms. Candidates also differed greater than a half standard deviation from population norms on Cooperativeness, a measure of Agreeableness, with a mean of 43 indicating a .7 standard deviation less than norms.

The physical predictors- Pushup, Sit-up, and 2-mile Run were sufficiently normally distributed based on the skewness and kurtosis statistics, but were slightly skewed towards higher scores. Candidates scored above Army norms on each of the events, with the Run score being the largest mean difference between the candidates and U.S. Army norms. The mean candidate run score was 93.5, out of a maximum score of 100, compared to an Army mean of 71.2 (Knapik et al., 1994). Candidates averaged 91.2 points on the sit-ups event compared to an Army mean of 70.0, and candidate mean pushup score was 87.4 compared to 71.0 Army-wide (Knapik et al., 1994). In terms of standardized effect sizes of differences, sit-ups score ($\underline{d}$ = 2.37) was the most significant difference, followed by run score ($\underline{d}$ = 1.84) and pushups score ($\underline{d}$ = 1.60). Full results of descriptive statistics expressed in T-scores are shown in Table 1.

4.2 Intercorrelation of Independent Variables

Intercorrelation of the predictor variables and checks for multi-collinearity were conducted using JMP Pro statistical software on the 746 candidate records. Pairwise correlations were low to moderate between the 3 controls and the 20 independent variables, with the strongest correlation ($\underline{r}$ = .56) between the personality scales Complexity and Breadth of Interest, both JPI-R sub-facets of Openness. Moderate to strong correlations were also found between Verbal and

Performance IQ (r = .53), and between the physical events Pushups and Sit-ups (r = .36). All pairwise correlations were substantially lower than the recommended cutoff of |.7| (Wilson et al., 2012), indicating a lack of multicollinearity concerns in modeling all independent variables. The Correlation Matrix is shown in Table 2.

4.3 By-week Attrition Profiles

Attrition profiles were developed by calculating the mean scores on each of the predictor variables for those candidates who attritted during week 1, 2, 3, 4, and weeks 5-8, as well as for passing candidates. Results are shown in Table 3. Analysis of the mean scores shows that those attritting candidates had consistently lower physical ability scores than passing candidates. Other predictors did not exhibit clear patterns in relation to passing candidate scores. These results suggest that greater physical ability allows candidates greater success during the entirety of the course, while other predictors' importance may vary by week, depending on the curriculum demands.

4.4 Validity Coefficients

Bivariate correlations of each of the independent variables with the dichotomous pass/fail outcome were computed for three timeframes: (1) for the full 8-week course, (2) for week 1 pass/fail only, and (3) for pass/fail results during weeks 2-8. Results are shown in Table 4. Regardless of the time period, the three physical predictors and the control variable of Class Pass Rate were always significant at $p < .01$, and the physical predictors were always the strongest bivariate predictors. For example, Pushups demonstrated the strongest correlation, ranging from a high for total attrition (r = .30), to a low during week 1 attrition (r = .24), while Run and Sit-ups ranged from a high of r= .26/ r = .24 to a low of r = .20/r = .19, respectively. Interestingly, Candidate Age was also significant at $p < .05$ or less at each time period, demonstrating small correlations (r = .12 to .09).

Beyond the physical events, the cognitive and personality validity coefficients were less consistent between the time periods and accounted for less variance than physical abilities in attrition rates. Performance IQ, for example, was the only significant cognitive predictor, and it was only significant for the full 8-week attrition (r = .10) and for attrition from weeks 2-8 (r = .10). Of the personality predictors, four scales were significant for all three attrition time frames. Organization, a sub-facet of Conscientiousness, had the strongest correlations (r = .19 to .07) and was significant during all three time periods. Energy level and Social Confidence, both components of Extraversion, were also significant for all three time periods, and demonstrated correlations ranging from r = .12 to.09, and r = .12 to .08, respectively. In addition, Sociability, another sub-facet of Extraversion, was significant during all but week 1, with a moderate correlation (r = .13) with attrition outcomes from the entire course.

None of the other predictors, whether controls, verbal IQ, or the other 11 personality scales, demonstrated significant validity coefficients for more than one timeframe. Differences in the correlational strength of bivariate correlations suggest that attrition from week 1 results from different reasons than attrition from weeks 2-8.

4.5 Full Logistic Regression Model

For the three time periods, attrition rates were regressed on all predictors—the three control variables, three physical predictors, two cognitive predictors, and 15 personality scales. Results from each of the three timeframes demonstrated that when all variables were considered, only Class Pass Rate, Pushup Score, Run Score and Sit-up Score were consistently significant at p <

.05. For the full 8-week timeframe, Performance IQ, and personality scales Organization and Tolerance were also significant at $p < .05$. For week 1 timeframe, personality scales Energy Level, Cooperativeness, Sociability, and Social Confidence were significant at $p < .10$, while none of the cognitive predictors were. For the Week 2-8 timeframe, personality scales Organization and Anxiety (negative correlation) were significant at $p < .05$, while cognitive predictors again were not. Across the three timeframes, many of the variables that were significant and relatively stronger in the bivariate analysis were no longer significant in the multivariate model, potentially indicating a problem with shared variance. Results of the full 8-week model are shown as an example in Table 5. A factor analysis was conducted to reduce the dimensionality of the personality scales.

4.6 Exploratory Factor Analyses

Exploratory factor analyses (EFAs) were conducted on the 15 personality scales to identify the best underlying factor structure and to compare the results from this unique military sample to previous factor analysis results (Jackson, 1977; Erdle et al., 1992; Paunonen and Jackson, 1996; Doster et al., 2000). Although prior factor analytic results for the JPI-R were based on item-level data, only scale-level data were available for the military sample. Maximum likelihood estimation with both orthogonal and oblique rotations were considered. Determining the similarities between the military sample and previous factor analyses of the JPI-R was a subjective criterion used when completing multiple iterations of factor analysis. As with the non-military samples, the initial goal was to recover the big five factor structure. In the current EFAs, solutions were objectively evaluated using the Tucker and Lewis Index (TLI; cut-off >.95), and the Root Mean-Squared Error of Approximation (RMSEA; cut-off < .06) (cf. Brown, 2015).

Analysis began with all 15 personality scales. The eigenvalues and scree plot for this factor analysis indicated support for 5 factors 64.3 percent of the variance, but demonstrated poor fit indices, with a TLI of .87 and a RMSEA of .07. Regardless of the rotation method used, the multiple personality scales cross-loaded across several factors. The second iteration was based on CFA finding of Paunonen and Jackson (1996); they found that eliminating three scales—Risk-Taking, Energy Level, and Orthodoxy—produced the best fitting five-factor structure. However, the EFAs with the military sample did not fit well with the five-factor solution (TLI: .90, RMSEA: .062).

After several different models were tested, a five-factor model with a Tucker and Lewis index of .954 and RMSEA of .039 was chosen as the best fitting model. A three-factor solution also produced an acceptable TLI and RMSEA, but the five-factor solution had slightly better fit statistics and was more consistent with the a priori expectations of the internal structure. The chosen solution did not include personality scales that failed to factor into theoretically supported clusters or that failed to demonstrate sufficiently high factor loadings (.4 or greater), and resulted in 11 scales being retained that accounted for 69.5% of the variance. Innovation, Anxiety, Social Confidence, and Responsibility were not included in the five-factor solution. This factor loading matrix is presented in Table 6.

In order to incorporate the factor analysis into logistic regression models, two composites were created. The first factor was similar to the Openness FFM dimension, and a composite was created by averaging scores for Complexity, Breadth of Interest, and Tolerance. The third factor was similar to the Conscientiousness FFM dimension, and a composite was created by averaging the scores for Scales of Energy Level, Organization, and Traditional Values. The scale for Responsibility did not load on this Conscientiousness factor, and that is consistent with previous studies of the JPI (Detwiler & Ramanaiah, 1996; Paunonen and Jackson, 1996). No other composites were created, based on factors two, four, and five only having individual scales with greater than a .4 loading, as well as little theoretical support for the low loading scales to be included in the factors. The use of two composite scores reduced the number of personality predictors from 15 to 11.

4.7 Logistic Regression Models Using Personality Composite Scores

To determine the final model, logistic regression models were fitted in an iterative fashion by adding and subtracting combinations of predictors to produce a solution with all predictors significant at $p < .05$ level, and the highest resulting ROC AUC for predicting successful outcomes. Given the future availability of additional data to cross-validate prediction models, the goal of the iterative process was to find the model for each time period with the ROC AUC closest to 1 for predicting candidate success, while including only significant predictors. For each considered time period, a full model consisted of 3 controls, 3 physical predictors, 2 cognitive predictors, and 11 personality predictors. After the first two iterations, Jackson's (1977; Paunonen & Jackson, 1996) previous work as well as the results of previous studies on attrition (Barrick & Zimmerman, 2009, e.g.) informed subsequent iterative attempts to try and find a model with an increase the ROC AUC through different combinations of significant predictors. For example, personality scales that were close to significance in the initial analysis, or that had significant bivariate relationships during the same attrition timeframe, or that had low shared variance with variables already in the model, but that had a theoretical logic for inclusion, were added and logistic regression model fit was analyzed. This process was continued until it was determined that any additional predictors would overfit the model, indicated by some predictor or several predictors becoming statistically insignificant.

4.7.1 Full 8-week Course. The first timeframe analyzed was success from the 8-week course where 338 candidates successfully completed the course and 408 candidates attritted for various reasons. Fitting a logistic regression model to predict candidate success resulted in a model with a ROC AUC of .788, with the significant predictors being Class Pass Rate (Odds Ratio (*OR) = 1.080),* Run Score (*OR* = 1.066), Pushup Score (*OR* = 1.060), Conscientiousness composite (*OR* = 1.055), Sit-ups score (*OR* = 1.034), and Performance IQ (*OR* = 1.022). This model had sufficient sample size and power, and results are found in Table 7. After several iterations of model fitting, the best model for the 8-week timeframe included 8 predictors: Class Pass Rate (*OR = 1.081*), Run Score (*OR = 1.068*), Pushup Score (*OR = 1.060*), personality Conscientiousness composite (*OR* = 1.060), Sit-ups Score (*OR* = 1.035), personality Openness composite (*OR* = 0.969), Performance IQ (*OR* = 1.024), and Sociability (*OR* = 1.019). The ROC AUC of this model was .784, a similar value to the model including all predictors, see Table 8**.**

Effects of the predictors ranged from weak but practically meaningful effects to strong effects, based on the use of range odds ratios for the sample, and the effect size criterion of 2 for a practically meaningful effect, 3 for a moderate effect, and 4 or larger for a strong effect (Ferguson, 2009). Weak effects were observed for Sociability scores (Range OR = 2.34), while moderate effects were observed for Openness composite scores (Range OR = .31/3.22*; * denotes the inverse odds ratio, listed for ease of comparison with variables with a positive relationship) and Sit-up scores (Range OR = 3.46). Strong effects were observed for Performance IQ (Range OR = 4.51), Run Scores (Range OR = 7.15), Conscientiousness composite scores (Range OR = 8.33), Pushup Scores (Range OR = 8.60), and Class Pass Rate (Range OR = 7.59). A confusion matrix demonstrating the predicted vs. actual outcomes for this model, assuming an equally-weighted error rate, is in Table 9. A model probability cutoff of 0.48

resulted in a balance of type I and type II prediction errors, and a model sensitivity, or recall, of 0.69, as well as a specificity, or selectivity rate of 0.74, for this sample. The F_1 score for this model and this cutoff was .69.

4.7.2 Week 1. For week 1, 601 candidates successfully progressed and 145 attritted. The full model had a ROC AUC of .747. See table 10 for results. The significance of predictors for this model was determined using a type I error rate of .10 due to power, and included Run Scores (*OR* = 1.064), Class Pass Rates (*OR* = 1.041), Responsibility scores (*OR* = 1.040), Openness composite scores (*OR* = 0.963), Pushup Scores (*OR* = 1.035), Cooperativeness scores (*OR* = 1.030), Sit-ups scores (*OR* = 1.027), and Social Confidence scores (*OR* = 1.026). Iterative attempts at optimizing the number of predictors for the highest resulting ROC AUC resulted in an eight-predictor model with a ROC AUC of .730. The reduction in predictors resulted in sufficient power, based on the sample size (746), base rate (.194), and final number of predictors (eight), and allowed sustaining a type I error rate of .05. Notably, the cognitive predictors were not significant in this time period and neither was the Conscientiousness composite. The results of this model are shown in Table 11.

Effect sizes based on range odds ratios (Ferguson, 2009) varied from small and practically meaningful such as Sit-up scores (Range OR = 2.63) and Class Pass Rates (Range OR = 2.85), to moderate effects from Social Confidence scores (Range OR = 3.05), Pushup scores (Range OR= 3.51), and Cooperativeness scores (Range OR = 3.61). Openness composite scores (Range OR = .241/4.15*), Responsibility scores (Range OR = 5.51) and Run scores (Range OR = 6.43) had strong effect sizes during this timeframe. A confusion matrix for this model is shown in table 12. For this timeframe, a model probability cutoff of .71 resulted in balanced type I and type II prediction errors, and model sensitivity, or recall, of 0.86, as well as a specificity, or selectivity rate of 0.45, for this sample. The F_1 score for this model and cutoff value was .86.

4.7.3 Weeks 2-8. The full logistic regression model for week 2-8 success resulted in a ROC AUC of .769. The model had sufficient statistical power. The significant predictors were Class Pass Rates (*OR* = 1.085), Pushup scores (*OR* = 1.057), Conscientiousness composite scores (*OR* = 1.055), Run scores (*OR* = 1.054), and Sit-ups scores (*OR* = 1.028). Results of this model are found in Table 13. Further iterations found that the personality scales of Anxiety and Cooperativeness could be included in the model, and if Performance IQ was added the exact probability of a Type I error = .059. Given the plans for cross-validation of the prediction model, it was decided to retain Performance IQ in the final, eight predictor model. The eight predictor model included Class Pass Rate (*OR* = 1.085), Run score (*OR* = 1.056), Pushup score (*OR* = 1.056), Conscientiousness composite (*OR* = 1.054), Anxiety (*OR* = 0.970), Sit-ups score (*OR* = 1.028), Cooperativeness (*OR* = 1.029), and Performance IQ score (*OR* = 1.107). This model had a ROC AUC of .763. Full results of this model's parameter estimates and odds ratios are found in Table 14.

During this timeframe, Performance IQ scores (Range OR = 2.83) and Sit-up Scores (Range OR = 2.67) had a small but practically meaningful effect, while Cooperativeness scores (Range OR = 3.37), and Anxiety scores (Range OR = .31/3.19*) had moderate effects. Large effect sizes were found for Run Scores (Range OR = 5.20), Conscientiousness composite scores (Range OR = 7.01), Pushup Scores (Range OR = 7.54) and Class Pass Rates (Range OR = 8.27). A confusion matrix for this model is shown in table 15. For this timeframe, a model probability cutoff of .53 resulted in balanced type I and type II prediction errors, and model sensitivity, or recall, of 0.75, as well as a specificity, or selectivity rate of 0.66, for this sample. The F_1 score for this model and cutoff was .85

4.8 Comparison of Best Models

Results of the logistic regression modeling over the three timeframes provide insights to research questions 1, 2, and 3, particularly in regards to the pattern of results of predictors in different models. First, it is interesting that all three physical ability scores were significant across all three timeframes given that week 1 training features no critical physical event tests. Second, cognitive abilities were not a consistent predictor of success; only performance IQ was retained in the final model for two timeframes: for 8-week success, and week 2-8 success. The models differed most in the different personality predictors included in the final models and their effect sizes. Week 1 success was predicted by Responsibility scores and the Openness composite scores (negative relationship) as well as Cooperativeness scores and Social Confidence scores. Success in completing weeks 2-8 was predicted by the Conscientiousness composite scores, as well as Anxiety scores and Cooperativeness scores. Meanwhile, eight-week success was predicted by the Openness composite scores (again, negative relationship), Conscientiousness composite scores, and Sociability scores; the latter scale scores were not significant in the other two prediction models. The models also differed in the optimum cutoff score used to balance equal outcomes of a type I vs. a type II prediction error, as well as their F_1 scores. The 8-week model's cutoff score of .53 balanced prediction errors and resulted in a true positive rate (TPR) of .69, a true negative rate (TNR) of .74, with a F_1 score of .69. The week 1 model's cutoff score of .71 resulted in a TPR of .86, a TNR of .45, and a F_1 score of .86. The week 2-8 model's cutoff score of .53 resulted in a TPR of .75 and a TNR of .66, with a F_1 score of .85.

4.9 Personality Profile Analysis

A trait profile analysis was conducted, following the inferential accuracy method used by Jackson (1972; Reed and Jackson, 1975; Hauenstein & Alexander, 1991), to predict attrition as a function of a candidates' similarity to personality norms. Attrition was modeled as a function of sensitivity scores operationalized as each candidate's correlation with normative scores on the 15 JPI-R personality dimensions. Threshold was operationalized as the sum of the absolute values of each candidate's mean deviation from the 15 JPI-R normative scores. Three sets of JPI-R norms were explored (See Table 16); mean JPI-R dimension scores for all candidates (n = 748), mean JPI-R scores for those candidates who successfully completed training (n = 340), and mean JPI-R scores for candidates from a separate organizational leader selection and training program (n= 1000), that all mid-level leaders must complete.

In terms of threshold, candidates' total mean differences in personality scales were normally distributed regardless of the comparison sample, with kurtosis ranging from 0.63 to 0.97 and skewness ranging from 0.65 to 0.71, spanning from a low mean difference of 41.0 to a high mean difference of 221.7 across the three sets of norms considered. Sensitivity was measured by analyzing the correlation of each candidates' personality T-scores to a normative mean T-score. Correlations were normally distributed with kurtosis ranging from -0.95 to 0.44 and skewness ranging from -0.90 to 0.50, a maximum r = .94, and a minimum of r = -.59. Figures 3-6 provide examples of how candidates might vary on inferential accuracy. For example, candidate #408 (see Figure 5) exhibited weak inferential accuracy with a moderate *negative* sensitivity score (r = -.44) and a high, less accurate, threshold score (d =132). Candidate #409 (see Figure 4) exhibited strong inferential accuracy with a high sensitivity score (r = .91) and low threshold score (d = 56). Candidate #557 (see Figure 3) exhibited mixed inferential accuracy a with a low, negative sensitivity score (r = -.23), and a relatively low threshold score (d = 83). Meanwhile, Candidate #708 (see Figure 6) exhibited opposite mixed results, with a high sensitivity score (r = .84), but also an inaccurate, high threshold score (123).

After descriptive statistical analyses, a bivariate correlation analysis was conducted for threshold and sensitivity, using the three sets of norms, and for each of the three time periods as shown in Table 17. Threshold was not a significant predictor for any set of comparison norms nor for any time period. In contrast, sensitivity was a significant predictor for all norms during the 8-week and weeks 2-8 time periods, but not for week 1. Sensitivity correlations were modest, ranging from a high (r = .120; p < .01) for passing candidate norms and for week 2-8 time period, to a low (r = .086; p < .05) for leader norm comparison for the 8-week period. Passing candidate norms produced the strongest correlations for each time period. The range of point biserial correlations for sensitivity was similar to the bivariate relationships between T-scores for personality scales relating to Extraversion and Conscientiousness, while they were considerably lower than the bivariate correlations for physical events, and similar to the bivariate correlations for performance IQ as shown in Table 4.

Multivariate logistic regression models were analyzed for each of the three timeframes, and for each set of comparison norms, to compare the personality trait profile models with the logistic regression models using personality T-scores. Results are shown in Table 18. These analyses substituted sensitivity and threshold for candidate personality T-scores, and incorporated all controls to develop a full model just as in the previous logistic regression models. A final model was developed that included only significant predictors. This was done to compare whether the use of personality scales or personality profile similarity predicted attrition better. A comparison of multivariate logistic regression models suggested that use of personality

T-scores predicts slightly better than use of threshold and sensitivity. When modeled using all predictors—controls, physical predictors, cognitive predictors, and the sensitivity and threshold values, the 8-week attrition profile model resulted in a ROC AUC of .771 in comparison with the T-score personality model with the highest ROC AUC of .784. A confusion matrix and resulting diagnostics is shown in Table 19. Results were comparable to the T-score personality model. The profile model had the same probability cutoff (.48) for equalized type I and II errors, and its TPR of .68, TNR of .72, and F_1 score of .67 were all similar. The difference in effect size between the two models is small (Rice & Harris, 2005), but potentially relevant. Operationalized, the difference is the accurate prediction of 19 successful graduates based on the annual average of 1485 candidates per year, assuming similar cutoffs are used.

4.10 Interaction Effect of Physical Ability and Personality

An interaction effect was analyzed between ability and personality, using the strongest ability predictor, which was physical ability. Full APFT score, an aggregate sum of a candidates' scores on the pushup, sit-up, and run event, was used as the ability predictor. For personality predictor, a candidates' sensitivity (correlation) coefficient was used because it was the most comprehensive single predictor representing personality. Multiple logistic regression analysis was conducted including the physical aggregate, the personality sensitivity, the class pass rate as a control variable, performance IQ, and an interaction variable between physical ability and personality. The interaction was non-significant and this model yielded a ROC AUC lower than any of the other methods examined.

5.0 Discussion

The results of this study suggest answers to the research questions areas for future research. It is worth highlighting that a goal of this study was to find the best predictive model, given the archival data on independent variables linked with attrition outcomes. In obtaining the best predictive model, possible over-fitting and risking slightly higher Type I error rates were accepted if inclusion of a predictor was logically and theoretically supported because the final prediction models will be cross-validated. To summarize the prediction model findings for all three time periods, Table 20 indicates the significant predictors for the full models and the predictors retained in the final models, where each independent variable's effect size is noted for the various models. Variables with weak effect sizes are annotated with "W", while moderate and strong effect sizes are represented by "M" and "S", respectively.

The final model using personality T-scores and personality composites based on factor analysis, was that of the 8-week timeframe, based on having a ROC AUC closest to 1, although this model had the lowest F_1 score of the three timeframes. The final model for week 1 attrition had the lowest ROC AUC, but the highest F_1 score. The week 2-8 final model had the second highest ROC AUC and a F_1 score nearly as high as the week 1 model. These findings are not surprising given that the attrition rate for the 8-week program was closest of the three timeframes to a 50% drop-out rate, and week 1 attrition was furthest from the 50% drop-out rate. For each of the timeframes, the final model had eight predictors, but both the predictors and the relative weights differed between the timeframes.

5.1 Research Question 1

The first research question addressed the extent to which predictors of week 1 attrition differ from predictors of attrition from weeks 2-8. This was of interest because of previous work studying police academy attrition found evidence that early vs. late attrition was due to different reasons (Ryan et al., 2000), and because week 1 in this study had the highest attrition among all weeks, representing almost 20 percent of the total attrition.

First, there were several predictors consistent across week 1 and weeks 2-8. For example, it was notable that all three physical variables predicted for week 1 and for weeks 2-8, despite a physical ability screening as a prerequisite for entry, and despite week 1 having no physical critical events eliminating candidates. The physical predictors are discussed in greater detail below in answering research question 3. The control variable of class pass rate also predicted for both time periods, with a weak effect size for prediction of week 1 attrition and strong a strong effect size for prediction week 2-8 attrition. Cooperativeness, a personality score associated with Agreeableness, also predicted during both timeframes, with a moderate effect size in each. A final similarity is that constructs related to conscientiousness were retained in both time frames. Responsibility scale scores had strong effect sizes in the week 1 model, and the conscientiousness composite scores had strong effect sizes in the week 2-8 model. These results were consistent with previous findings regarding the consistent relationship between conscientiousness and continuance commitment (Erdheim et al., 2006)

There were interesting differences between the predictors of each time period in the cognitive and personality scores. Performance IQ, typically associated with training performance (Schmidt & Hunter, 2004), but not attrition (Topp and Kardash, 1986; Griffeth et al., 2000) did not predict week 1 attrition, but it did have a weak effect size on attrition from weeks 2-8. This is understandable, given the focus on training and learning new skills during weeks 5-8 of the course. Further differences existed in the personality predictors. Week 1 success was strongly predicted by the openness composite scores (negative relationship) and moderately predicted by social confidence scores, neither one of which predicted for weeks 2-8. These results are in line with previous research finding a negative relationship between openness with organizational commitment (Erdheim et al., 2006). Meanwhile, success from weeks 2-8 was moderately predicted by anxiety scores (negative relationship).

The meaningful differences in personality predictors between the two time periods suggest the causes of attrition vary with the point in time that attrition occurs. While the personality differences alone are insufficient to confidently conclude the mediating constructs, the moderate effect of social confidence scores in predicting week 1 success is not inconsistent this those of earlier studies that found perceptions of fit and of self-confidence (Schmitt & Ryan, 1997; Kristoff-Brown, Zimmerman, & Johnson, 2005; Moran et al., 2011; Kiernan, 2005) influence attrition early in selection programs. Further, the strong effect of the openness composite in predicting week 1 attrition supports previous findings that perceived alternatives (Schmitt & Ryan, 1997) and openness to experience (Zimmerman, 2008) are related to voluntary attrition and negatively related to organizational commitment (Erdheim et al., 2006). In this study, the combination of openness and social confidence in predicting week 1 attrition, even when no training events have yet eliminated candidates, may also suggest that candidates' self-efficacy (Moran et al., 2011) influences week 1 attrition. In Bandura's (1977, 1982, 1986) model of self-efficacy, previous performance accomplishments most strongly influence an individual's confidence that they can succeed in a future task. Therefore, it is likely that candidates who perform well in the pre-selection physical diagnostic events and in preparation events have higher self-efficacy, while those lower scoring candidates have lower levels of confidence.

Because week 1 is on the eve of the first physical tests, low efficacy candidates may be more likely to voluntarily attrit to pre-empt failure. A specific self-efficacy study is needed to test this conclusion.

5.2 Research Question 2

Research Question 2 addressed how well a single model of total attrition compares to dual-stage models of attrition in which attrition from week 1 and attrition from weeks 2-8 are modeled separately. The results of the study provide limited insight due to the analysis used. Because of the different base rates of success between week 1, weeks 2-8, and the full 8-week course, a simple comparison of the ROC AUC of each model is insufficient to answer research question 2. The 8-week model had a higher ROC AUC (.784) than either the week 1 (.730) or the week 2-8 model (.763), which is likely explained by statistical power advantages rather than practical advantages in prediction. For example, the 8-week timeframe had the largest sample and the success vs. attrition rate closest to 50%. Therefore, a direct comparison of ROC AUC was determined to be insufficient to answer research question 2 with confidence.

However, when comparing models using prediction cutoffs for equivalent type I and type II prediction errors, a different pattern emerged. The week 1 model had the best F_1 score (.86), followed by the week 2-8 model (.85), while the 8-week model had the lowest F_1 score (.69). The results indicate that the precision and recall of the models for week 1 and weeks 2-8 are superior to the overall model, and may be a more effective way of predicting a candidates' success or failure, although cross-validation and a simulation would be required to empirically test this method.

While a single stage model may be more practically meaningful to the organization, since all attrition represents an opportunity cost, there are reasons why a more thorough future analysis may be helpful. First, any increase in predictive value will aid decisions on pre-selection for the program. Second, a dual-stage model may allow a deeper understanding of the construct differences between week 1 and later attrition, supporting program changes or behavioral interventions to increase success. Finally, the use of multiple models versus a single model of attrition would allow more flexibility in determining model cutoffs. For example, although there was no a priori rationale for a cutoff other than equal type I and type II prediction errors, it may benefit the organization to adjust the probability cutoff scores based on the number of potential candidates available and the number of classes upcoming. Further, because of the way this specific program is structured, it would likely be advantageous to minimize type II errors (false negatives) for week 1, knowing that all the critical testing occurs during weeks 2-8. The intent is to conduct a simulation to address the utility of a single vs. a dual-stage selection decision model after the current prediction models are cross-validated.

5.3 Research Question 3

The third research question addressed the relative contributions of cognitive abilities, physical abilities, and personality in the prediction of attrition. Although intelligence and personality predictors are commonly examined in selection research, it is rare to have a job family where physical abilities are so critical to successful performance. Cognitive ability scores broadly predict both job and training performance well (Hunter, 1986; Schmidt & Hunter, 1998; Schmidt & Hunter, 2004), but cognitive ability scores have not similarly predicted attrition and turnover (Topp & Kardash, 1986; Griffeth et al., 2000). Personality has been shown to predict both an intent to quit as well as actual quitting in meta-analysis of employee attrition (Zimmerman, 2008). However, the military selection context provided a unique opportunity to analyze all three dimensions in a single study.

5.3.1 Physical abilities. As seen in Table 20, physical ability was a strong and consistent predictor of candidate success. Physical ability scores predicted across all three time periods and were the strongest predictors of the three categories of predictors. High levels of physical abilities are critical characteristics of successful military performance, but the strong, consistent predictive ability of all physical scores across the three time periods is surprising given the range restriction on physical ability scores that resulted from pre-screening. Further, week 1 training activities lack a physical critical event test, meaning no candidates are involuntarily failed because of physical performance during week 1. It is clearer that post-week 1 training events tax physical abilities. For example, the land navigation exercise, the multi-day field training exercise, and the combat-focused fitness tests occur during weeks 2-5, but more research is required to understand the relationship between physical abilities and week 1 attrition.

5.3.2 Cognitive abilities. Cognitive ability scores were weighted less than expected in the final models, and they were the weakest category of predictor. Because the 8-week program included training and new skill development, a stronger relationship with cognitive abilities was expected, particularly during weeks 2-8. However, verbal IQ did not predict for any time period, and performance IQ had a weak effect for weeks 2-8, and did not predict for week 1, although it had a strong effect size over the full 8-week course. The results suggest that the course may lack sufficient cognitive loading to allow variance in aptitude to manifest as variance in training performance. For example, although there are knowledge and skill development requirements during the course, the pace, volume, or complexity of events may be too easy to allow performance to be differentiated as a function of cognitive abilities. Further, a review of the course's critical events highlights that there are only two critical events that are primarily cognitively loaded—a written test on unit standards and history, and land navigation (orienteering). Dimensions of learning and problem-solving, highly valued abilities in the job specifications for special forces personnel (U.S. Office of Personnel Management, 2015), may not be adequately reflected in training exercises given that success on most training exercises require superior physical rather than superior cognitive abilities.

5.3.3 Personality. The predictive accuracy of personality scores was greater than typically found when predicting job performance, both in terms of effects sizes and the broad array of significant personality predictors. In every timeframe there was a personality predictor with a strong effect size. In this sample personality predicted better than cognitive abilities, which supports previous research on attrition (Griffith et al., 2000).

Conscientiousness, which factored into separate Conscientiousness composite scores and Responsibility scores, predicted with a strong effect size for each time period, which was consistent with previous research demonstrating a relationship between Conscientiousness and attrition (Zimmerman, 2008; Barrick & Zimmerman, 2009). This is interesting partly because the military sample demonstrated higher T-scores than U.S. population norms on Conscientiousness-related scores, meaning the relationship with attrition was strong, despite range restriction.

Other personality facets returned surprising results that contrasted with previous research. Most surprisingly, unlike prior research (e.g., Zimmerman, 2008), personality scales associated with Neuroticism did not consistently predict attrition. Anxiety, a JPI-R facet of Neuroticism, was the only related scale that significantly predicted attrition, and it was only during the week 2-8 timeframe (Range OR = .26/3.85 inverse), for a moderate effect.

The Openness to experience composite, meanwhile, predicted moderately for the 8-week timeframe and strong (negative relationship) for week 1 success. This was interesting, and in contrast to typical findings that Openness to experience is typically a poor predictor of organizational criteria. However, this supports previous findings by Zimmerman (2008, p. 333) that high openness does relate to higher quitting, suggesting that those high in Openness may be more prone to seek out other opportunities. It may also be grounds for caution; it is possible that those candidates with the highest Openness scores self-select out at higher rates due to boredom or dissatisfaction with the rote requirements of the program, and those candidates high on openness may perform well at higher levels of the organization (Li et al., 2014).

Agreeableness and Extraversion scores were less predictive in general, but with some exceptions. For example, Social Confidence moderately predicted week 1 success, while Sociability weakly predicted success from over the 8-week timeframe. These two scales related to extraversion were the only extraversion sub-facets that predicted, and only during one of the three potential timeframes. Meanwhile, Cooperativeness was the only scale score related to agreeableness that predicted candidate success. However, it predicted with a moderate effect size for both week 1 and for week 2-8 timeframes. The inconsistent relationships found with those scales related to agreeableness and extraversion support previous research findings that those facets' relationships with attrition are less generalizable than conscientiousness and emotional stability (Zimmerman, 2008).

Overall, the personality findings are stronger than found in most organizational contexts. First, the moderate to strong prediction of the Conscientiousness and Openness composites suggest that properly factored personality composites are more effective predictors than cognitive ability in this context. Second, the different personality scales that predict for different time periods suggest that attrition in different weeks results from different psychological reasons, although more research is needed to determine what constructs mediate attrition.

5.4 Research Question 4

The fourth research question addresses how well a candidates' personality similarity to an organizational exemplar predicts attrition. The question was posed to explore the implications of Schneider's (1987) A-S-A model, where individuals with similar personality traits are attracted and selected at higher rates than others (Kristoff, 1996; Oh et al., 2017). Personality scale scores narrowly predicted success better than the profile metrics from the inferential accuracy model (Jackson, 1972; Reed and Jackson, 1975). When personality profile sensitivity and threshold were used to predict success, the highest ROC AUC was slightly lower than the models that used T-scores (ROC AUC .771 vs. .784). Other model diagnostics were also similar. This model had the same .48 probability cutoff for equalized type I and type II prediction errors, and its TPR of .68, TNR of .72, and F_1 score of .67 are comparable to the model using personality T-scores. The similarity in the ROC AUC values suggests that there may be value in a personality trait profile approach. Using a more complex type of profile analysis, or using a better sample for an organizational exemplar personality profile could more conclusively support Schneider's A-S-A Model (Schneider, 1987; Schneider et al., 1998), and subsequent studies demonstrating the importance of candidates' similarity to organizational norms in the selection and socialization phase of employment (Kristoff, 1996; Oh et al., 2017). Despite the simplicity of the inferential accuracy model and the availability of more complex methods of profile analysis, the conclusions potentially suggest that in this sample, processes contributing to organizational homogeneity are already occurring during assessment and training.

5.5 Measurement Structure of JPI-R

An unintended but potentially interesting contribution of this study lies in the factor analysis results. The JPI-R has been validated in previous studies on several samples (Jackson, 1977, Erdle et al., 1992; Detwiler & Ramanaiah, 1996; Paunonen & Jackson, 1996; Doster et al., 2000), and found to factor into 5 or 6 facets strongly resembling the NEO-PI-R five factor model. However, previous factor analyses used item-level data, whereas this study relied on scale-level data. Of note, the military sample scales did not factor similarly into the five personality clusters—Analytical, Emotional, Extroverted, Opportunistic, Dependable—generated by the JPI-R (Paunonen & Jackson, 1996). Although personality scales mean differences between military samples and the U.S. population are well known, an interesting question that has not been examined is whether personality scales are invariant across the general population and the military sub-population. Regardless of the cause, the differences in factor structure suggest that there are interesting personality differences between the military sample and typical study populations, and imply opportunities for future research into personality traits that present in both the military as a whole, and special operations forces as an even smaller sub-population.

5.6 Implications

The study offers implications for the military selection context. The military may benefit from a re-evaluation of its pre-screening relative to job specifications, to determine if the first hurdle in the assessment process captures the knowledge, skills, abilities and other characteristics (KSAOs) needed for success in the organization, and to determine if the assessments are psychometrically sound. For example, the organization collects many individual differences measures on candidates during pre-screening only to fail to use these assessments in decision-making. Meanwhile, during the selection through training process, i.e., using training performance as a second stage selection hurdle, it may be useful to evaluate the extent to which the curriculum is too heavily weighted toward a subset of KSAOs, such as physical events. Overweighting specific KSAOs during training creates criterion deficiencies that likely increase false negative rates. Stated another way, it increases the risk that candidates who would be ultimately successful in the organization drop out during training due to relative weaknesses in only one particular area.

Alternatively, if the current assessment and training program accurately weights the KSAOs necessary for successful organizational performance, an increase in success rates will be achieved if more stringent screening cut-offs are used on the physical ability assessments. For example, the range odds ratio between high physical scores and low physical scores all exceed 7, indicating those currently allowed at the bottom of physical ability scores are more than seven times less likely to pass than the physically strongest candidates. Delaying the entry of candidates until they can meet a higher physical performance threshold would result in lower attrition.

A more direct implication for the military selection context is the potential to use a multidimensional composite score as either a risk management cut-off or a top-down pre-selection tool, once cross-validation of the prediction models is complete. For example, if a pool of 250 potential candidates is on the ground awaiting the next class, once pre-screening is complete, the 8-week model prediction formula could be used to select which 165 candidates gain entry:

$1 / (1+e^{(-(-23.5+0.078*(\text{classpassrate})+0.058*(\text{pushup\_score})+0.034*(\text{situps\_score})+0.066*(\text{run\_score})+0.024*(\text{performance\_IQ})-0.031*(\text{factor1\_openness})+0.058*(\text{factor3\_consientiousness})+0.019*(\text{MTS8\_sociability})))})$. In this respect, the prediction model's weighted inclusion of physical, cognitive, and personality factors would represent a mechanical prediction tool that could be used to mitigate clinical prediction errors (Dawes, Faust

& Meehl, 1989), and likely predict better over a large sample (Grove et al., 2000) than the expert clinical prediction currently used.

In its potential use as a pre-screening tool, the model probability cutoff could be adjusted based on the number of candidates available relative to course capacity. For example, during times when a large number of candidates are awaiting a course, a higher probability cutoff could be utilized because of cost of a type I error (false positive) is higher. In contrast, during times when fewer candidates are awaiting a course, the cost of a type II error (false negative) may be higher, particularly if it means not utilizing full course capacity. After cross-validation, a table of cutoffs and resulting confusion matrices and F_1 scores will be created to inform an organizational utility vs. risk decision on cutoffs.

Implications of the current study extend beyond the military context. Particularly in light of the current emphasis on police training, the results demonstrate that personality scales are effective predictors of success vs. attrition, even when selection and training include a significantly weighted physical component, or a component of risk and hardship. However, previous studies of attrition from police selection and training programs (Topp & Kardash, 1986; Ryan et al., 2000; Hewgley, 2013) have been inconsistent in finding relationships between personality measures, performance, and attrition. Further, they rarely incorporate physical predictors and criterion into their analysis, leaving open the possibility that significant results have been found in the absence of controls for the physical job requirements.

5.7 Limitations

The lack of item-level responses for the JPI-R prevented a finer examination of the specific items' contributions to the scales' measurement structure. This limitation also prevented measurement invariance analysis, which may have provided insights into how prospective candidates conceptualize the personality facets differently than other studies' samples. Based on mean differences between this sample and U.S. population norms, and the inability to reproduce the JPI-R factor structure, it is possible that the JPI-R is measuring different constructs in a military candidate population.

A second limitation related to the archival data was the ability to reliably analyze reasons for attrition. Reasons for attrition were included in the data records, but these data were not trustworthy. Reliable data regarding reasons for attrition would allow greater understanding of the attrition process for two reasons. First, candidates sometimes must dropout due to physical injuries, an unknown number of which are faking physical injury rather than quitting. Although physical abilities may be related to avoidance of injury, the more interesting issue is the proportion of voluntary attrition that is recorded as a physical injury or another reason that does not require the candidate to openly admit quitting. Second, reliable coding of reasons for voluntary attrition would have allowed testing mediational models to clarify the psychological processes that connect personality and aptitude to attrition.

A third limitation was related to the analysis methods. Although separate attrition models were developed for 8-week attrition, week 1 attrition, and attrition from weeks 2-8, a simple comparison of the ROC AUC and F_1 scores of these three models based on a common cutoff score is insufficient to conclude whether a single model or a dual- stage model of attrition would best predict. Because of the differential attrition rates between week 1 and weeks 2-8, and resulting statistical power, a simulation on the data set and a subsequent validation sample would be more appropriate to compare a single versus a dual-stage model.

5.8 Future Directions

Despite the richness of physical ability, cognitive ability, and personality data on candidates, there remains a significant amount of unexplained variance in predicting attrition from the special operations selection through training program. Other potential predictors of attrition should be explored and include bio data (Barrick & Zimmerman, 2005; Hewgley, 2013), self-efficacy (Bandura & Adams, 1977; Bandura, 1986), and mental toughness/hardiness/grit (Kelly et al., 2014), which has been shown to be predictive of West Point cadet performance during initial training and through the first year.

In addition to investigating the usefulness of additional predictors, future work should establish psychometrically sound criterion measures for the selection through training program. While the attrition data was a highly reliable dependent variable, the dichotomous outcome reduced the statistical power to find relationships that may have existed with continuous variable criterion. Establishment and implementation of continuous criterion based on job-related performance scales would allow more powerful statistical analyses to be used, potentially uncovering additional small, but practically meaningful relationships.

Finally, future work could focus on the design and implementation of practical interventions to increase candidates' success rates without changing job-related course standards. An example of this might be a self-efficacy intervention in the weeks prior to candidates' course start, where candidates incrementally gain confidence performing similar tasks, or through vicarious experiences of watching others, or through deliberate positive reinforcement from cadre. Though not the same context, a similar intervention improved candidate self-efficacy and improved recruitment for the Israeli Defense Force Special Forces (Eden & Kinnar, 1991). Another example may be a pre-course resiliency intervention in which candidates are taught goal-setting and mental self-regulation facets of the Army's Master Resilience Training (Reivich et al., 2011). More can be done to better understand, as well as better influence, the attrition rates of military, police, and other high-risk occupations.

6.0 Conclusion

An analysis of archival data from a special operations assessment and training program provided answers to several research questions related to the prediction of attrition. Although there were more similarities than differences in the prediction of early versus late attrition, there was ample evidence of differences in terms of personality predictors reinforcing claims that there are different causes of early vs. late attrition in a selection through training context. The study's incorporation of physical ability predictors and unique physical job characteristics demonstrated the potential for physical ability to dominate prediction in contexts where it is heavily tested and consistently featured. With regard to the usefulness of personality trait profile similarity, the study showed that a nonlinear prediction model should not be discounted. Potentially the most interesting conclusion is that personality scale scores, particularly openness and conscientiousness, may be more important than cognitive ability in a context predicting success in military training.

References

Allen, D. G., Bryant, P. C., & Vardaman, J. M. (2010). Retaining talent: Replacing misconceptions with evidence-based strategies. *The Academy of Management Perspectives*, *24*, 48–64.

Bandura, A., & Adams, N.E. (1977). Analysis of Self-Efficacy Theory of Behavioral Change. *Cognitive Therapy and Research*, 1 (4), 287-310.

Bandura, A. (1986). The Explanatory and Predictive Scope of Self- Efficacy Theory. *Journal of Social and Clinical Psychology*, 4 (1), 359-373.

Barrick, M. R., & Zimmerman, R. D. (2005). Reducing voluntary, avoidable turnover through selection. *Journal of Applied Psychology, 90,* 159–166.

Barrick, M. R., & Zimmerman, R. D. (2009). Hiring for retention and performance. *Human Resource Management*, 48, 183–206.

Binsch, O., Banko, K.M., Bertil, V.J., and Valk, P.J.L. (2015). Examining the Relationship Between Mental, Physical, and Organizational Factors Associated with Attrition During Maritime Forces Training. *Journal of Strength and Conditioning Research*, 29 (11), 187-191.

Brown, T. A. (2015). Confirmatory Factor Analysis for Applied Research, Second Edition. United Kingdom: Guilford Publications.

Carless, S.A. (2000). The Validity of Scores on the Multidimensional Aptitude Battery. *Educational and Psychological Measurement*, 60 (4), 592-603.

Chappelle, W., et al. (2010). Multiple Aptitude Battery – II Normative Intelligence Test Data that Distinguish U.S. Air Force AC-130 Gunship Sensor Operators. *Air Force Research Laboratory-SA-BR-TR-2010-0006*.

Chen, G., Casper, W. J., & Cortina, J. M. (2001). The roles of self-efficacy and task complexity in the relationships among cognitive ability, conscientiousness, and work-related performance: A meta-analytic examination. *Human Performance*, 14, 209–230.

Collins, J.M. (1987). *U.S. and Soviet Special Operations*. Congressional Research Service Report. Retrieved from Website: https://babel.hathitrust.org/cgi/pt?id=mdp.39015039055655&view=1up&seq=5

Colosio, A.L., Fontana, F.Y., & Pogliaghi, S. (2016). Attrition in Italian Ranger Trainees during Special Forces Training Program: A Preliminary Investigation. *Sport Science Health*, 12, 479-483.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PIR) and NEO Five Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674.

Detwiler, F.R.J., & Ramanaiah, N.V. (1996). Structure of the Jackson Personality Inventory from the Perspective of the Five-Factor Model. *Psychological Reports*, 79, 411-416.

Diemer, H. (1917). Causes of "turnover" among college faculties. *The Annals of the American Academy of Political and Social Science, 71,* 216–224.

Douglas, P. H. (1918). The problem of labor turnover. *The American Economic Review*, 8, 306–316.

Doster, J.A., et al. (2000). Stability and Factor Structure of the Jackson Personality Inventory-Revised. *Psychological Reports*, 86, 421-428.

Eden, D. & Kinnar, J. (1991). Modeling Galatea: Boosting Self- Efficacy to Increase Volunteering. *Journal of Applied Psychology*, 76 (6), 770-780.

Erdheim, J., Wang, M., & Zickar, M. J. 2006. Linking the Big Five personality constructs to organizational commitment. *Personality and Individual Differences*, 41: 959-970.

Ferguson, C.J. (2009). An Effect Size Primer: A Guide for Clinicians and Researchers. *Professional Psychology Research and Practice*, 40 (5), 532-538.

Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26, 463–488.

Grove, W.M., et al. (2000). Clinical Versus Mechanical Prediction: A Meta-Analysis. *Psychological Assessment*, 12 (1), 19-30.

Gubata, M.E., et al. (2012). A Noncognitive Temperament Test to Predict Risk of Mental Disorders and Attrition in U.S. Army Recruits. *Military Medicine*, 177 (4), 374-379.

Hardy, J.H., et al. (2017). Are Applicants More Likely to Quit Longer Assessments? Examining the Effect of Assessment Length on Applicant Attrition Behavior. *Journal of Applied Psychology*, 102 (7), 1148-1158.

Hauenstein, N.M.A., & Alexander, R.A. (1991). Rating Ability in Performance Judgments: The Joint Influence of Implicit Theories and Intelligence. *Organizational Behavior and Human Decision Processes*, 50, 300-323.

Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. (2013). Causes and consequences of collective turnover: A meta-analytic review. *Journal of Applied Psychology, 98,* 412–453.

Hom, P. W., & Griffeth, R. W. (1995). *Employee turnover*. Cincinnati, OH: South-Western College Publishing.

Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology, 102*, 530-545.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29(3), 340-62.

Jackson, D.N. (1972). A Model for Inferential Accuracy. *The Canadian Psychologist*, 13 (3), 185-195.

Kelly, D. R., Matthews, M. D., & Bartone, P. T. (2014). Grit and hardiness as predictors of performance among West Point cadets. *Military Psychology, 26*(4), 327–342. https://doi.org/10.1037/mil0000050

Kiernan, M.D., Repper, J., & Arthur, A. (2015). Why do they fail? A Qualitative follow up Study of 1000 Recruits to the British Army Infantry to Understand High Levels of Attrition. *Work*, 52 (4), 921-934.

Knapik, J., et al. (1994). Army Physical Fitness Test (APFT): Normative Data on 6022 Soldiers. *U.S. Army Research Institute of Environmental Medicine*. Report No T94-7.

Kristof, A.L. (1996). Person-Organization Fit: An Integrative Review of Its Conceptualizations, Measurement, and Implications. *Personnel Psychology*, 49. 1-49.

Lee, T.W. & Mitchell, T.R (1994). An alternative approach: the unfolding model of voluntary employee turnover. *Academy of Management Review, 19*, 51-89.

Li, et al. (2014). Retaining the Productive Employee: The Role of Personality. *The Academy of Management Annals*, 8 (1), 347-395.

Maltarich, M. A., Nyberg, A. J., & Reilly, G. (2010). A Conceptual and Empirical Analysis of the cognitive ability-voluntary turnover relationship. *Journal of Applied Psychology*, 95, 1058–1070.

Maertz, C. P., & Campion, M. A. (1998). 25 years of voluntary turnover research: A review and critique. In C. Cooper & I. Roberson (Eds.), *International review of industrial and organizational psychology* (Vol.13, pp. 49–81). New York, NY: Wiley.

Maertz, C. P., & Campion, M. A. (2004). Profiles in quitting: Integrating content and process turnover theory. *Academy of Management Journal*, *47,* 566–582.

Mischel, W., & Shoda, Y. (2010). The situated person. In B. Mesquita, L. F. Barrett, & E. R. Smith (Eds.), *The mind in context* (p. 149–173). Guilford Press.

Mitchell, T.R., Holtom, B.C., Lee, T.W., Sablynski, C.J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal,* 44, 1102-1121.

Moran, D.S., et al. (2011). Prediction Model for Attrition from a Combat Unit Training Program. *Journal of Strength and Conditioning Research*, 25 (11), 2963-2970.

Oh, I., et al. (2017). Do Birds of a Feather Flow, Fly, and Continue to Fly Together? The Differential and Cumulative Effects of Attraction, Selection, and Attrition on Personality-based within-organization homogeneity and between-organization heterogeneity progression over time. *Journal of Organizational Behavior*, 39, 1347-1366.

Peduzzi P, et al. (1996) A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis. *Journal of Clinical Epidemiology,* 49, 1373-1379.

Pope, R.P., Herbert, R., Kirwan, J.D., and Graham, B.J. (1999). Predicting Attrition in Basic Military Training. *Military Medicine*, 164 (10), 710-714.

Rapuano, S.K., Trujillo, A.M., & Crowder, T.A. (2016). The Science of Sit-Ups: An Assessment of Total Physical Fitness. *Infantry*, January-March 2016, 47-51.

Reed, P.L., & Jackson, D. N. (1975). Clinical Judgment of Psychopathology: A Model for Inferential Accuracy. *Journal of Abnormal Psychology*, 84 (5), 475-482.

Reivich, K.J., et al. (2011). Master Resilience Training in the U.S. Army. *American Psychologist*, 66 (1), 25-34.

Rice, M.E., & Harris, G.T. (2005). Comparing Effect Sizes in Follow-Up Studies: ROC Area, Cohen's *d*, and *r*. *Law and Human Behavior*, 29 (5), 615-620.

Ryan, A.M., et al., (2000). Applicant Self-Selection: Correlates of Withdrawal from a Multiple Hurdle Process. *Journal of Applied Psychology*, 85 (2), 163-179.

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

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162-173.

Schmit, M. J. & Ryan, A.M. (1997). Applicant Withdrawal: The Role of Test-Taking Attitudes and Racial Differences. *Personnel Psychology*, 50, 855-876.

Schneider, B. (1987). The People Make the Place. *Personnel Psychology*, 40, 437-453.

Schneider, B., Smith, D. B., Taylor, S., & Fleenor, J. (1998). Personality and organizations: A test of the homogeneity of personality hypothesis. *Journal of Applied Psychology*, 83, 462–470.

Tharwat, A. (2018). Classification Assessment Methods. *Applied Computing and Informatics*, DOI 10.1016/j.aci.2018.08.003

Topp, B. W., & Kardash, C. A. (1986). Personality, achievement, and attrition: Validation in a multiple-jurisdiction police academy. *Journal of Police Science & Administration*, 14(3), 234–241.

U.S. Army Field Manual (FM) 7-22, Army Physical Readiness Training (Washington, D.C.: Department of the Army, 2012)

U.S. Office of Personnel Management Human Resources Solutions. (2015). Job Analysis Technical Report Prepared for the United States Army Special Operations Command.

West, S.G., Finch, J.F., & Curran, P.J. Structural equation models with nonnormal variables: problems and remedies. In: *Hoyle RH, editor*. *Structural equation modeling: Concepts, issues and applications*. Newbery Park, CA: Sage; 1995. pp. 56–75.

Wilson, J.H., Keating, B.P., Beal-Hodges, M. (2012). Regression Analysis: Understanding and Building Business and Economic Models Using Excel (1st ed., Ser. Quantitative approaches to decision making collection). Business Expert Press. https://doi.org/10.4128/9781606494356

Zeidner, M., & Matthews, G. (2000). Intelligence and Personality. In R. Sternberg (Ed.), *Handbook of Intelligence* (pp. 581-610). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511807947.027

Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model. *Personnel Psychology, 61*, 309–348.

Zimmerman, R.D., Swider, B.W., Woo, S.E., & Allen, D.G. (2016). Who withdraws? Psychological individual differences and employee withdrawal behaviors. Under second round review at *Journal of Applied Psychology, 101*, 498-519.

Zimmerman, R.D., Swider, B.W., & Boswell, W.R. (2019). Synthesizing Content Models of Employee Turnover. *Human Resource Management*, 58, 99-114.

Tables and Figures

Table 1

Mean, Standard Deviations, Range, Skewness and Kurtosis statistics of the three control variables and the 20 independent variables

	M	*SD*	*Min*	*Max*	Range	Skewness	Kurtosis
N = 746							
PR	46.9	8.32	33	59	26	-0.14	-1.31
Age	22.4	3.27	18	35	17	1.09	0.97
Time	36.6	22.16	7	189	182	1.57	4.45
PU	87.4	8.98	63	100	37	-0.27	-0.63
SU	91.3	8.04	64	100	36	-0.70	-0.15
Run	93.5	6.85	70	100	30	-1.00	0.50
Verbal	104.2	10.90	75	139	64	0.18	-0.00
Perf	105.0	10.90	71	134	63	-0.05	-0.05
MTS 1	47.4	8.28	28	76	48	0.49	0.10
MTS 2	49.3	9.27	28	70	42	-0.02	-0.76
MTS 3	45.8	8.69	21	64	43	-0.53	-0.09
MTS 4	51.2	10.11	25	71	46	-0.18	-0.61
MTS 5	43.7	8.43	25	70	45	0.27	-0.36
MTS 6	42.2	7.53	27	66	39	0.44	-0.22
MTS 7	42.5	7.65	30	73	43	0.76	0.53
MTS 8	50.8	10.0	29	74	45	-0.17	-0.69
MTS 9	51.9	8.31	22	65	43	-0.85	0.39

Table 1 (continued)

Mean, Standard Deviations, Range, Skewness and Kurtosis statistics of the three control variables and the 20 independent variables

	M	*SD*	*Min*	*Max*	Range	Skewness	Kurtosis
N = 746							
MTS 10	53.0	7.62	27	68	41	-0.66	0.22
MTS 11	46.3	8.73	27	74	47	0.35	-0.20
MTS 12	49.3	9.30	29	72	43	0.02	-0.67
MTS 13	57.5	7.75	24	71	47	-0.83	0.81
MTS 14	55.6	8.05	28	72	44	-0.53	0.19
MTS 15	54.6	7.16	26	69	43	-0.83	0.66

Note. PR = Class pass rate for candidate's class, Age = candidate's age in years, Time = amount of time candidate spent awaiting admission in days, PU = pushups, SU = sit-ups, Run = Run score, Perf= Performance IQ, Verbal = Verbal IQ; Personality Scale labels: MTS 1 = Complexity, MTS 2 = Breadth of Interest, MTS 3 = Innovation, MTS 4 = Tolerance, MTS 5 = Empathy, MTS 6 = Anxiety, MTS 7 = Cooperativeness, MTS 8 = Sociability, MTS 9 = Social Confidence, MTS 10 = Energy Level, MTS 11 = Social Astuteness, MTS 12 = Risk-Taking, MTS 13 = Organization, MTS 14 – Traditional Values, MTS 15 = Responsibility

Table 2

Independent Variables Inter-correlations

	PR	Age	Time	PU	SU	Run	Verbal	Perf	MTS1	MTS2	MTS3	MTS4	MTS5	MTS6	MTS7
PR	1														
Age	.080	1													
Time	-.225	.029	1												
PU	.106	.181	.023	1											
SU	.085	.113	.037	.358	1										
Run	.073	.053	.061	.197	.252	1									
Verbal	-.041	-.059	.071	-.108	.044	.085	1								
Perf	-.008	.155	.093	-.021	.059	.082	.534	1							
MTS1	.006	.067	.034	.008	.025	-.006	.234	.156	1						
MTS2	.003	.221	.035	-.021	.036	-.019	.192	.207	.562	1					
MTS3	-.055	.095	.057	.030	-.050	-.012	.038	.176	.400	.460	1				
MTS4	-.030	.178	.025	.036	.014	.010	.041	.109	.255	.470	.210	1			
MTS5	.025	.000	-.053	.040	-.004	.056	.017	-.046	.177	.181	.116	.088	1		
MTS6	.094	-.200	-.083	-.039	-.071	-.042	-.164	-.227	-.019	-.195	-.101	-.294	.339	1	

Table 2 (Continued)

Independent Variables Inter-correlations

	PR	Age	Time	PU	SU	Run	Verbal	Perf	MTS1	MTS2	MTS3	MTS4	MTS5	MTS6	MTS7
MTS7	.048	-.080	-.035	-.051	-.028	.009	-.041	-.166	-.088	-.059	-.145	-.205	.204	.426	1
MTS8	-.007	.038	-.008	.065	.035	.112	-.034	.033	-.008	.170	.152	.246	.288	-.137	.071
MTS9	-.034	.138	-.003	.090	.059	.123	.069	.123	.034	.216	.317	.233	.075	-.348	-.328
MTS10	-.023	.129	.066	.087	.063	.106	.095	.110	.156	.353	.326	.25`	.054	-.306	-.195
MTS11	-.026	-.091	-.020	-.098	-.044	-.037	.099	.054	.166	.120	.183	-.089	.119	.192	.232
MTS12	.028	-.034	.005	.023	-.011	-.002	.055	.017	.213	.114	.218	-.109	.015	-.001	-.172
MTS13	.008	.230	.001	.113	.077	.101	-.027	.012	-.049	.084	.039	.042	-.049	-.175	-.096
MTS14	-.041	-.007	-.020	-.018	-.022	-.105	-.222	-.101	-.241	-.083	-.073	-.051	.042	-.079	.079
MTS15	-.044	.247	.049	.047	.066	.031	.039	.107	.161	.352	.207	.347	.129	-.274	-.123

Note. PR = Class pass rate for candidate's class, Age = candidate's age in years, Time = amount of time candidate spent awaiting admission in days, PU = pushups, SU = sit-ups, Run = Run score, Perf= Performance IQ, Verbal = Verbal IQ; Personality Scale labels: MTS 1 = Complexity, MTS 2 = Breadth of Interest, MTS 3 = Innovation, MTS 4 = Tolerance, MTS 5 = Empathy, MTS 6 = Anxiety, MTS 7 = Cooperativeness, MTS 8 = Sociability, MTS 9 = Social Confidence, MTS 10 = Energy Level, MTS 11 = Social Astuteness, MTS 12 = Risk-Taking, MTS 13 = Organization, MTS 14 – Traditional Values, MTS 15 = Responsibility

Table 2 (Continued)

Independent Variables Inter-correlations

	MTS8	MTS9	MTS10	MTS11	MTS12	MTS13	MTS14	MTS15
MTS8	1							
MTS9	.457	1						
MTS10	.215	.411	1					
MTS11	-.018	.009	.008	1				
MTS12	.024	.166	.059	.188	1			
MTS13	.149	.299	.315	-.093	-.131	1		
MTS14	.157	.128	.143	-.120	-2.68	.242	1	
MTS15	.163	.191	.365	-.148	-.223	.271	.268	1

Note. PR = Class pass rate for candidate's class, Age = candidate's age in years, Time = amount of time candidate spent awaiting admission in days, PU = pushups, SU = sit-ups, Run = Run score, Perf= Performance IQ, Verbal = Verbal IQ; Personality Scale labels: MTS 1 = Complexity, MTS 2 = Breadth of Interest, MTS 3 = Innovation, MTS 4 = Tolerance, MTS 5 = Empathy, MTS 6 = Anxiety, MTS 7 = Cooperativeness, MTS 8 = Sociability, MTS 9 = Social Confidence, MTS 10 = Energy Level, MTS 11 = Social Astuteness, MTS 12 = Risk-Taking, MTS 13 = Organization, MTS 14 – Traditional Values, MTS 15 = Responsibility

Table 3

By-week attrition profiles

	N	*PU*	*SU*	*Run*	*V-IQ*	*P-IQ*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Wk 1	145	83.7	88.2	90.0	104.2	104.1	47.8	50.1	46.6	52.1	42.7	41.9	41.8	49.7	50.5	51.2	47.2	49.6	56.4	55.6	53.4
Wk 2	57	86.7	88.5	91.7	102.2	100.9	46.7	47.3	45.4	52.0	44.3	41.7	41.6	51.6	53.6	51.9	45.8	51.5	55.4	55.3	54.9
Wk 3	58	83.7	91.7	93.1	106.7	105.5	46.3	49.0	47.3	50.5	43.7	42.2	42.5	49.0	50.5	54.0	46.9	47.9	57.1	55.6	56.1
Wk 4	88	85.9	90.4	93.6	103.6	105.4	47.6	49.0	44.9	51.8	43.0	43.7	41.9	48.3	50.6	51.4	46.3	49.9	56.3	54.6	54.1
Wk 5-8	60	85.8	90.1	92.7	101.9	103.8	49.7	50.1	46.0	50.2	44.7	44.5	44.5	50.4	51.0	53.8	47.3	48.8	55.9	55.9	54.4
Pass	340	90.3	93.4	95.3	104.6	106.3	47.0	49.2	45.5	50.9	44.0	41.6	42.8	52.2	53.0	54.0	45.6	48.9	59.0	55.9	55.0

Note: wk = week of attrition. Pass = passing candidates. PU= pushups score, SU=sit-ups score, Run= run score, V-IQ= verbal IQ from MAB-II, P-IQ= performance IQ from MAB-II, numbers 1-15 correspond to the 15 personality scales of the Jackson Personality Inventory-Revised (JPI-R), 1= Complexity of Thought, 2= Breadth of Interest, 3=Innovation, 4=Tolerance, 5=Empathy, 6=Anxiety, 7=Cooperativeness, 8=Sociability, 9= Social Confidence, 10=Energy Level, 11=Social Astuteness, 12=Risk-Taking, 13=Organization, 14= Traditional Values, 15=Responsibility

Table 4

Point-Biserial Correlations of Independent Variables with Candidate Success for the full 8-week course, for success from week 1, and for success from weeks 2-8

Time Period	8-week (N=746)	week 1 (N=746)	weeks 2-8 (N=601)
	r	*r*	*r*
PR	.271**	.162**	.261**
Age	*.115***	.097**	.089*
Time	-.047	-.047	-.033
PU	.300**	.199**	.267**
SU	.243**	.186**	.206**
Run	.257**	.244**	.199**
Verbal	.039	-.001	.049
Perf	.100**	.043	.099*
MTS1	-.036	-.026	-.030
MTS2	-.008	-.047	.017
MTS3	-.035	-.44	-.018
MTS4	-.033	-.045	-.016
MTS5	.034	.060	.008
MTS6	-.070	.021	-.099*

Table 4 (Continued)

Point-Biserial Correlations of Independent Variables with Candidate Success for the full 8-week course, for success from week 1, and for success from weeks 2-8

Time Period	8-week (N= 746)	week 1 (N = 746)	weeks 2-8 (N=601)
	r	*r*	*r*
MTS7	.035	.048	.016
MTS8	*.127***	.055	.128**
MTS9	.115**	.082*	.097*
MTS10	.119**	.111**	.088*
MTS11	-.069	-.056	-.055
MTS12	-.033	-.015	-.032
MTS13	.181**	.074*	.185**
MTS14	.029	.001	,036
MTS15	.048	.083*	.014

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

Table 5

LogWorth and significance of predictors in multivariate logistic regression

	Log Worth	*P-Value*
Class Pass Rate	11.69	0.000
Pushup Score	7.48	0.000
Run Score	4.97	0.000
Sit-ups Score	2.21	0.006
Organization	2.12	0.007
Performance IQ	1.76	0.017
Tolerance	1.31	0.049
Cooperativeness	1.27	0.054
Anxiety	0.93	0.119
Energy Level	0.89	0.129
Sociability	0.74	0.182
Innovation	0.52	0.300
Social Astuteness	0.50	0.317
Social Confidence	0.38	0.415
Empathy	0.29	0.513
Risk Taking	0.23	0.595
Time on Ground	0.18	0.662
Breadth of Interest	0.17	0.674
Traditional Values	0.13	0.749

Table 5 (Continued)

LogWorth and Significance of Predictors in Multivariate Logistic Regression

	Log Worth	*P-Value*
Responsibility	0.11	0.771
Complexity	0.10	0.789
Age	0.10	0.794
Verbal IQ	0.03	0.926

Table 6

Factor loading matrix: 11-scale, 5-factor, quartermax rotation

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
MTS2	**0.8422**	0.0746	0.1407	0.0735	0.0586
MTS1	**0.6579**	-0.0575	-0.0988	0.1568	0.1865
MTS4	**0.5714**	0.2396	0.0104	-0.2939	-0.2163
MTS8	0.0842	**0.8676**	0.2182	0.0252	0.0217
MTS5	0.1787	**0.3074**	-0.0128	0.2790	-0.0074
MTS13	0.0172	0.0248	**0.5698**	-0.0827	-0.0604
MTS10	0.3344	0.0886	**0.5199**	-0.1300	0.1141
MTS14	-0.1856	0.0981	**0.4365**	0.0471	-0.2503
MTS7	-0.1147	0.0989	-0.0882	**0.7188**	-0.2151
MTS11	0.1135	-0.0270	-0.0714	**0.3820**	0.2166
MTS12	0.096	0.0331	-0.1349	-0.0096	**0.7877**

Note: Scales for Innovation, Anxiety, Social Confidence, and Responsibility were removed. MTS 1= Complexity, MTS2=Breadth of Interest, MTS4=Tolerance, MTS5=Empathy, MTS7= Cooperativeness, MTS8=Sociability, MTS10=Energy Level, MTS11=Social Astuteness, MTS12=Risk Taking, MTS13=Organization, MTS14=Traditional Values. Bold factor loadings identify the underlying factor structure.

Table 7

Logistic Regression Results of 8-Week Success: 19 Predictors

	Parameter Estimate	Std Error	*p*
Class Pass Rate	**0.077**	**0.012**	**.000**
Age	0.018	0.029	.535
Time	-0.002	0.004	.677
PU score	**0.059**	**0.011**	**.000**
SU score	**0.033**	**0.012**	**0.005**
Run score	**0.064**	**0.014**	**.000**
Verbal IQ	0.005	0.011	.649
Performance IQ	**0.022**	**0.010**	**.027**
Openness	-0.028	0.015	.0608
Conscientiousness	**0.053**	**0.020**	**.006**
Innovation	-0.009	0.012	.472
Empathy	0.006	0.012	.632
Anxiety	-0.020	0.015	.195
Cooperativeness	0.026	0.014	.060

Table 7 (continued)

Logistic Regression Results of 8-Week Success: 19 Predictors

	Parameter Estimate	Std Error	*p*
Sociability	0.012	0.011	0.263
Social Confidence	0.012	0.014	.417
Social Astuteness	0.009	0.011	.405
Risk Taking	0.001	0.010	.918
Responsibility	0.007	0.015	.655

Note: Model Diagnostics: AICc- 863.0; BIC 954.1; ROC AUC- .788; Lack of Fit Probability > χ^2= .007

Table 8

Logistic regression results of 8-week success; final model - eight predictors

	Parameter Estimate	*Std Error*	*p*	*OR*	*Range OR*
Class Pass Rate	0.078	0.011	.000	1.081	7.59
Run Score	0.066	0.014	.000	1.068	7.15
Pushup Score	0.058	0.010	.000	1.060	8.60
Conscientiousness	0.058	0.016	.000	1.060	8.33
Sit-up Score	0.034	0.012	.003	1.035	3.46
Openness	-0.031	0.012	.011	0.969	0.31/ 3.22
Performance IQ	0.024	.008	.003	1.024	4.51
Sociability	0.019	.009	.032	1.019	2.34

Note: Model Diagnostics: AICc- 847.6; BIC- 888.9; ROC AUC- .784; Lack of Fit Probability > χ^2= .009

Table 9

Confusion Matrix: 8-week final model, optimized for equivalent type I and type II prediction errors

		Predicted	
		Pass	Fail
Actual	Pass	233	105
	Fail	107	301

Note: Model probability cutoff of .48 yielded balanced type I and type II error rates. Actual number of passing candidates was 338; predicted number of passing candidates based on model was 340. True positive rate (sensitivity, recall, e.g.) = 0.69. True negative rate (specificity, selectivity, e.g.) = 0.74. F1 score =.69

Table 10

Logistic regression results of week 1 success; full model: 19 predictors

	Parameter Estimate	*Std Error*	*p*
Class Pass Rate	**0.034**	**0.012**	**.006**
Age	0.037	0.035	.280
Time	-0.005	0.004	.247
Pushup Score	**0.033**	**0.012**	**.008**
Sit-up Score	**0.026**	**0.013**	**.045**
Run Score	**0.062**	**0.014**	**.000**
Cooperativeness	**0.033**	**0.017**	**.060**
Verbal IQ	-0.017	-0.012	.773
Performance IQ	0.017	0.012	.156
Innovation	-0.017	0.014	.243
Empathy	0.011	0.014	.455
Anxiety	0.018	0.018	.323
Sociability	-0.008	0.012	.507
Social Confidence	**0.030**	**0.016**	**.064**

Table 10 (continued)

Logistic regression results of week 1 success; full model: 19 predictors

	Parameter Estimate	*Std Error*	*p*
Social Astuteness	-0.013	0.013	.313
Risk Taking	0.012	0.012	.338
Responsibility	**0.038**	**0.017**	**.029**
Openness	**-0.034**	**0.017**	**.048**
Conscientiousness	0.015	0.022	.494

Note: Model Diagnostics: AICc- 674.6; BIC 765.7; ROC AUC- .747; Lack of Fit Probability > χ^2= .994

Table 11

Logistic regression results of week 1 success; final model - eight predictors

	Parameter Estimate	*Std Error*	*p*	*OR*	*Range OR*
Run Score	0.062	0.014	.000	1.064	6.43
Class Pass Rate	0.040	0.012	.000	1.041	2.85
Responsibility	0.040	0.014	.006	1.040	5.51
Openness	-0.038	0.015	.013	0.963	0.241/4.15*
Pushup Score	0.034	0.012	.005	1.035	3.51
Cooperativeness	0.030	0.015	.040	1.030	3.61
Sit-up Score	0.027	0.013	.040	1.027	2.63
Social Confidence	0.026	0.013	.044	1.026	3.05

Note: Model Diagnostics: AICc- 661.0; BIC- 702.2; ROC AUC- .730; Lack of Fit Probability > χ^2= .995; * inverse range odds ratio depicted for comparison purposes.

Table 12

Confusion Matrix: week 1 final model, optimized for equivalent type I and type II prediction errors

		Predicted	
		Pass	Fail
Actual	Pass	518	83
	Fail	80	65

Note: Model probability cutoff of .71 yielded balanced type I and type II error rates. Actual number of passing candidates was 601; predicted number of passing candidates based on model was 598. True positive rate (sensitivity, recall, e.g.) = 0.86. True negative rate (specificity, selectivity, e.g.) = 0.45. F_1 score=.86

Table 13

Logistic regression results of week 2-8 Success; 19 predictors

	Parameter Estimate	*Std Error*	*p*
Class Pass Rate	**0.082**	**0.013**	**.000**
Age	0.005	0.032	.872
Time	0.001	0.005	.880
Pushup Score	**0.055**	**0.012**	**.000**
Sit-up Score	**0.027**	**0.013**	**.034**
Run Score	**0.053**	**0.015**	**.000**
Verbal IQ	0.006	0.011	.622
Performance IQ	0.017	0.011	.113
Innovation	-0.005	0.013	.700
Empathy	0.001	0.013	.938
Anxiety	-0.032	0.017	.056
Cooperativeness	0.025	0.015	.094
Sociability	0.014	0.012	.228
Social Confidence	0.007	0.015	.658

Table 13 (continued)

Logistic regression results of week 2-8 Success; 19 predictors

	Parameter Estimate	*Std Error*	*p*
Social Astuteness	-0.003	0.012	.813
Risk Taking	-0.005	0.011	.679
Responsibility	-0.015	0.017	.367
Openness	-0.025	0.016	.121
Conscientiousness	**0.054**	**0.021**	**.012**

Note: Model Diagnostics: AICc- 723.7; BIC 810.2; ROC AUC- .769; Lack of Fit Probability > χ^2= .002

Table 14

Logistic regression results of week 2-8 success; final model – eight predictors

	Parameter Estimate	*Std Error*	*p*	*OR*	*Range OR*
Class Pass Rate	0.081	0.012	.000	1.085	8.27
Run Score	0.055	0.015	.000	1.056	5.20
Pushup Score	0.055	0.011	.000	1.056	7.54
Conscientiousness	0.053	0.018	.004	1.055	7.01
Anxiety	-0.030	0.014	.038	0.971	0.31/*3.19
Cooperativeness	0.028	0.013	.033	1.029	3.37
Sit-up Score	0.027	0.013	.033	1.028	2.67
Performance IQ	0.016	0.009	*.059	1.017	2.83

Note: Model Diagnostics: AICc- 708.8; BIC- 748.1; ROC AUC- .763; Lack of Fit Probability > χ^2= .003. * Inverted Range Odds Ratio (negative effect) shown for comparison.

Table 15

Confusion Matrix: week 2-8 final model, optimized for equivalent type I and type II prediction errors

		Predicted	
		Pass	Fail
Actual	Pass	254	86
	Fail	90	173

Note: Model probability cutoff of .53 yielded balanced type I and type II error rates. Actual number of passing candidates was 340; predicted number of passing candidates based on model was 344. True positive rate (sensitivity, recall, e.g.) = 0.75. True negative rate (specificity, selectivity, e.g.) = 0.66. F_1 score =.85

Table 16

Mean Personality Scale T-Scores, three sample norms

	All Candidates (N = 748)	*Passing Candidates* (N = 340)	*Leader Selection Candidates* *(N = 1000)*
Complexity	47.0	47.4	47.8
Breadth of Interest	49.2	49.2	52.5
Innovation	45.5	45.8	48.9
Tolerance	50.9	51.2	51.6
Empathy	44.0	43.7	43.1
Anxiety	**41.6**	**42.2**	**42.5**
Cooperativeness	42.8	42.5	45.2
Sociability	52.2	50.8	49.4
Social Confidence	53.0	51.9	55.4
Energy Level	54.0	52.9	56.6
Social Astuteness	45.6	46.3	48.2
Risk Taking	48.9	49.2	47.8
Organization	**59.0**	**57.5**	**59.3**
Traditional Values	55.9	55.6	56.9
Responsibility	55.0	54.6	57.6

Note: The highest and the lowest scale T-scores for each normative comparison are in **Bold**.

Table 17

Bivariate correlations of threshold and sensitivity across three time periods, and three comparison norms

	8-week	week 1	weeks 2-8
Leader Norms			
Threshold	-.001	.024	-.015
Sensitivity	.086*	.028	.093*
Total Candidate Norms			
Threshold	.029	.050	.008
Sensitivity	.100**	.014	.117**
Passing Candidate Norms			
Threshold	-.001	.030	-.017
Sensitivity	.112**	.039	.120**

Note: * significant at $p < .05$; ** significant at $p < .01$;

Table 18

Multivariate Logistic Regression Results-ROC AUC for three time periods, three sets of trait norms

	8-week	week 1	weeks 2-8
Leader Norms			
Full Model	.770	.716	.751
Final Model	.769	.712	.750
Total Candidate Norms			
Full Model	.772	.720	.753
Final Model	.769	.712	.751
Passing Candidate Norms			
Full Model	.773	.720	.754
Final Model	.771	.712	.752

Note: All models included Class Pass Rate control variable, pushup score, sit-up score, run score. 8-week and weeks 2-8 final models included performance IQ and personality sensitivity. Week 1 final models only included class pass rate control and the three physical predictors.

Table 19

Confusion Matrix: 8-week final personality profile model, optimized for equivalent type I and type II prediction errors

		Predicted	
		Pass	Fail
Actual	Pass	230	108
	Fail	114	293

Note: Model probability cutoff of .48 yielded balanced type I and type II error rates. Actual number of passing candidates was 338; predicted number of passing candidates based on model was 344. True positive rate (sensitivity, recall, e.g.) = 0.68. True negative rate (specificity, selectivity, e.g.) = 0.72. F1 score =.67

Table 20

Side-by-side comparison of models and predictor effects, three considered timeframes

	8-week		*week 1*		*weeks 2-8*	
	Full Model	*Final Model*	*Full Model*	*Final Model*	*Full Model*	*Final Model*
Control Variables						
Class Pass Rate	*S*	*S*	*W*	*W*	*S*	*S*
Age						
Time on Ground						
Physical Predictors						
Pushup score	*S*	*S*	*M*	*M*	*S*	*S*
Sit-up score	*M*	*M*	*W*	*W*	*W*	*W*
Run score	*S*	*S*	*S*	*S*	*S*	*S*
Cognitive Predictors						
Verbal IQ						
Performance IQ	*M*	*S*	*W*			*W**
Openness related Predictors						
Openness composite	*W*	*M*	*M*	*S*		
Innovation						
Conscientiousness related Predictors						
Conscientiousness composite	*S*	*S*			*S*	*S*
Responsibility			*S*	*S*		
Emotional Stability related predictors						
Empathy						
Anxiety						*M*

Table 20 (continued)

Side-by-side comparison of models and predictor effects, three considered timeframes

	8-week		*week 1*		*weeks 2-8*	
	Full Model	*Final Model*	*Full Model*	*Final Model*	*Full Model*	*Final Model*
Risk-Taking						
Extraversion related predictors						
Sociability		*W*				
Social Confidence				*M*		
Energy Level						
Agreeableness related predictors						
Cooperativeness				*M*		*M*
Social Astuteness						

Note: W =Weak (Recommended Minimum Practically Significant Effect Size), M = Moderate effect size, S= Strong effect, using Ferguson's (2009) conventions of 2, 3, and 4 as Odds Ratio guidelines, and predictor Range Odds Ratios. * Performance IQ was kept in the model at p = .059; all other predictors were significant at p < .05.

Figure 1

Overview of 8-week SOF selection through training program, with Fiscal Year 2019 overall entrance, completion, and attrition numbers

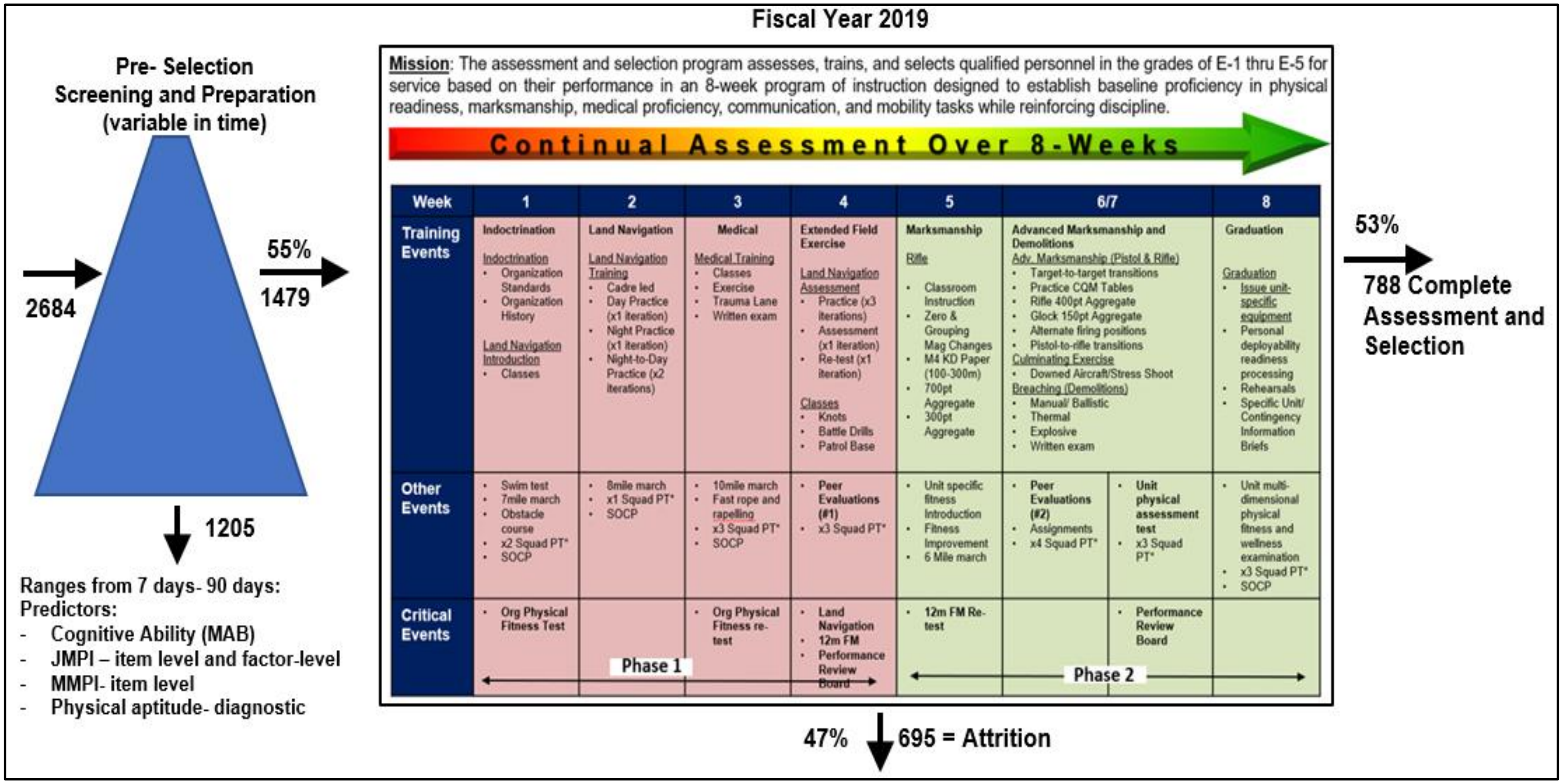
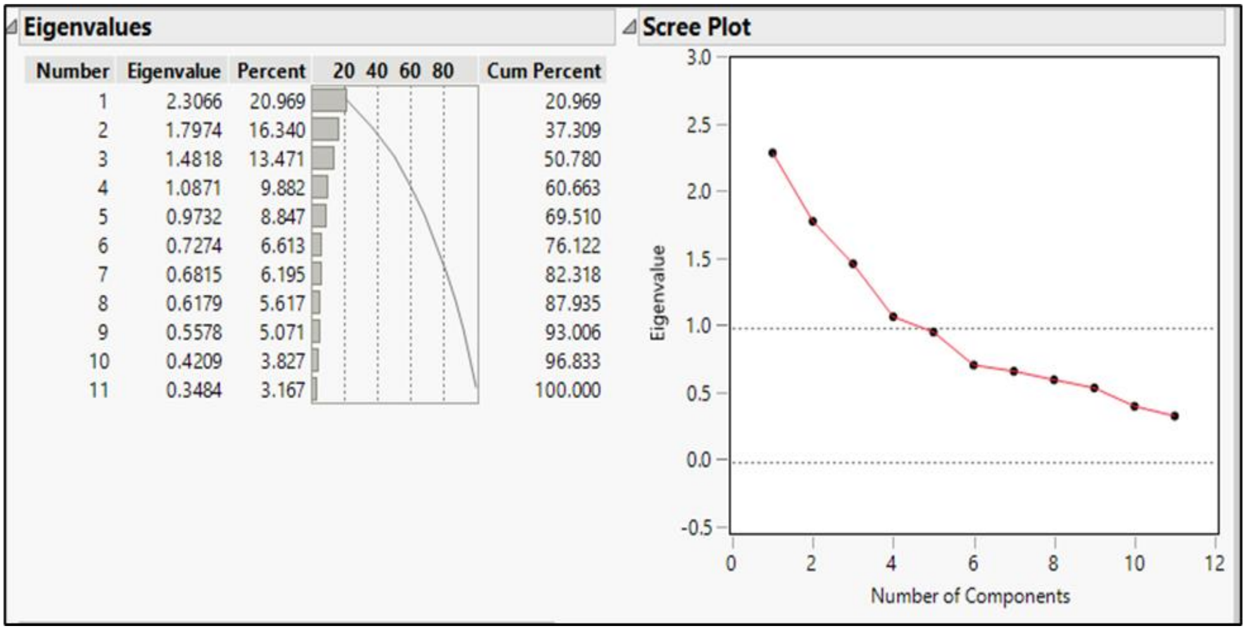


Figure 2

Eigenvalues and scree plot of 11 personality scales (MTS 3-Innovation, 6- Anxiety, 9-Social Confidence, 15-Responsibility, excluded)

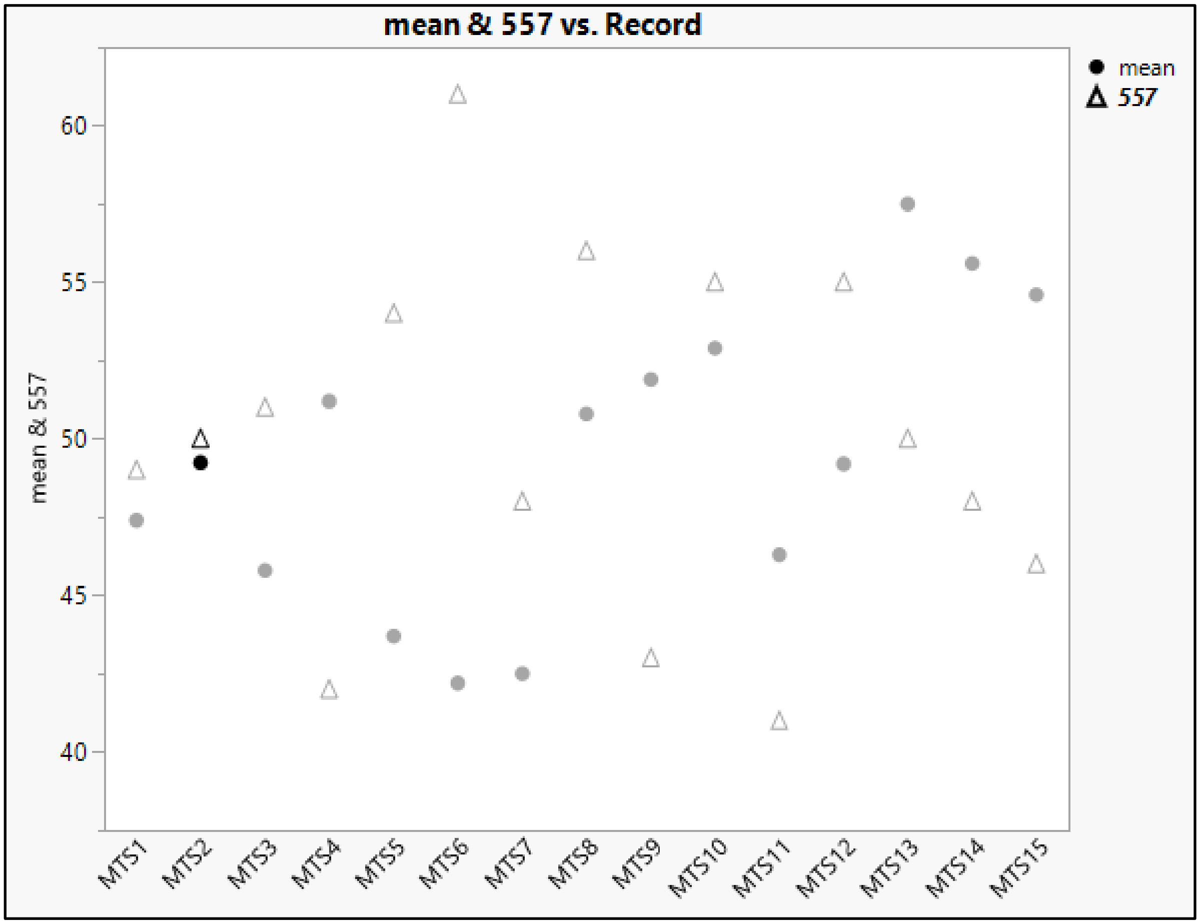


Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent
1	2.3066	20.969		20.969
2	1.7974	16.340		37.309
3	1.4818	13.471		50.780
4	1.0871	9.882		60.663
5	0.9732	8.847		69.510
6	0.7274	6.613		76.122
7	0.6815	6.195		82.318
8	0.6179	5.617		87.935
9	0.5578	5.071		93.006
10	0.4209	3.827		96.833
11	0.3484	3.167		100.000

Figure 3

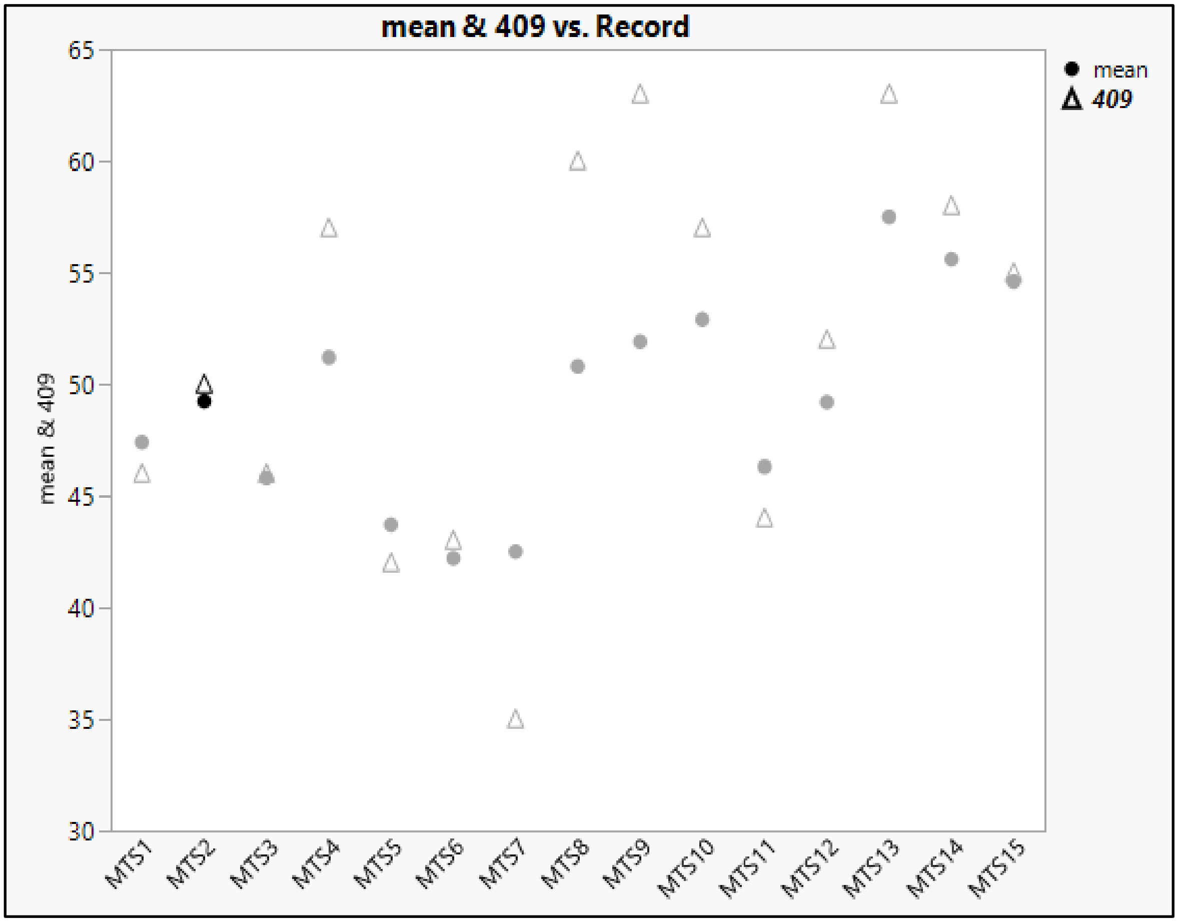
Personality Profile Graph: Low Threshold, Low Sensitivity, Record 557



Note: Threshold: 83; Sensitivity $\underline{r}$ = -.23

Figure 4

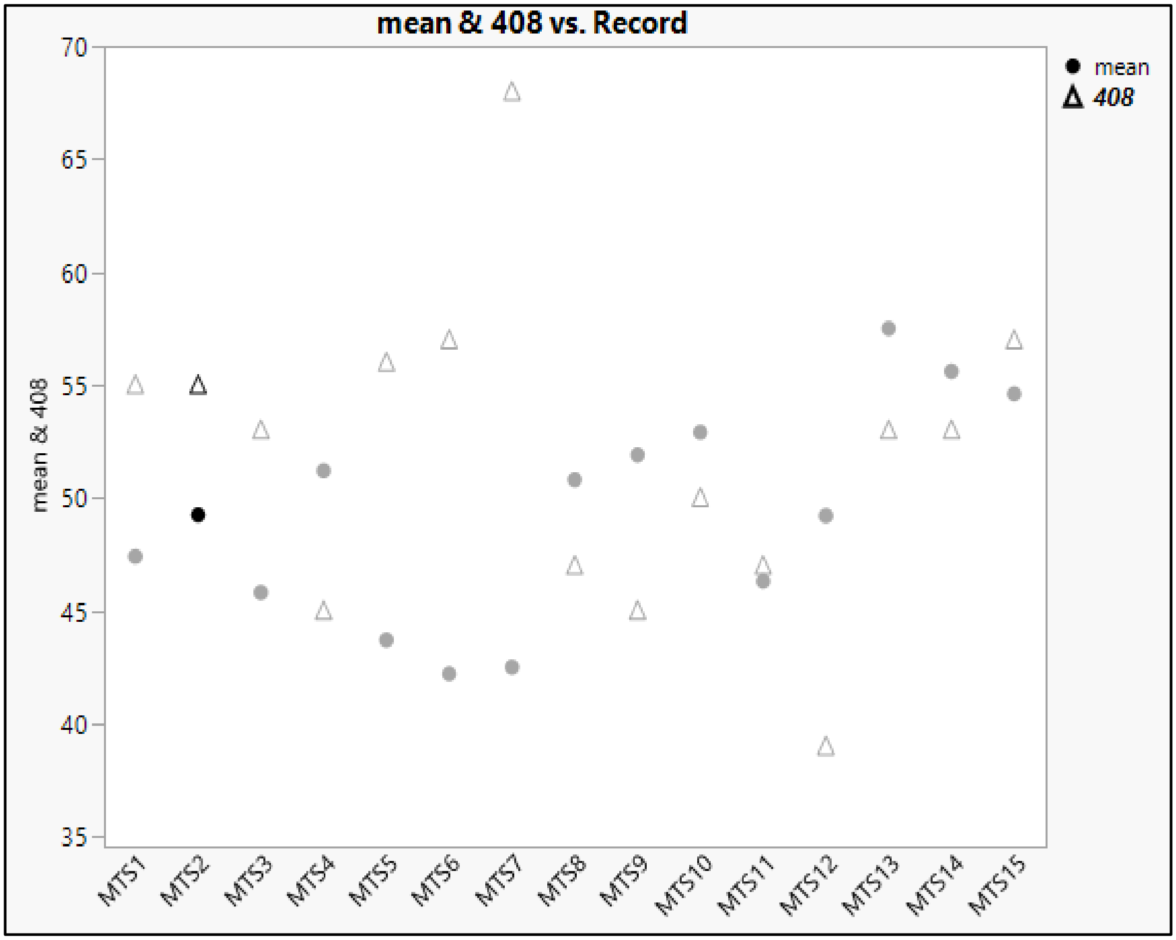
Personality Profile Graph: Low Threshold, High Sensitivity, Record 409



Note: Threshold: 56; Sensitivity r = .91

Figure 5

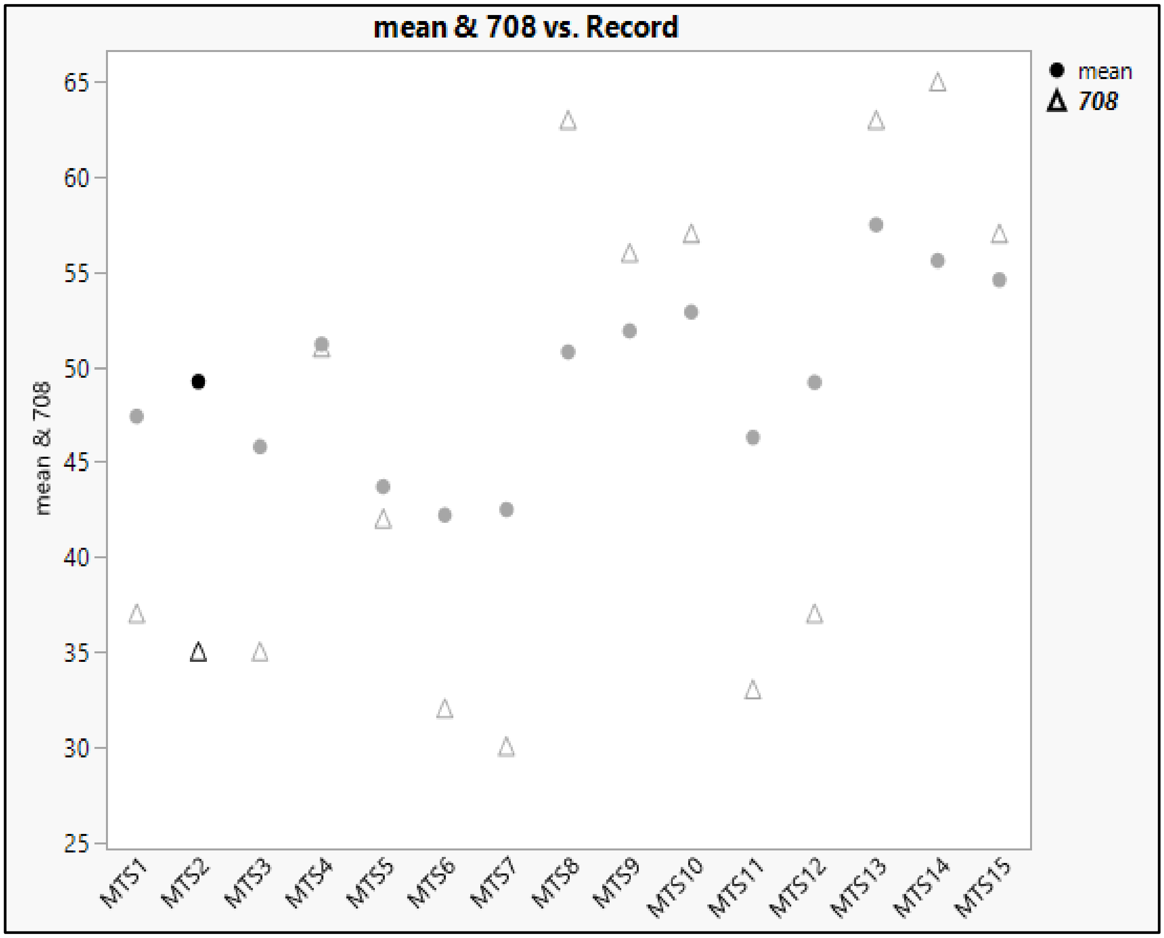
Personality Profile Graph: High Threshold, Low Sensitivity, Record 408



Note: Threshold: 132; Sensitivity r = -.44

Figure 6

Personality Profile Graph: High Threshold, High Sensitivity, Record 708



Note: Threshold: 123; Sensitivity $\underline{r}$ = .84