USE OF A PRE-EMPLOYMENT MULTITASKING INSTRUMENT AND NEUROPSYCHOLOGICAL MEASURES TO PREDICT FIELD PERFORMANCE RATINGS IN POLICE OFFICERS

APPROVED BY SUPERVISORY COMMITTEE

C. Munro Cullum, Ph.D. ABPP

Antoinette McGarrahan, Ph.D.

Linda Hynan, Ph.D.

Timothy Proctor, Ph.D. ABPP

Stephen Bishopp, Ph.D.

DEDICATION

I owe my deepest gratitude to many individuals, for without their support, this project would not have been possible. I would like to thank the members of my Graduate Committee for their time and patience in taking this project from inception to completion. To Dr. Hynan, for her tireless assistance and encouragement while I muddled through the statistical analyses, to Dr. McGarrahan for the original idea, ceaseless encouragement, and building the relationships with the Grand Prairie and Carrollton Police Departments, to Dr. Bishopp and Dr. Proctor for your feedback and thoughtful questions, and finally, to my mentor, Dr. Cullum, for the countless meetings, support, and patience as I focused my interests and found my path. Your direction and guidance has been paramount in my development as a scientist and clinician. I extend my deepest gratitude to Chief Dye, the Grand Prairie Police Department, Chief Redden and the Carrollton Police Department for their assistance with this project. I would also like to thank Elida Godbey for her assistance and guidance building the database in REDcap and to Doris Escobedo for her administrative support and scheduling and rescheduling meeting after meeting. I would also like to thank the staff of the UTSW Neuropsychology Clinic with special thanks to Judy Shaw for patiently correcting my scoring errors and imparting her wisdom regarding the assessment process. She was always available and never grew tired of my myriad questions. And thank you for always remembering birthdays and keeping up with each individual's dessert preference. A special thank you is owed to my mentors and friends, Dr. Andrew Schmitt and Dr. Paula Lundberg-Love. Dr. Schmitt offered me an opportunity to assist with his research the first day of my undergraduate career and shortly thereafter, Dr. Lundberg-Love took a chance and allowed me to take the lead on a book chapter. I was intimidated and overwhelmed, but with

their guidance, I successfully published my first manuscript. Thank you both for your patience and commitment to my learning. And to Dr. Lynette Abrams-Silva, your friendship and mentoring has been invaluable. Thank for your sharing your excitement and vision of neuropsychology, the many challenging fact-finding sessions, and for being my friend and confidante.

Finally, I owe the biggest debt of gratitude to my loving family, who have supported me in every way possible without fail and never declined when I needed to practice my testing skills. To my father and mother, David and Rebecca Galusha, thank you for instilling in me an insatiable drive and curiosity to never stop learning and growing. To my son, Tristan, and my daughter, Haley, thank you for walking this path with me. We make an incredible team. I know it has been hard, but we did it, we made it to the finish line! And, to my brother, the late Dr. Jeremy W. Galusha, thank you for never giving up on me and believing in me even when I didn't believe in myself.

USE OF A PRE-EMPLOYMENT MULTITASKING INSTRUMENT AND NEUROPSYCHOLOGICAL MEASURES TO PREDICT FIELD PERFORMANCE IN POLICE OFFICERS

by

JEANINE MARIE GALUSHA

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

August, 2017

USE OF A PRE-EMPLOYMENT MULTITASKING INSTRUMENT AND NEUROPSYCHOLOGICAL MEASURES TO PREDICT FIELD PERFORMANCE RATINGS IN POLICE OFFICERS

Publication No. ___________________

Jeanine Marie Galusha, Ph.D.

The University of Texas Southwestern Medical Center, 2017

Supervising Professor: C. Munro Cullum, Ph.D., ABPP

The role of the police officer has evolved drastically since the inception of pre-employment psychological evaluations in the 1960s and increasingly relies on adequate multitasking ability to keep up with the demands of the job. However, though necessary, multitasking is not directly assessed. Conversely, some agencies make use of a multitasking test (*CritiCall*) to evaluate prospective 911 operators during the application process. Thus, the goals of this study were to evaluate the utility of standard neuropsychological tests and *CritiCall* in predicting police officer success and to evaluate the relationship between the multitasking instrument and neuropsychological measures. North Texas police officers were administered the NIH Toolbox Cognition and Emotion Batteries along with several traditional neuropsychological measures of attention, working memory, and mental flexibility, and a subset completed the *CritiCall* test. Each officers' direct supervisor completed a standard Supervisor Survey which served as the primary outcome measure. Spearman correlations were used to compare performance on *CritiCall* to the NIH Toolbox Cognition Battery and standard neuropsychological test scores, and all scores were tested for predictive ability in relation to the Supervisor Survey via stepwise linear regression. Aspects of the multitasking instrument were associated with performance on measures of processing speed, attention, working memory, mental flexibility, and crystallized cognitive abilities. No subtests or composite scores from the NIH Toolbox Cognition Battery were predictive of the Supervisor Survey, whereas several subtests from the NIH Toolbox Emotion Battery were significant predictors. Standard neuropsychological tests combined with subtests from the Emotion Battery were found to be some of the strongest predictors, specifically the combination of a measure of perceived stress and working memory. Primary results provide evidence for a relationship between neuropsychological factors and police field performance ratings. Continued research is needed to further evaluate and validate the Supervisor Survey and to confirm these findings in additional settings.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER ONE
Introduction

Today more than ever, police officers are tasked with functioning in an increasingly complicated environment. Officers find themselves in patrol cars bombarded with continuous stimuli. In addition to lights and a siren, a radio to monitor and answer, and frequently, a mobile data terminal (MDT) with which to interact, officers often perform these and other tasks simultaneously while driving (see e.g. Yager, Dinakar, Sanagaram, & Ferris, 2015). Officers are also tasked with synthesizing large amounts of information and making rapid decisions under stressful circumstances. Occasionally, the decisions they make mean the difference between life and death for officers and citizens. Lower abilities in certain cognitive skills may increase the incidence of errors in judgement, especially in times of great stress or when the officer is distracted. The need for higher level cognitive functioning and the ability to multitask is vitally important in policing. Surprisingly, there is a dearth of research examining the cognitive skills associated with successful police performance.

Further, as with many other professions, there are currently no agreed upon standards or requirements in place within police agencies to assess desired cognitive skills (Cochrane, Tett, & Vandacreek, 2003). At this time, in addition to a general medical exam and passing a drug test, the only other requirement for licensure as a Texas Peace Officer is for an applicant to be in satisfactory psychological and emotional health as measured by two instruments, one to assess personality and one to evaluate psychopathology. No further guidance is afforded by Texas state law; thus, a psychologist has great latitude regarding the instruments utilized for this purpose.

Intuitively, the best predictor of job success is the ability to perform the necessary functions of the job. Beyond the selection process, satisfactory completion of the police academy and passing the state licensing exam are the current methods for assessing a new police officers' ability. However, the dynamic nature of policing and the complexities of human behavior mean that it is not possible to assess all of the officers' abilities during the academy (see also Fyfe, 1999). For instance, the pressure of making decisions under duress, constant bombardment of stimuli inside the patrol car as well as out, and most assuredly, risking one's life are job functions that are unmeasurable in a sterile academy setting. These very abilities are inextricably connected to successful policing but least observable during recruit training (also see White, 2008). Additionally, research findings have been mixed in terms of the relationship between academy performance and later outcomes, i.e. success in the field (Henson, Reyns, Klahm, & Frank, 2010). Thus, the selection process where physical and personality evaluations are currently performed is the first opportunity to help identify those best suited to police work, yet cognitive skills are not measured. This may be related to several factors including a lack of information regarding which cognitive skills are most important in law enforcement.

Intelligence testing has a long history of utilization for job placement. During World War I, Robert Yerkes and a committee designed the Alpha and Beta Army intelligence tests to determine if Army recruits were suited for military service (Boake, 2002). Both tests were created for group administration and the Alpha test was used with literate English speakers and the Beta for illiterate or non-proficient English speakers. The Alpha test included analogies, synonyms and antonyms, and number-completion whereas the Beta test used incomplete-pictures and coding with all tests being timed (Army Alpha and Beta tests). In a meta-analysis, Hunter and Hunter (1984) analyzed different predictors of job performance and found cognitive ability had a mean validity of .53 in predicting job performance as measured by supervisor ratings for entry-level jobs and Hunter (1986) reported validity of .58 for predicting performance in high complexity jobs, .51 for medium complexity, and .40 for low complexity jobs.

Beyond the use of IQ testing, the assessment of more detailed cognitive skills has been useful in predicting real world functioning. In one of the earliest investigations of the relationship between neuropsychological test performance and employment, Heaton, Chelune, and Lehman (1978) found that neuropsychological tests commonly used in neurological populations successfully identified patients at higher risk for unemployment. Their sample included normal controls, people with pesticide poisoning, referrals from neurology/ neurosurgery, individuals in vocational rehabilitation, and people from other social agencies and physicians. The tests that appeared to discriminate the best between employed and unemployed groups were the ones that assessed current adaptive abilities (i.e. performance on the Halstead-Reitan Battery and Wechsler Adult Intelligence Scale performance subtests) as compared to tests related to past experience and education (i.e., WAIS verbal subtests and Peabody Individual Achievement Test) (Heaton et al., 1978). More recently, global cognitive functioning predicted successful training outcomes with pilots (see King et al., 2013) and was significantly associated with aviator performance over and above personality testing (Paullin, Katz, Bruskiewicz, Houston, & Damos, 2006). Psychometric *g* (general intelligence) as measured by the Air Force Officer Qualifying Test Form O, was also found to be the best predictor of pilot and navigator success, with the addition of specific job knowledge adding little predictive ability beyond *g* (Olea & Ree, 1994). Similarly, Ree and Earles (1992) reported measured intelligence was the best overall predictor of job training success, irrespective of job difficulty based on studies involving pilots, navigators, and other airmen. In addition, in a meta-analytic study covering 85 years of research in the area of personnel selection, Schmidt and Hunter (1998) noted that general mental ability (i.e. intelligence as measured by undefined commercially available tests) was a strong predictor of job performance in both medium complexity and professional-managerial jobs.

While several studies and meta-analyses report a significant association between cognitive ability and job performance, little is known about any such relationship among predictors of police performance. An investigation by Higgins, Peterson, Pihl, and Lee (2007) sought to examine if the relationship between "prefrontal cognitive ability" and intelligence. Additionally, they tested cognitive ability as a predictor of job performance among high-complexity administrative and low-complexity factory floor jobs. They defined their measures as assessing "dorsolateral prefrontal cognitive ability" based on performance on a computerized battery of seven cognitive tasks designed to assess conditional associative learning, working memory, and word fluency. Job performance was measured by supervisor and self-rated evaluations on a 16-point rating scale that included the following dimensions: quality, quantity, knowledge, versatility, judgment, communications, human relations, professionalism, responsiveness, punctuality, and attendance. Overall performance on their cognitive battery was significantly related to job performance (corrected r's = .52 to .72) as rated by supervisors (Higgins et al., 2007).

While neuropsychological testing has been shown to be useful in predicting job performance in several populations as noted above, it has yet to be utilized in the area of police selection or evaluation, though this has been recommended. As early as 1917, Terman reasoned general intelligence was the most important characteristic (next to integrity) in ascertaining fitness for duty as a police officer or firefighter and investigated the use of an abbreviated form of the Stanford-Binet intelligence scale as one aspect of the hiring process for 30 applicants to the San Jose, California fire and police departments. He also recommended further investigation correlating results of mental tests with later job success. Thurstone (1922) evaluated the use of the Army Alpha test in a sample of Detroit police officers and recommended steps be taken to attract individuals of "superior mentality" to the police service. Simply stated, he recognized the importance of intelligence to success as a police officer. More recently, Meier, Farmer, and Maxwell (1987) recommended intelligence and cognitive function testing as part of the psychological screening of potential officers due to the face validity and importance of these types of measures in other jobs. Interestingly, some law enforcement agencies are utilizing a pre-employment multitasking test, *CritiCall*, to assess 911 operators prior to employment to ensure applicants can effectively multitask and fulfill the necessary role. Despite the aforementioned need for police officers to effectively multitask, this measure has not been studied in police populations. Furthermore, *CritiCall* has not been examined with respect to the cognitive skills required for success or how scores might help predict in-field performance. The purpose of this study is to answer these questions.

CHAPTER TWO
Review of the Literature

History of Law Enforcement Pre-Employment Testing

Psychological evaluations of prospective police officers dates back to 1967 following a statement by the President's Commission on Law Enforcement and Administration of Justice in which it was recommended emotional stability of officers be determined through psychological tests prior to employment (Simmers, Bowers, & Ruiz, 2003)). This was followed by a statement by the National Advisory Commission on Criminal Justice Standards and Goals recommending police agencies retain a psychiatrist or psychologist to conduct said testing by 1975 and a statement by the president of the International Association of Chiefs of Police (IACP) in 1986 recognizing psychological services as an integral part of a well-run agency (Simmers et al., 2003). The ultimate goal of psychological testing is to evaluate an officer's fitness for duty and to protect the agency and the public. Police officers wield immense power. They are granted the authority to enforce laws, detain, search, and arrest individuals. Moreover, they are legally mandated to use the amount of force required to enact a lawful objective, up to and including deadly force. Bittner (1970) noted that the role of police is defined by their ability to use force. Indeed, case law supports police using authority to carry out their mandate of protecting the public and providing security to the community (see also Simmers at al., 2003). Decisions made by police officers, some occurring on their first day on the job, can greatly impact people and entire communities. Hence, it is appropriate that the ultimate goal of psychological testing is to determine, *a priori*, who is the best fit for a career in law enforcement.

Since the inception of psychological testing for law enforcement officers, a number of instruments for assessing mental and emotional responses have emerged; however, the research has been mixed regarding the association of different instruments with outcome variables including officer performance. Thus far, there is no consensus regarding which test(s) to utilize or how best to measure outcomes (Dantzker, 2011). In a survey of 17 municipal police agencies and the Texas Department of Public Safety, Danztker and McCoy (2006) found all agencies used at least two tests (up to six), with the most popular (78%) being the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), a common measure of adult psychopathology. The second most commonly used (33% of respondents) was the 16 Personality Factor Questionnaire (16-PF) (Dantzker & McCoy, 2006). Similar findings were reported in a national survey of police agencies with the MMPI-2 being the most frequently used measure, utilized in 71.6% of agencies (Cochrane at al., 2003). The MMPI has a long history of utilization in pre-employment psychological evaluations of police officers (Azen, Snibbe, & Montgomery, 1973; Saccuzzo, Higgins, & Lewandowski, 1974; Saxe & Reiser, 1976; Schoenfeld, Kobos, & Phinney, 1980; Bartol, 1982; Mills & Stratton, 1982; Beutler, Storm, Kirkish, Scogin, & Gaines, 1985; Mullins & McMains, 1995; Weiss, Serafino, Serafino, Willson, & Knoll, 1998; Tarescavage, Brewster, Corey & Ben-Porath, 2015; Tarescavage, Corey & Ben-Porath, 2015). Another self-report instrument, the Inwald Personality Inventory (IPI; Inwald, Knatz, & Shusman, 1980) was developed specifically for law enforcement populations. Simmers and colleagues (2003) evaluated the use of the MMPI, the MMPI-2, and the IPI and their relationship to future police functioning in 18 studies spanning from 1980 to 2000, and reported that all three provided at least “modest” correlations (MMPI/ MMPI-2 range = .05 - .34; IPI range = .22 - .44) and effect size relationships (MMPI/ MMPI-2 range = .10 - .88; IPI range = .49 - 1.03) depending on performance variables examined. In contrast, in an investigation with the Los Angeles Sheriff’s Department with an unknown number of subjects, Mills and Stratton (1982) found no relationship between performance on the MMPI and success as a police officer defined at three levels: academy acceptance, graduation, and field employment.

Tarescavage, Brewster, Corey, and Ben-Porath (2015) examined the ability of the MMPI-2-RF to predict police supervisor performance ratings (N=131) and found scales from the emotional dysfunction and interpersonal domains were most related to post-hire supervisor ratings, with small but statistically significant zero-order correlations r’s < .30. An additional exploratory study by Tarescavage, Corey, and Ben-Porath (2015) examined the predictive validity of the MMPI-2-RF for problem behavior among police officers and again found (among the hundreds of correlations reported) a statistically significant relationship with outcome variables, specifically between the emotional dysfunction and interpersonal domain (zero-order correlations r’s < .30). Of particular interest, the cognitive complaints scale, a self-report of subjective memory problems and concentration difficulties, showed correlations ranging from r=.18 to r=.25 and was associated with problems related to failure to control conflict, assertiveness, and routine tasks such as radio usage, decision-making, multitasking (Tarescavage, Corey, & Ben-Porath, 2015). The reported relationship between the cognitive complaints scale and associated problem areas may indicate an officer’s perceived difficulty managing and processing information in not only conflict situations but also during routine tasks. Likewise, Black (2000) found the Pl/Pq Higher Test, a commercially available measure of cognitive ability, had the highest correlation with overall training performance in a sample of police recruits. Another personality assessment instrument, the NEO Personality Inventory-Revised (NEO-PI-R; Costa & McRae, 1992) has been utilized in police officers as well. Detrick, Chibnall, and Luebbert (2004) found select facet scores (values, excitement-seeking, anxiety, deliberation, fantasy, activity, ideas, values, self-consciousness, altruism, feelings, order, positive emotions, and vulnerability) were related to some aspects of police academy performance (such as graduation, firearms, and physical training) but these authors did not examine scores in relation to later job performance. Furthermore, Sanders (2008) found the Big Five personality traits (extroversion, neuroticism, agreeableness, conscientiousness, and openness) as measured by the online test, Big Five Inventory, were not predictive of police performance.

Other methods besides self-report personality scales have been examined for use in predicting police officer success and include police academy performance, civil service test scores, and completion of a college degree or a certain number of college hours. Unfortunately, as with personality testing, research examining the validity of each of these methods has also been mixed. White (2008) found hours of college education per se was not a significant predictor of police academy performance. Similarly, in a study of officers from a Midwestern police department, Henson and colleagues (2010) found college education was not related to measures of academy or on the job success whereas civil service exam score was related to academy success. Additionally, civil service exam scores predicted higher supervisor evaluations after the 2nd year and across the three-year average. Finally, the overall academy score (based on average grade performance) was related to officers' 1st year evaluation and their 3-year evaluation average (Henson et al., 2010). Despite these somewhat contradictory findings regarding what tests or combination of tests predict police officer performance, no one has of yet focused on the component cognitive processes that underlie success, although several areas of function appear to be especially promising.

Multitasking

In addition to personality factors that may relate to police suitability/performance, the ability to multitask is inherent to policing, as noted, and the need has continued to grow with the addition of multi-media devices to the patrol car. A typical patrol unit has a minimum of a police radio and switch panels to control the lights and siren. Frequently, a mobile data terminal (MDT) is added, plus the officer's cell phone is used, and most patrol units are equipped with a dash camera and radar detector. When operating a vehicle one must maintain awareness of the driving environment as well as manage the multiple streams of information inside the vehicle. Though many people believe they are capable of multitasking successfully, research indicates this is a false belief and in reality, only about 2% of the population, sometimes deemed "supertaskers," are able to multitask without experiencing a subsequent decrease in performance (Gustafson, 2015; Watson & Strayer, 2010). Thus, the majority of the population is actually engaged in task switching, which in simple cognitive tasks is associated with an increase in reaction time and error rate (Rogers & Monsell, 1995). While the addition of various electronics to the patrol unit are designed to aid officers in performing the requirements of their job, they come at a potential price. There is the ever present possibility of overwhelming an officer with so many stimuli that they miss vital information in their environment. Oftentimes, officers must manage this steady influx of information and respond accordingly, via various modalities in stressful and potentially dangerous circumstances. There is little room for error when one's life or the lives of others may be at stake.

Garrison, Williams and Carruth (2012) studied law enforcement officers in a simulated driving task which required participants to interact with a semi-automated dispatch system and an MDT. Task demands were varied between conditions and eye movements toward the MDT were analyzed. They found on average, 25-30% of the time when participants attended to the MDT, they were concurrently driving (Garrison et al., 2012). This is obviously problematic given that this is less time spent attending to the task of driving and observation of activity outside the patrol vehicle, thereby requiring multitasking. While the information available from the MDT can help ease cognitive loads, it can also serve as a powerful distraction. Ample research has demonstrated that multitasking, even in the form of conversing on a cell phone or texting while operating a vehicle is associated with an increase in driving errors (Drews, Pasupathi, & Strayer, 2008; Strayer & Drews, 2004; Strayer, Drews, & Johnston, 2003; Drew, Yazdani, Godfrey, Cooper, & Strayer, 2009; Hosking, Young, & Regan, 2007). For police officers, add in an MDT and in-car controls, and the risk may increase exponentially.

Additionally, research suggests the more one engages in multitasking the more they are susceptible to interference from irrelevant stimuli (Ophir, Nass, & Wagner, 2009). Officers need to be attending appropriately to the information they are receiving but avoid being distracted when their immediate response is not required. Unfortunately, those who believe they are adept at multitasking also tend to be the ones that are the least efficient (Sanbonmatsu, Strayer, Medeiros-Ward, & Watson, 2013), which could lead to officers with poor multitasking ability being overly distracted and missing important information in their environment.

Specific Cognitive Skills Relevant to Police Officer Functioning

Despite the need for police officers to possess a high level of cognitive functioning, little research has been devoted to delineating these requisite skills. Skills hypothesized to be important due to their relationship with multitasking and cognitive efficiency include working memory, attention, and processing speed, all components of fluid intelligence. Working memory is a limited capacity system designed to briefly hold information (i.e., several seconds) and perform mental operations on it. This information can stem from sensory input or may be recalled from long term memory (Strauss, Sherman, & Spreen, 2006). This particular ability is important to police officers because it is necessary for them to observe dynamic situations, apply their training and knowledge, and then make decisions accordingly. Generally, attention is defined as the mental ability to filter information and monitor responses to various stimuli (Strauss et al., 2006) and processing speed is the ability to mentally process and perform cognitive tasks quickly. Working memory (as well as attention and fluid intelligence) has also been noted to be an important predictor of multitasking ability (König, Bühner, & Mürling, 2005; Colom, Martinez-Molina, Shih, & Santacreu, 2010).

Limits on working memory may result in a more impulsive decision making style, which in police officers, could be catastrophic (Hinson, Jameson, & Whitney, 2003). Kleider, Parrot, and King (2009) focused on the impact of working memory and negative emotionality and the influence on police officer shoot/ don't shoot decisions. They found lower working memory capacity was associated with increased likelihood of shooting errors. Additionally, the interaction of lower working memory capacity and negative emotionality was significant and the presence of lower working memory capacity predicted lower discriminability among participants with higher negative emotionality (Kleider et al., 2003). Pressure and anxiety, a natural part of law enforcement can also impact attentional control resources. Nieuwenhuys, Savelsbergh, and Oudejans (2012) demonstrated that anxiety negatively impacted police officers' shooting decisions and posited that it was the impact of the anxiety on attentional control that explained the effect. As such, in this investigation, the officer, in a heightened state of anxiety due to the task at hand, responded to the presented scenario through reliance on threat-related inferences as opposed to objective visual information.

Officers, faced with a constant barrage of stimuli and anxiety provoking situations must have adequate attentional control, working memory, processing speed and the ability to multitask in order to perform the required job related functions optimally. Due to the aforementioned cognitive demands of the position, quantitative measures assessing these areas will be used to evaluate police officers and determine which, if any, predict performance as measured by Supervisor Survey.

Aims/ Hypotheses

Aim I: To examine whether neurocognitive performance will predict field performance ratings in police officers.

Hypothesis: The NIH Toolbox Fluid Composite score and subtest level scores will predict field performance ratings.

Aim II: To examine whether performance on the *CritiCall* multitasking assessment can predict field performance ratings.

Hypothesis: Performance on *CritiCall* will predict field performance ratings, with higher *CritiCall* scores associated with higher field performance ratings.

Aim III: To investigate the relationship between *CritiCall* module scores and the NIH Toolbox Cognition Battery composite and subtest scores and standard neuropsychological tests.

Hypothesis I: *CritiCall* module scores will be significantly correlated with the NIH Toolbox Cognition Battery composite scores and subtest scores.

Hypothesis II: *CritiCall* module scores will be significantly correlated with standard neuropsychological test scores.

Aim IV: To examine whether police officers' scores on the NIH Toolbox Emotion Battery combined with NIH Toolbox Cognition Battery composite scores significantly predict field performance ratings.

Hypothesis: A combination of scores from the NIH Toolbox Emotion Battery and composite scores from the NIH Toolbox Cognition Battery will predict field performance ratings.

CHAPTER THREE
Methodology

Participants

Police officers were recruited from two North Texas municipal police departments and a total of 46 (out of approximately 80 approached) consented to participate (26 from Agency 1 and 20 from Agency 2). Inclusion criteria included employment as a licensed police officer with one of the participating agencies and English fluency. There were seven females and 39 males between 22 and 51 years of age (M = 34.33, SD = 8.57) and 89% (n = 41/46) were Caucasian. Fourteen participants (30%) reported military service and of those, eight had been deployed (57%). Demographic characteristics of the sample are provided in Table 1. Independent t-tests revealed that the two agencies did not significantly differ by age [$t(44) = -0.67$, $p = .505$] but did differ by education [$t(44) = -2.60$, $p = .013$]. Education was evaluated as a covariate in analyses using non-education corrected scores (NIH Toolbox Emotion Battery and *CritiCall*). It was not found to be significant in any model. Of these subjects, 46 completed the NIH Toolbox Cognitive and Emotional Battery and 37 completed all aspects of the study. Of the nine that did not take *CritiCall*, there were eight males and one female with ages between 22 and 47. Three were African American, one was Hispanic, and five were Caucasian. Participation drop out appeared to be random.

Table 1
Demographics

	Agency 1 N = 26	Agency 2 N = 20	Agencies combined N = 46
Age (years)			
Mean (SD)	33.58 (8.95)	35.30 (8.18)	34.33 (8.57)
Range	22-51	22-47	22-51
Education (years)			
Mean (SD)	15.15 (1.52)	16.20 (1.11)	15.61 (1.44)
Range	11-16	14-18	11-18
	N (%)	*N* (%)	*N* (%)
Sex (*n* male, %)	23 (89)	16 (80)	39 (85)
Handedness (n right, %)	25 (96)	18 (90)	43 (94)
Race (n Caucasian, %)	21 (21)	20 (100)	41 (89)
Ethnicity (n Non-Hispanic, %)	22 (85)	19 (95)	41 (89)

Materials

Standard neuropsychological measures were selected to assess cognitive domains thought to be most important in police officer functioning. Specifically, measures of attention, working memory, processing speed, multitasking ability and emotional functioning were included. *CritiCall* is a pre-employment screening device used with prospective dispatchers to assess multitasking ability and has been shown in a content-related validation study to have a significant relationship to supervisor ratings of police dispatchers (Biddle Consulting Group, 2013). It has not been evaluated as a pre-employment tool in police officers. The NIH Toolbox core cognitive measures were chosen as they provide a brief, global measure of general cognitive functioning as well as a fluid intelligence composite composed of the skills theorized to be necessary for multitasking, decision making, and successful task switching. The NIH Toolbox emotional core measures were chosen to provide a brief assessment of the officers' emotional functioning, as some studies have suggested a relationship between aspects of psychological functioning and fitness for duty. Finally, three standard neuropsychological tests that assess attention and working memory (*WAIS-IV Digit Span*), executive functioning (*Neuropsychological Assessment Battery Categories*), and processing speed (*Symbol Digit Modalities Test*). These standard measures and the NIH Toolbox cognitive battery were compared to *CritiCall* module scores in order to ascertain which cognitive abilities are being tapped by the multitasking instrument.

CritiCall

CritiCall is a computer based, pre-employment public safety dispatcher testing program designed to assess an applicant's ability to perform tasks in a simulated dispatch environment which requires a high level of multitasking. For applicants, it provides a realistic job environment preview, and for agencies, it provides a sample of the applicant's ability to manage the work environment demands. Applicants are required to complete various tasks including dispatching the appropriate agency (e.g. fire/ EMS/ police) to simulated emergencies, data entry, recall of entered data, and basic map skills. The corrected validity coefficient between composite *CritiCall* test scores and job performance as measured by the employees' supervisors was reported as 0.44 (Biddle Consulting Group, 2013). *CritiCall* offers multiple customizable modules, of which Agency 1 Communications utilizes nine, including: *Decision Making*, *Data Entry MT*, *Data Entry MT (audio)*, *Call Summarization 2 MT*, *Character Comparison*, *Memory Recall*, *Memory Recall-Numeric (audio)*, *Prioritization*, and *Map Reading*. *Decision Making* yields an individual score as a percentage correct. *Data Entry MT* and *Data Entry MT (audio)* yield two scores, a typing score in key strokes per hour and an emergency response score that indicates the respondent's accuracy on the decision making scenarios. The remainder of the modules yield individual scores as a percentage correct. The North Texas Emergency Communications Center (NTECC), Agency 2 dispatchers utilize the same modules as Agency 1, in addition to four others. The additional modules are *Call Summarization 1*, which yields a primary score for percentage correct and the emergency response score, *Cross Referencing (audio)*, *Spelling*, and *Sentence Clarity* (also yielding a percentage correct score). Both agencies obtain two average scores. One is an average of the percentage correct across all modules (this is the *Overall Non-Data Entry* score) and the other is an average of keystrokes per hour for all data entry modules (*Overall Data Entry* score). Only modules used by both agencies were included in the analyses. See Appendix A for descriptions of the *CritiCall* modules and the psychometric properties of the test.

National Institutes of Health (NIH) Toolbox Cognition Battery

The NIH Toolbox is a computerized test battery developed by the NIH to assess cognition, emotional, motor, and sensory functioning in a variety of populations ages 3 to 85 (Weintraub et al., 2013). The Cognition Battery assesses the following cognitive domains: language, episodic memory, executive function, working memory and processing speed. The seven core measures available are the *Picture Vocabulary Test*, *Flanker Inhibitory Control and Attention Test*, *Dimensional Change Card Sort Test*, *Picture Sequence Memory Test*, *List Sorting Working Memory Test*, *Pattern Comparison Processing Speed Test*, and *Oral Reading Recognition Test*. From the seven core tests, three composite scores are produced: Crystallized

Cognition, Fluid Cognition, and Cognitive Function. Further descriptions of the composite scores and individual tests, including reliability and validity information can be found in Appendix A.

NIH Toolbox Emotion Battery

The NIH Toolbox Emotion Battery assesses four main components of emotional functioning including psychological well-being, social relationships, stress and self-efficacy, and negative affect (Slotkin et al., 2012). Additional information regarding each component and individual surveys, including reliability and validity data, can be found in Appendix A.

Neuropsychological Assessment Battery (NAB) Categories Test

The *NAB Categories* test is designed to assess components of executive functioning, including: mental flexibility, ability to sort and classify, novel problem solving, and generativity (Stern & White, NAB manual, 2009). These skills are important to assess due to their hypothesized relationship to decision making and multitasking in police officers. There are two cards and the test stimuli include both visual (photographs) and verbal (written details) information to be used in sorting, with six pictures per test card. Additional information can be found in Appendix A.

Symbol Digit Modalities Test (SDMT)

SDMT assesses attention, visual scanning, and motor speed and is administered in a written format (Smith, 1991). Attention and processing speed are important abilities related to multitasking, decision making, and functioning in a complex, dynamic environment such as law enforcement. Further information can be found in Appendix A.

WAIS-IV Digit Span

The *WAIS-IV Digit Span* was selected because it assesses simple attention and working memory, cognitive abilities important for decision making and multitasking ability. Additional information can be found in Appendix A.

Supervisor Survey

The Supervisor Survey is one of the few available police-specific rating scales and has been used to evaluate the predictive validity of the MMPI-2-RF in police officers (Tarescavage, Brewster, Corey, & Ben-Porath, 2015) and was used as the primary outcome measure to provide an indirect assessment of field performance. It consists of 11 domains in addition to a global rating: “emotional control and stress tolerance problems,” “routine task performance,” “decision making and judgment problems,” “feedback acceptance problems,” “assertiveness problems,” “social competence and teamwork problems,” “integrity problems,” “conscientiousness and commitment problems,” “substance use problems,” “impulse control problems,” and “potential for inappropriate aggression.”[1]

Procedure

This study was approved by the University of Texas Southwestern Medical Center Institutional Review Board and each participant provided written informed consent prior to participation. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Texas Southwestern Medical Center (Harris et al., 2009). Subjects were administered *CritiCall*, the National Institute of Health (NIH) Toolbox Cognition and Emotion Batteries, the *Neuropsychological Assessment Battery (NAB) Categories* subtest,

---

[1] Supervisory and field training officer evaluations have validity among police commanders as acceptable measures of police officer performance (Sanders, 2008).

Symbol Digit Modalities Test (SDMT), and the *Wechsler Adult Intelligence Scale, 4th edition (WAIS-IV) Digit Span* subtest. Neuropsychological tests were administered in the following order: *WAIS-IV Digit Span, SDMT*, *NAB Categories,* NIH Toolbox Cognition Battery, and Emotion Battery. Only the *CritiCall* modules used by both police agencies were used in the statistical analyses. Test order was not counterbalanced, though in all but four participants, test administration was separated by greater than one week. The Supervisor Survey was sent via email to each officer's respective first line supervisor or field training officer. Once completed, a total score was calculated for each domain. Missing data in the Supervisor Survey was handled by calculating a mean item score for that individual and imputing that value into missing fields. Totals were then calculated for each domain and summed for a total score. The domain "potential for inappropriate aggression" was not included in the final score due to redundancy with an item in the "emotional control and stress problems" domain. Likewise, the "global rating" was not included as it would have been redundant. Outliers were examined in each analysis for influence. The following Supervisor Survey domains were used as outcome variables for the regression analyses: "emotional control and stress problems," "routine task performance," "decision making and judgment problems," "feedback acceptance problems," "assertiveness problems," social competence and teamwork problems," "integrity problems," "conscientiousness and commitment problems," "impulse control problems," and "supervisor evaluation total." "Substance use problems" was not tested due to no variability in the scores. Spearman correlations were calculated to compare *CritiCall* module and overall scores to the NIH Toolbox Cognitive Battery (subtest and composite corrected t-scores) and to the standard neuropsychological tests corrected t-scores. Stepwise linear regressions were used to test for significant predictors of the outcome variable. Analyses were performed using IBM SPSS Statistics 24.

CHAPTER FOUR

Results

Means and standard deviations for the total sample and by agency for all outcome variables examined can be found in Tables B1-B5 in Appendix B. The two police agencies differed significantly on two variables, NIH Toolbox Emotion Battery subtest Meaning & Purpose and NAB Categories. Meaning & Purpose was not found to be an important predictor in any model. In the models that included NAB Categories, agency was entered as a covariate and examined for significance.

Aim I hypothesis that the NIH Toolbox Fluid Composite score would predict field performance ratings as rated by the Supervisor Survey, was not supported. Regression results did not find that the Fluid Composite was a predictor of the total or individual domain scores from the Supervisor Survey ratings. The subtests of the Fluid Composite (*Flanker Inhibitory Control and Attention Test; Dimensional Change Card Sort Test, Picture Sequence Memory Test, Pattern Comparison Processing Speed Test; and List Sorting Working Memory Test*) were subsequently analyzed and were also not predictive of “supervisor evaluation total” or domain scores.

Aim II hypothesis stated that performance on *CritiCall* would predict scores on the Supervisor Survey. A significant model emerged for “conscientiousness and commitment problems” domain. *Map Reading* was a significant predictor: $F(1, 34) = 6.46$, $p = .016$, and accounted for 13.5% of the variance. See Table 4 for the regression coefficients. No other significant models emerged for other domains or the supervisor evaluation total score.

Aim III hypothesis I stated the module scores from *CritiCall* would be significantly correlated with the NIH Toolbox Cognition Domain composite and subtest scores. This hypothesis was supported and several modules from *CritiCall* were found to correlate significantly with both subtest and composite scores from the NIH Toolbox Cognition Battery that assess processing speed and crystallized abilities, with modest positive correlations ranging from r = .34 to r = .44 (p's < .05). See Table 2 for significant results and Table C1 in Appendix C for the correlation matrix of all scores.

Table 2: Significant Correlations between CritiCall and NIH Toolbox Cognition Battery

	Dimensional Change Card Sort	Fluid Composite	Cognitive Function Composite	Crystallized Composite	Oral Reading Recognition
Data Entry MT (audio)	0.44 *p* = .006	0.40 *p* = .013	0.35 *p* = 0.036		
Overall Data Entry	0.34 *p* = .038				
Map Reading				0.35 *p* = .032	
Overall Non-Data Entry					0.41 *p* = .012

Aim III hypothesis II stated that *CritiCall* module scores would be significantly correlated with standard neuropsychological test scores. Similar to hypothesis I, this was supported, as several scores showed modest correlations with measures of processing speed, working memory, and executive function. See Table 3 for significant results and Table C2 in Appendix C for the full correlation matrix.

Table 3: Significant Correlations between CritiCall and Standard Neuropsychological Tests

	Longest Digit Forward t-score	Longest Digit Backward t-score	Digit Span Total Scaled Score	SDMT Written t-score	NAB t-score
Memory Recall-Numeric (Audio)	0.62 *p* < .001	0.37 *p* = 0.023	0.70 *p* < .001		
Map Reading	0.45 *p* = .005		0.37 *p* = .022		
Overall Non-Data Entry	0.43 *p* = .008		0.51 *p* = .001		
Call Summarization 2 MT		0.33 *p* = .048			
Overall Data Entry				0.36 *p* = 0.025	
Data Entry MT				0.42 *p* = .009	
Character Comparison					-0.51 *p* = .001

Aim IV hypothesis stated that a combination of scores from the NIH Toolbox Emotion Battery and composite scores from the Cognition Battery would predict field officer performance ratings. Using the stepwise method, a significant model emerged using *Self-Efficacy* as a predictor for “routine task performance” and “assertiveness problems.” Perceived Stress was a significant predictor for “decision making and judgment problems,” “impulse control problems,” “social competence and teamwork problems,” and the “supervisor evaluation total.” *General Life Satisfaction* was a predictor for “conscientiousness and commitment problems” and the “overall rating.” *Instrumental Support* was a significant predictors of “integrity problems.” Finally, *Emotional Support* was a significant predictor of "feedback acceptance problems." These predictors were in the expected direction. See Table 4 for the regression coefficients. Subtest level scores from the NIH Toolbox Cognition Battery were also examined but none was predictive of the "supervisor evaluation total" score or domain scores.

Table 4: Regression Coefficients for the Variables Entered into the Models

Outcome	Predictor	B	SE B	β	*p*	Adj R^2
Conscientiousness and commitment problems	Map Reading	0.062	0.025	0.4	0.016	0.135
Conscientiousness and commitment problems	General Life Satisfaction	-0.219	0.065	-0.458	0.002	0.192
Overall Rating	General Life Satisfaction	-0.087	0.024	-0.479	0.001	0.212
Supervisor evaluation total	Perceived Stress	1.602	0.551	0.405	0.006	0.145
Social competence and teamwork problems	Perceived Stress	0.314	0.104	0.414	0.004	0.152
Impulse control problems	Perceived Stress	0.077	0.027	0.397	0.006	0.138
Decision making and judgment problems	Perceived Stress	0.517	0.192	0.375	0.01	0.121
Integrity problems	Instrumental Support	-0.066	0.027	-0.36	0.017	0.109
Assertiveness problems	Self-Efficacy	-0.069	0.031	-0.322	0.029	0.083
Routine task performance	Self-Efficacy	-0.305	0.137	-0.319	0.031	0.081

Feedback acceptance problems	Emotional Support	-0.05	0.018	-0.396	0.007	0.137

CHAPTER FIVE

Conclusions and Recommendations

Discussion

Police officers are required to function in a complicated environment while integrating multiple sources of stimuli and responding appropriately in a dynamic environment. Pre-employment psychological evaluations are a requirement to become licensed as a police officer in the State of Texas and most states in the country and many instruments have been used and investigated for this purpose, often yielding mixed results. For example, the study by Mills and Stranton (1982) found no relationship between performance on the MMPI and success as a police officer. In contrast, another study by Simmers and colleagues (2013) found the MMPI MMPI-2, and IPI showed "modest" correlations (r = .05- .44) depending on performance variables evaluated. Various tests are used in the evaluation of prospective police officers, depending on the clinician, and the State of Texas offers little guidance beyond the requirement that the officer be in "satisfactory psychological and emotional health." In addition, the testing procedures currently in use have not evolved in decades, while the job has increased in scope and complexity. Thus, the goals of this investigation were to examine whether standard measures of multitasking and the component cognitive skills (working memory, attention, and processing speed) could be used to predict success in law enforcement.

In the current study, police officers' skills were evaluated by the NIH Toolbox Cognition Battery as well as standard neuropsychological tests and pre-employment tests normally used for dispatchers, i.e., *CritiCall*. Officers were also administered the NIH Toolbox Emotional Battery as a brief assessment of psychological functioning. The outcome measure was the Supervisor Survey, an instrument designed and used in a previous investigation to evaluate police officers' performance (Tarescavage, Brewster, Corey, & Ben-Porath, 2015). Relationships between *CritiCall* and the NIH Toolbox Cognition Battery and standard neuropsychological tests were also tested and several significant correlations were observed.

There were a number of important findings in this research. For instance, using linear regression modeling, the cognitive tests or *CritiCall* module scores were not predictive of the total score on the Supervisor Survey. However, one *CritiCall* module score, *Map Reading*, significantly predicted the "conscientiousness and commitment problems" domain, accounting for 13.5% of the variance. It is thought this was a significant predictor due to *Map Reading* relying more on crystallized cognitive abilities, which subsequently predicted more difficulties in this area. The "conscientiousness and commitment problems" domain included items such as proper documentation of officer activities and finishing assignments, things postulated to be more related to knowledge gained through training and experience, similar to *Map Reading*, which relies on knowledge of traffic laws and the ability to navigate the most direct route.

One score from the emotional battery, *General Life Satisfaction*, was found to be a significant predictor of "conscientiousness and commitment problems" (accounting for 19.2% of the variance) and the "overall rating," (21.2% of the variance). This suggests that those who evaluate their lives and find less satisfaction are rated as performing worse overall on the supervisor survey. In addition, they are perceived as having less ambition, less initiative, and more difficulty completing tasks and paying attention to detail. *Instrumental Support* was also found to be a significant predictor of “integrity problems,” though only accounting for 10.9% of the variance. It appears that those who perceive they have more support from social contacts are viewed as more trustworthy and ethical than officers with lower scores on *Instrumental Support*.

Self-Efficacy significantly predicted “routine task performance” (accounting for 8.1% of the variance) and “assertiveness problems” (8.3% of the variance). *Self-Efficacy* evaluates how much one believes they are able to function and have control over events in their life. Thus, officers with higher beliefs in their own efficacy are evaluated as being more assertive compared to officers with lower self-efficacy and evaluated to have less problems with routine task performance including radio operation, paperwork, and other routine tasks. Greater scores on *Emotional Support*, or the perception one has an empathic social network available if needed, were associated with less reported problems accepting feedback. Thus, officers who feel they have social contacts available to them may find it easier to receive constructive criticism regarding their activities.

Finally, higher scores on the *Perceived Stress* measure were found to significantly predict “decision making and judgment problems” (12.1% of the variance), “social competence and teamwork problems” (15.2 % of the variance), “impulse control problems” (13.8% of the variance), and the “supervisor evaluation total” (14.5% of the variance). These results suggest that the more perceived stress an officer has and the less coping resources, the more difficulties they have in these domains, including more problems overall. This may include problems following instructions, making decisions, not filing appropriate criminal charges, poor common sense, poor social skills, and acting without thinking things through. This also supports the study by Kleider and colleagues (2009) that found lower working memory combined with higher stress was associated with an increase in shooting errors.

Overall, the variance accounted for in the supervisor survey ratings was less than 25% for these emotional measures, which is generally consistent with existing, albeit limited research into the relationship between psychological measures and police officer functioning (Tarescavage, Brewster, Corey, & Ben-Porath 2015). Additional stepwise linear regressions were subsequently conducted using scores from the standard neuropsychological tests to evaluate if any were predictive of the Supervisor Survey domains and, if so, if any added to the predictive ability of the emotional measures. Several significant predictors were found. *Longest Digit Backward*, a measure of attention and working memory, was found to significantly predict the following domains: “emotional control and stress problems” (6.5% of the variance), “social competence and teamwork problems” (10.8% of the variance), and the “supervisor evaluation total” (7.6% of the variance). *NAB Categories* significantly predicted “routine task performance” (7.9% of the variance). Because Agency 1 and Agency 2 differed significantly on the *NAB Categories* score, agency was entered as an additional predictor but was not significant. *SDMT* and *Longest Digit Backward* were significant predictors of “decision making and judgment problems” (17.4% of the variance) and “impulse control problems” (15.5% of the variance). These results suggest that lower scores on measures of working memory, mental flexibility, problem solving, and processing speed are related to more problems in these domains. *Longest Digit Forward* and *Digit Span Scale* predicted “assertiveness problems” (16.8% of the variance), with higher scores on *Longest Digit Forward* and lower scores on *Digit Span Scale* being associated with a more passivity. Regression coefficients can be found in Table D1 in Appendix D.

The next step was to evaluate the standard neuropsychological tests used in this study alongside the emotional measures to determine if the combination would yield increased predictive ability. Again, stepwise linear regressions were done and produced some interesting findings. A combination of the variables *Perceived Stress* and *Longest Digit Backward* significantly predicted "decision making and judgment problems," "social competence and teamwork problems," "impulse control problems," and "supervisor evaluation total" accounting for 29%-35% of the total variance. Regression coefficients can be found in Table D2 in Appendix D. This seems to suggest that lower scores on a task of working memory combined with higher levels of perceived stress are associated with more problems in the above areas, as rated by the officers' respective supervisors. Research has demonstrated a relationship between limits on working memory and more impulsive decision making (Hinson et al., 2003) as well as the impact of working memory and negative emotionality on officers' shoot/ don't shoot decisions (Kleider et al., 2009). An increased likelihood of shooting errors was associated with lower working memory capacity and a significant interaction effect was found between lower working memory and higher negative emotionality (Kleider et al., 2009). Current results are generally in keeping with what has been found in these other studies.

Several modules from *CritiCall* were found to correlate significantly with both subtest and composite scores from the NIH Toolbox Cognition Battery and standard neuropsychological tests that assess attention, working memory, processing speed, and crystallized abilities, with positive correlations ranging from $r = 0.33$ to 0.41 (p's $< .05$) and r's $= 0.42$ to 0.70 (p's $<.01$). One negative correlation was revealed between *Character Comparison* and the *NAB Categories* total raw score ($p < .01$). This is an interesting finding and may represent a spurious correlation; however, one possibility is that *Character Comparison*, on its face, may have appeared to be a relatively easy task. Thus, some participants may have devoted less attention to it as compared to other, more difficult tasks, like *NAB Categories* which requires more sustained mental effort to complete. Ergo, this easy task with a low demand on cognitive resources may not have been deemed as important as the other, more demanding tasks, resulting in decreased attention and accuracy (e.g. see Kahneman, 1973).

CritiCall is marketed as a task to evaluate a prospective dispatcher's ability to multitask, an ability that is actually driven by one's working memory (as well as attention, processing speed, and fluid intelligence) (König et al., 2005; Colom et al., 2010). As such, *CritiCall* modules were predicted to correlate with similar measures of the NIH Toolbox and standard neuropsychological measures. *Data Entry MT* is said to assess the ability to read written data and enter it into entry fields while responding to scenarios that appear at intervals in the lower right of the screen. *Data Entry MT (audio)* is a similar task but involves the ability to hear audible data and enter it accurately while also responding to scenarios. These scenarios involve a depicted emergency and the test taker must select the appropriate agency to dispatch according to rules they have been provided. *Data Entry MT (audio)* was moderately correlated with *Dimensional Change Card Sort*, Fluid Composite, and the Cognitive Function Composite. *Dimensional Change Card Sort* assesses cognitive flexibility (set shifting) and attention. Thus, this suggests that *Data Entry MT (audio)* is related to the ability to attend and shift between tasks and respond accurately and quickly. *Data Entry MT* was moderately correlated with *SDMT* written t-score and likely represents it assessing aspects of processing speed. *CritiCall Memory Recall-Numeric (audio)* was moderately correlated with t-scores from *Longest Digit Span Forward* and *Backward* and strongly correlated with the overall *Digit Span* scale score. This module assesses one's ability to hear, memorize, and recall 7-digit phone numbers, a task very similar to *Digit Span*, and likely assesses attention and working memory. *Map Reading*, the ability to examine a simple map and plan out the most direct route while obeying all traffic laws was moderately correlated with the *NIH Crystallized Composite*, *Longest Digit Forward* t-score and the overall *Digit Span* scale score. Driving and knowledge of traffic laws, especially in police officers, are overlearned abilities acquired through both practical and classroom training and experience. Thus, its relationship to *Crystallized Cognition* likely reflects them using these skills and knowledge to complete the task. The correlation with *Longest Digit Span* and *Digit Span* total scale score suggests a relationship with attentional resources. *CritiCall Call Summarization 2 MT* measures the ability to hear, understand, and then use audible information while entering it accurately. It also contains the emergency response scenarios presented at various intervals during the test. The score represents a combination of data accuracy and responses to the decision making and multiple choice items. It was moderately correlated with *Longest Digit Span Backward* t-score, which may represent the need for adequate working memory in order to perform well on the task. The *CritiCall Overall Non-Data Entry* score was moderately related to *Longest Digit Span Forward* t-score, *Digit Span* scale score, and *NIH Oral Reading Recognition*.

This is not a surprising finding as the *Overall Non-Data Entry* score is an average of all non-data entry modules including: *Map Reading*, *Memory Recall-Numeric (audio)*, and *Call Summarization 2 MT*. Finally, the *CritiCall Overall Data Entry* score, an average of the data entry modules and including *Data Entry MT* and *Data Entry MT (audio)*, was moderately correlated with *NIH Toolbox Dimensional Change Card Sort* and *SDMT* written t-score, reflecting its relationship with processing speed. These results suggest that, as predicted, aspects of *CritiCall* are significantly related to subtests and composites scores of the NIH Toolbox Cognition Battery.

Conclusions

This study represents some of the first research into the hypothesized cognitive skills that are associated with success in police officers as measured by supervisor ratings. The goals of this study were to ascertain if the NIH Toolbox Cognition Battery and/or Emotional Battery and/or *CritiCall* predicted the total or individual domain scores from the rating instrument. The NIH Toolbox Fluid Composite score and subtest level scores were not found to be predictive of the supervisor survey total or domain scores, and one module score of *CritiCall*, *Map Reading,* significantly predicted “conscientiousness and commitment problems.” Of the NIH Toolbox Emotional Battery, *General Life Satisfaction* predicted conscientiousness and commitment problems,” and the “overall rating.” *Instrumental Support* predicted “integrity problems.” *Self-Efficacy* predicted “assertiveness problems” and “routine task performance” and *Emotional Support* predicted “feedback acceptance problems” Finally, *Perceived Stress* predicted “social competence and teamwork problems,” “decision making and judgment problems,” “impulse control problems," and the "supervisor evaluation total." Additional regressions were conducted and found that several standard neuropsychological variables were significant predictors of some domains of the Supervisor Survey and, when combined with emotional measures, provided even stronger predictive ability. *Perceived Stress* and *Longest Digit Backward* were significant predictors of "decision making and judgment problems," "social competence and teamwork problems," "impulse control problems," and "supervisor evaluation total" accounting for 29% - 35% of the variance, suggesting that increased stress in one's life, combined with lower working memory, is associated with increased difficulties in these areas. Another goal was to explore the relationships between *CritiCall* module scores and the NIH Toolbox Cognition Battery and the standard neuropsychological tests. These analyses revealed moderate to strong correlations between some *CritiCall* module scores and tests that assessed attention, working memory, processing speed, and crystallized cognitive abilities.

Current pre-employment psychological evaluations for police officers have seen few updates since their inception even though the nature of policing has seen many changes. Police officers are required to operate in an increasingly distracting environment (the patrol car) and still maintain vigilance. The ability to multitask is required; however, the requisite cognitive skills needed to multitask effectively are not directly evaluated prior to employment, which can lead to attrition during the police academy and training and to potentially serious mistakes in judgment. These preliminary results suggest that the addition of two relatively brief measures (to evaluate perceived stress and working memory) to the current testing may be useful for agencies in order to determine which applicants would be best suited for the position. While lower scores in these areas may not preclude one from becoming a police officer, it may help with placement. For example, an applicant with lower working memory scores may be better suited to a smaller agency with a lower call volume and crime rates. Future research is needed to continue to evaluate these measures for their predictive ability.

Limitations

Limitations of this investigation include the relatively small sample size of primarily Caucasian males. It is comprised of two metropolitan police agencies, which may limit generalization to less urban agencies. Also, many module scores on *CritiCall* had limited variability (>50% of scores were 90-100), and as such, correlations between those modules and the neuropsychological tests may have been constrained. In addition, the outcome measure, the Supervisor Survey, has not been validated or studied extensively in a large sample; however, it has been used as a primary outcome measure in previous research with the MMPI-2 RF (Tarescavage, Brewster, Corey, & Ben-Porath, 2015), and at the time of this investigation, represented one of the few options to assess the outcome ratings. There are no validated police-specific outcome measures available, and the two agencies used in this study collect different information for yearly performance evaluations, which did not lend itself to scientific analysis. Further research is needed to evaluate the Supervisor Survey and should include a factor analysis in larger samples. Additionally, this line of research exploring the relationship between cognitive variables and police officer outcomes should be expanded to other police agencies of varying sizes, both rural and metropolitan, and seek to recruit diverse participants.

APPENDIX A
Test Characteristics and Psychometric Properties

CritiCall

According to the *CritiCall* website (Public-Safety Dispatcher, 2017), new dispatchers, poorly matched to the position and/or agency, quit early in employment, frequently prior to the end of the probationary period. This employee turnover costs the hiring agency in both time and money. *CritiCall* has been successful in streamlining the hiring process and ensuring applicants have the requisite skills and aptitude to be successful (Kilday, 2003). Overall reliability of the *CritiCall* test, as used for pre-employment testing of dispatchers, was reported as 0.92, and the uncorrected validity coefficient between composite CritiCall test scores and job performance as measured by the employees' supervisors, was 0.41 (corrected = 0.44) (Public-Safety Dispatcher, 2017). Biddle Consulting Group (2013) reported that per the United States Department of Labor, this is classified as "very beneficial."

CritiCall offers multiple, customizable test modules. Current modules utilized by Agency 1 Communications include: *Decision Making*, *Data Entry Multitasking (MT)*, *Data Entry MT (audio)*, *Call Summarization 2 MT*, *Character Comparison*, *Memory Recall*, *Memory Recall-Numeric (audio)*, *Prioritization*, and *Map Reading*. North Texas Emergency Communications Center (NTECC) use the same modules plus four more: *Call Summarization 1* (audio), *Cross-Referencing* (audio), *Spelling* (audio), and *Sentence Clarity*. Agencies set their own pass/ fail criteria.

The *Decision Making* module involves presenting various emergency situations to the examinee to which they must dispatch the most appropriate agency (police, fire, ems, or public utilities). Tasks from this module are used throughout testing to simulate the multi-tasking nature of the job and assess the respondent's ability to make decisions quickly and accurately according to the stated decision rules (CritiCall Test Descriptions, 2017). The *Data Entry* module evaluates the ability of the test taker to read written data and enter it correctly. Similarly, the *Data Entry* (audio) module assesses his/ her ability to hear information vocally and enter it correctly. *Call Summarization 1* assesses one's ability to hear, comprehend, and summarize a short story and then answer multiple choice questions about the information. *Call Summarization 2* is a similar task but respondent listens to a telephone call and enters details. Multi-tasking ability is assessed through the decision-making task. *Call Summarization 2 MT* (multi-tasking) is similar to *Call Summarization 2* but contains more multiple choice and decision making items. The score reflects the respondent's accuracy of data entered plus responses to multitasking and multiple choice items (CritiCall Test Descriptions, 2017). *Cross Referencing* (audio) evaluates one's ability to locate information that is requested audibly and answer questions based on it. *Character Comparison* tests how well the respondent can compare and contrast written text by presenting him/ her with characters and multiple choice items. He/ she must then select the matching text from a group of similar character sequences. *Memory Recall* is simply the ability to learn and recognize words. The respondent is shown word pairs, which subsequently disappear from the screen. They are then shown one of the words from the pair and asked to identify the other word. *Memory Recall Numeric* presents seven-digit phone numbers to test takers, which he/ she is then asked to enter a short time later. *Prioritization* assesses the ability to use decision rules to properly prioritize incidents as they are presented. *Map Reading* requires the respondent to determine the most direct route while obeying all traffic rules and to answer general questions about routes and maps. *Spelling* measures the ability to correctly spell similar sounding words, which have different spellings and meanings based on the context in which they are used. In *Sentence Clarity*, respondents must choose the passage with the most clearly communicated meaning.

NIH Toolbox Cognition Battery

The NIH Toolbox Cognition Battery is made up of seven core subtests that contribute to three composite scores: Fluid Cognition Composite, Crystallized Composite, and Cognitive Function composite. The Fluid Cognition Composite is comprised of scores from the following tests: *Flanker Inhibitory Control and Attention*, *Dimensional Change Card Sort*, *Picture Sequence Memory*, *List Sorting Working Memory*, and *Pattern Comparison Processing Speed* with higher scores indicating higher level of functioning. These measures are postulated to assess "fluid" abilities, or the skills needed to solve problems, reason, think and act quickly, encode memories, and adapt and respond to novel experiences and are thought to be more dependent on innate ability as opposed to gained from education and experience (Slotkin et al, 2012). See below for a brief description of each subtest and the domain assessed:

- *Flanker Inhibitory Control and Attention* assesses attention and executive functioning. The participant is presented with a series of arrows pointing different directions and is tasked with choosing the button that corresponds with the direction the middle arrow is pointing. At times, the arrows point the same way and others, the middle arrow points the opposite. Thus, one must inhibit their attention to the surrounding stimuli and only respond according to the middle stimulus.

- *Dimensional Change Card Sort* measures cognitive flexibility and also attention. Participants are presented with two stimuli that vary by shape and color and instructed to sort the stimuli per the word given on the screen.
- *Picture Sequence Memory* is a measure of episodic memory. Test takers are shown a sequence of pictures and then asked to reproduce the sequence. There are two trials, the second with more pictures added.
- *List Sorting Working Memory* assesses working memory by presenting participants with pictures of items which they then must put in size order from smallest to biggest. The first trial presents one category of objects (food or animals) while the second trial presents both categories with the instructions that the participant put the food in size order and then the animals.
- *Pattern Comparison Processing Speed* evaluates processing speed by presenting test takers with two pictures side by side. They then must select if the pictures are the same or not as quickly as possible.

The Crystallized Cognition Composite includes the *Picture Vocabulary Test* and *Oral Reading Recognition* and represents a more global assessment of verbal reasoning theorized to be a product of one's educational and cultural exposure (Slotkin et al., 2012). The *Picture Vocabulary Test* measures receptive vocabulary by presenting test takers with four pictures and then stating a word. The test taker is instructed to pick the picture that matches the meaning of the word. *Oral Reading Recognition* is a word reading test, in which the test taker reads each word presented and the examiner selects if correct or not correct. The Cognitive Function Composite includes all core measures and represents a global measure of general cognitive functioning similar to a full scale score derived in other available tests (e.g. WAIS-IV). Each subtest and composite score is reported as a fully corrected t-score (age, education, sex, and ethnicity) and an age corrected standard score. T-scores were used for all analyses.

NIH Toolbox Emotion Battery

The psychological well-being domain is comprised of the following surveys: *Positive Affect*, *General Life Satisfaction*, and *Meaning and Purpose* and evaluates both the subjective, experiential and the evaluative aspects of well-being (Slotkin et al., 2012). The social relationships domain focuses on these three aspects in adults: perceived social support (*Emotional Support* and *Instrumental Support*), companionship (*Friendship*, and *Loneliness*), and social distress (*Perceived Hostility* and *Perceived Rejection*) (Slotkin et al., 2012). The stress and self-efficacy domain uses the *Perceived Stress* survey and *Self-Efficacy* survey to evaluate an individual's perception about events and their perceived coping resources (Slotkin et al., 2012). The negative affect domain is comprised of measures assessing three main negative emotions: anger, fear, and sadness through the use of the following surveys: *Anger-Physical Aggression*, *Anger-Hostility*, *Anger-Affect*, *Fear-Affect*, *Fear-Somatic Arousal*, and *Sadness*.

Neuropsychological Assessment Battery (NAB) Categories Test

NAB Categories is a test of executive function normed for adults ages 18-97 and was standardized on citizens or residents of the United States whose primary language was English

(Stern & White, 2009: NAB Manual). Normative data is provided for demographically corrected for age, sex, and education (recommended) and U.S. census matched. The stability coefficient was .54 and the G coefficient (a measurement of equivalent forms reliability) was .89. Validity studies indicated NAB categories is assessing executive functioning as evidenced by moderate correlations with similar assessments.

Symbol Digit Modalities Test (SDMT)

The SDMT was originally published in 1973 and was developed by Adam Smith in order to screen children for cerebral dysfunction (Smith, 1973). A relatively simple test requiring less than five minutes to administer, it requires the examinee to convert geometric shapes into written and/ or oral responses and to do it as fast as they can in 90 seconds thus assessing attention, visual scanning, tracking, and motor speed (Strauss et al., 2006). The maximum raw score is 110 on both the written and oral forms. Digit symbol coding tasks have repeatedly been shown to be highly sensitive to cerebral dysfunction in children and adults (Smith, 1991: SDMT manual). Test-retest reliability for the written form is 0.80 (Smith, 1991).

WAIS-IV Digit Span

Digit Span is a core working memory subtest of the Wechsler Adult Intelligence Scales-IV (WAIS-IV). The test consists of three parts: Digit Span Forward, Digit Span Backward and Digit Span Sequencing. Altogether, this subtest assesses simple attention and working memory. Across age groups, split-half reliability is 0.93 and the stability coefficient is 0.83 (Wechsler, 2008: WAIS-IV Manual).

- Digit Span Forward, the test taker is instructed to repeat a string of digits. The string increases by one digit each trial, to a maximum of nine.
- Digit Span Backward, the test taker is given a string of numbers (up to eight) and must repeat them in reverse order.
- Digit Span Sequencing, the individual is given a string of random single-digit numbers (up to nine) and must repeat them in ascending order.

APPENDIX B
Total Sample and Subgroup Descriptive Statistics

Table B1: CritiCall Scores

		Agency 1	Agency 2	Agencies combined
Decision Making E Resp.	*Mean (SD)*	95.22 (9.47)	95.38 (6.60)	95.28 (8.45)
	Range	60-100	80-100	60-100
	N	23	13	36
Data Entry MT E Resp.	*Mean (SD)*	91.30 (19.38)	96.92 (6.30)	93.33 (16.04)
	Range	10-100	80-100	10-100
	N	23	13	36
Data Entry MT (audio) E Resp.	*Mean (SD)*	90.00 (10.87)	96.15 (6.50)	92.22 (9.89)
	Range	60-100	80-100	60-100
	N	23	13	36
Data Entry MT	*Mean (SD)*	4074.00 (1080.83)	3982.15 (641.97)	4041.73 (941.12)
	Range	2724-7536	2640-4692	2640-7536
	N	24	13	37
Data Entry MT (audio)	*Mean (SD)*	2920.00 (552.85)	2971.38 (377.46)	2938.05 (493.36)
	Range	1596-3852	2436-3588	1596-3852
	N	24	13	37
Call Summarization 2 MT	*Mean (SD)*	81.38 (7.34)	86.46 (8.53)	83.16 (8.05)
	Range	70-97	68-97	68-97
	N	24	13	37
Character Comparison	*Mean (SD)*	92.50 (9.44)	95.38 (7.76)	93.51 (8.89)
	Range	70-100	80-100	70-100
	N	24	13	37
Memory Recall	*Mean (SD)*	95.42 (6.58)	96.15 (6.50)	95.68 (6.47)
	Range	80-100	80-100	80-100
	N	24	13	37
Memory Recall-Numeric (audio)	*Mean (SD)*	86.08 (12.05)	85.54 (10.55)	85.89 (11.40)
	Range	60-100	73-100	60-100

	N	24	13	37
Prioritization	*Mean (SD)*	90.42 (10.42)	91.54 (8.99)	90.81(9.83)
	Range	60-100	70-100	60-100
	N	24	13	37
Map Reading	*Mean (SD)*	78.25 (21.85)	88.00 (22.41)	81.68 (22.24)
	Range	17-100	17-100	17-100
	N	24	13	37
Average Non-Data Entry	*Mean (SD)*	86.04 (5.94)	90.08 (5.60)	87.46 (6.07)
	Range	75-96	76-96	75-96
	N	24	13	37
Average Data Entry	*Mean (SD)*	3497.00 (642.66)	3476.77 (432.09)	3489.89 (571.14)
	Range	2610-5238	2682-4140	2610-5238
	N	24	13	37

Table B2: NIH Toolbox Cognition Battery Scores

		Agency 1 *N* = 26	Agency 2 *N* = 20	Agencies Combined *N* = 46
Picture Vocabulary	*Mean (SD)*	54.58 (5.79)	54.00 (7.31)	54.33 (6.42)
	Range	44-64	36-64	36.64
Flanker	*Mean (SD)*	46.38 (9.63)	42.60 (9.29)	44.74 (9.57)
	Range	32-65	25-58	25-65
List Sorting	*Mean (SD)*	53.77 (7.04)	51.45 (10.56)	52.76 (8.72)
	Range	43-67	37-74	37-74
DCCS	*Mean (SD)*	59.85 (10.78)	55.60 (11.31)	58.00 (11.10)
	Range	34-79	35-77	34-79
Pattern Comparison	*Mean (SD)*	50.65 (14.09)	49.00 (16.81)	49.93 (15.18)
	Range	29-77	21-78	21-78
Picture Sequence Memory	*Mean (SD)*	57.12 (10.20)	50.55 (11.88)	54.26 (11.32)
	Range	40-82	31-74	31-82
Oral Reading Recognition	*Mean (SD)*	54.42 (7.33)	56.05 (8.20)	55.13 (7.68)
	Range	39-71	43-71	39-71
Fluid Composite	*Mean (SD)*	55.23 (9.15)	49.60 (13.15)	52.78 (11.29)
	Range	40-72	27-72	27-72
Crystallized Composite	*Mean (SD)*	55.04 (6.47)	55.55 (7.66)	55.26 (6.93)
	Range	40-69	39-66	39-69
Cognitive Function Composite	*Mean (SD)*	56.04 (7.31)	52.95 (10.05)	54.70 (8.65)
	Range	43-72	33-70	33-72

Table B3: NIH Toolbox Emotion Battery Scores

		Agency 1 N = 26	Agency 2 N = 20	Agencies Combined N = 46
Positive Affect	*Mean (SD)*	48.27 (5.79)	47.30 (5.59)	47.85 (5.66)
	Range	36-64	33-59	33-64
General Life Satisfaction	*Mean (SD)*	55.15 (7.99)	53.15 (10.41)	54.28 (9.07)
	Range	41-75	21-65	21-75
Meaning & Purpose	*Mean (SD)*	57.42 (8.24)	52.05 (5.73)	55.09 (7.67)
	Range	39-68	39-64	39-68
Emotional Support	*Mean (SD)*	48.38 (10.20)	51.75 (9.91)	49.85 (10.11)
	Range	25-63	31-62	25-63
Instrumental Support	*Mean (SD)*	45.46 (8.23)	47.45 (8.80)	46.33 (8.45)
	Range	26-57	31-63	26-63
Friendship	*Mean (SD)*	50.85 (7.42)	51.20 (8.92)	51.00 (8.01)
	Range	35-67	36-67	35-67
Loneliness	*Mean (SD)*	51.96 (8.52)	49.15 (9.03)	50.74 (8.76)
	Range	37-68	37-65	37-68
Perceived Rejection	*Mean (SD)*	50.31 (9.77)	47.95 (8.77)	49.29 (9.33)
	Range	36-72	36-63	36-72
Perceived Hostility	*Mean (SD)*	53.00 (7.05)	52.15 (8.07)	52.63 (7.44)
	Range	42-73	33-65	33-73
Self-Efficacy	*Mean (SD)*	56.42 (7.79)	54.70 (6.33)	55.67 (7.17)
	Range	41-68	36-68	36-68
Perceived Stress	*Mean (SD)*	45.65 (7.46)	46.80 (7.24)	46.15 (7.31)
	Range	28-55	34-61	28-61
Fear-Affect	*Mean (SD)*	52.92 (5.21)	50.20 (6.68)	51.74 (5.98)
	Range	42-62	34-61	34-62
Fear-Somatic Arousal	*Mean (SD)*	51.38 (10.09)	47.10 (8.08)	49.52 (9.42)
	Range	38-79	38-64	38-79
Sadness	*Mean (SD)*	46.38 (7.65)	44.60 (9.03)	45.61 (8.23)
	Range	34-62	30-62	30-62
Anger-Affect	*Mean (SD)*	50.00 (6.50)	48.05 (7.92)	49.15 (7.14)
	Range	35-67	31-64	31-67
Anger-Hostility	*Mean (SD)*	46.38 (9.10)	47.60 (8.96)	46.91 (8.96)
	Range	36-63	36-62	36-63

Anger-Physical Aggression	*Mean (SD)*	49.92 (7.20)	53.15 (8.80)	51.33 (8.00)
	Range	42-64	42-70	42-70

Table B4: Standard Neuropsychological Test Scores

		Agency 1 *N* = 26	Agency 2 *N* = 20	Agencies Combined *N* = 46
NAB Categories	*Mean (SD)*	52.00 (8.49)	46.45 (7.49)	49.59 (8.46)
	Range	35-70	35-61	35-70
Longest Digit Forward	*Mean (SD)*	53.69 (9.66)	53.75 (7.74)	53.72 (8.78)
	Range	37-66	37-66	37-66
Longest Digit Backward	*Mean (SD)*	49.46 (7.03)	48.65 (7.41)	49.11 (7.13)
	Range	38-65	34-64	34-65
Longest Digit Sequence	*Mean (SD)*	50.15 (8.53)	51.65 (9.52)	50.80 (8.90)
	Range	36-72	35-72	35-72
Digit Span Scale	*Mean (SD)*	10.23 (1.97)	10.40 (2.41)	10.30 (2.15)
	Range	6-14	7-17	6-17
SDMT Written	*Mean (SD)*	53.21 (7.84)	51.92 (9.35)	52.65 (8.45)
	Range	38-67	36-76	36-76

Table B5: Supervisor Survey Domain Scores

	Agency 1 *N* = 26	Agency 2 *N* = 20	Agencies Combined N = 46
Emotional control and stress tolerance problems			
Mean (SD)	10.31 (5.53)	8.55 (4.07)	9.54 (4.98)
Range	5-22	5-22	5-22
Routine task performance			
Mean (SD)	12.73 (6.58)	14.4 (7.26)	13.46 (6.86)
Range	7-29	7-32	7-32
Decision making and judgment problems			
Mean (SD)	18.50 (11.18)	17.40 (8.64)	18.02 (10.06)
Range	9-59	9-47	9-59
Feedback acceptance problems			
Mean (SD)	1.96 (1.78)	1.60 (1.05)	1.80 (1.50)
Range	1-7	1-5	1-7
Assertiveness problems			
Mean (SD)	3.15 (1.71)	2.40 (1.19)	2.83 (1.54)
Range	1-7	1-5	1-7
Social competence and teamwork problems			
Mean (SD)	8.92 (5.85)	9.95 (5.22)	9.37 (5.55)
Range	5-29	5-23	5-29
Integrity problems			
Mean (SD)	3.23 (2.37)	3.10 (2.29)	3.17 (2.31)
Range	2-11	2-12	2-12
Conscientiousness and commitment problems			
Mean (SD)	11.31 (6.29)	11.45 (4.47)	11.37 (5.51)
Range	5-34	6-25	5-34
Supervisor evaluation total			
Mean (SD)	72.96 (37.48)	71.85 (32.27)	72.48 (34.94)
Range	39-205	42-179	39-205

APPENDIX C
Correlation Matrices

Table C1: Correlations Between CritiCall and NIH Toolbox Cognition Battery T-Scores

		Picture Vocabulary	Flanker	List Sorting	DCCS	Pattern Comparison	Picture Sequence Memory	Oral Reading Recognition	Fluid Composite	Crystallized Composite	Cognitive Function Composite
Decision Making E Resp Score	*r*	-0.057	-0.121	-0.135	0.213	-0.161	0.063	-0.214	-0.031	-0.156	-0.121
	p	0.739	0.481	0.433	0.213	0.348	0.716	0.211	0.860	0.363	0.482
Data Entry MT E Resp Score	*r*	0.136	-0.043	0.209	0.079	-0.144	0.037	0.112	0.012	0.139	0.073
	p	0.427	0.803	0.221	0.645	0.402	0.832	0.516	0.947	0.418	0.671
Data Entry MT (Audio) E Resp Score	*r*	0.242	0.054	-0.133	-0.067	-0.097	-0.159	-0.080	-0.115	0.110	-0.113
	p	0.155	0.755	0.440	0.697	0.572	0.354	0.641	0.506	0.523	0.513
Data Entry MT	*r*	0.086	0.014	-0.181	0.189	-0.102	0.054	-0.075	0.000	-0.025	-0.045
	p	0.612	0.934	0.282	0.262	0.549	0.753	0.661	1.000	0.884	0.792
Data Entry MT (audio)	*r*	-0.066	0.240	0.245	.443**	0.268	0.110	0.087	.404*	0.007	.345*
	p	0.698	0.152	0.144	0.006	0.109	0.515	0.607	0.013	0.966	0.036
Call Summarization 2 MT	*r*	0.090	0.195	0.091	0.115	0.191	0.089	0.170	0.242	0.134	0.256
	p	0.597	0.247	0.590	0.498	0.258	0.599	0.314	0.148	0.430	0.127
Character Comparison	*r*	-0.150	-0.115	0.097	-0.051	-0.123	0.176	0.019	-0.020	-0.034	-0.057
	p	0.376	0.497	0.570	0.766	0.467	0.297	0.912	0.908	0.840	0.739
Memory Recall	*r*	-0.018	0.077	0.002	-0.005	-0.095	-0.079	0.033	-0.004	-0.043	-0.011
	p	0.916	0.652	0.991	0.977	0.577	0.644	0.846	0.981	0.802	0.949
Memory Recall-Numeric (Audio)	*r*	-0.077	0.113	0.323	-0.095	0.187	0.067	0.217	0.151	0.160	0.186
	p	0.652	0.505	0.051	0.577	0.267	0.694	0.197	0.373	0.344	0.271

Prioritization	*r*	-0.058	-0.148	0.054	-0.123	0.016	-0.098	0.044	-0.106	-0.016	-0.096
	p	0.733	0.383	0.749	0.467	0.927	0.562	0.794	0.531	0.926	0.572
Map Reading	*r*	0.249	-0.012	0.003	-0.071	-0.040	0.195	0.320	0.008	.354*	0.153
	p	0.137	0.945	0.984	0.676	0.815	0.248	0.053	0.962	0.032	0.366
Overall Non-Data Entry	*r*	0.086	-0.035	0.167	-0.075	0.126	0.120	.408*	0.089	0.315	0.234
	p	0.613	0.839	0.324	0.657	0.459	0.480	0.012	0.600	0.058	0.162
Overall Data Entry	*r*	0.007	0.137	-0.008	.343*	-0.004	0.151	-0.028	0.189	-0.036	0.119
	p	0.968	0.418	0.962	0.038	0.981	0.374	0.869	0.264	0.833	0.484

*p < 05

**p < .01

N = 37

Table C2: Correlations Between CritiCall and Standard Neuropsychological Tests

		NAB T-Score	Longest Digit Forward T-Score	Longest Digit Backward T-Score	Longest Digit Sequence T-Score	Digit Span Scaled Score	SDMT Written T-Score
Decision Making E Resp Score	*r*	-0.151	0.102	-0.106	-0.203	0.062	-0.056
	p	0.378	0.555	0.540	0.235	0.719	0.747
Data Entry MT E Resp Score	*r*	0.060	0.259	0.072	0.146	0.183	-0.108
	p	0.729	0.127	0.678	0.395	0.284	0.530
Data Entry MT (Audio) E Resp Score	*r*	0.205	0.041	0.181	-0.056	-0.017	-0.115
	p	0.231	0.813	0.290	0.745	0.922	0.506
Data Entry MT	*r*	0.173	-0.135	0.090	-0.025	-0.123	.422**
	p	0.307	0.427	0.598	0.882	0.470	0.009
Data Entry MT (audio)	*r*	0.056	0.045	0.303	-0.023	0.190	0.262
	p	0.742	0.792	0.068	0.892	0.260	0.117
Call Summarization 2 MT	*r*	-0.126	0.061	.327*	0.202	0.224	0.193
	p	0.458	0.720	0.048	0.230	0.183	0.252
Character Comparison	*r*	-.506**	0.222	0.009	-0.057	0.130	-0.203
	p	0.001	0.186	0.957	0.740	0.444	0.229
Memory Recall	*r*	0.267	0.174	-0.001	-0.024	0.161	0.007
	p	0.110	0.304	0.995	0.888	0.341	0.967
Memory Recall-Numeric (Audio)	*r*	-0.074	.616**	.374*	-0.063	.700**	-0.077
	p	0.662	0.000	0.023	0.710	0.000	0.651
Prioritization	*r*	0.188	-0.084	0.158	-0.200	-0.070	-0.083
	p	0.264	0.622	0.352	0.236	0.679	0.626
Map Reading	*r*	-0.128	.449**	0.137	0.035	.376*	-0.107
	p	0.450	0.005	0.419	0.839	0.022	0.530

Overall Non-Data Entry	*r*	-0.176	.433**	0.235	0.046	.510**	-0.022
	p	0.298	0.008	0.161	0.789	0.001	0.899
Overall Data Entry	*r*	0.114	-0.027	0.187	0.095	0.006	.366*
	p	0.503	0.874	0.269	0.577	0.972	0.026

*p < 05

**p < .01

N = 37

APPENDIX D
Regression Coefficients

Table D1: Regression Coefficients for Standard Neuropsychological Variables

Outcome	Predictor	B	SE B	β	*p*	Adjusted R^2
Emotional control and stress problems	Longest Digit Backward	-0.205	0.101	-0.294	0.048	0.065
Routine task performance	NAB Categories	-0.256	0.116	-0.316	0.032	0.079
Decision making and judgment problems	SDMT Written	-0.315	0.129	-0.334	0.019	0.174
	Longest Digit Backward	-0.343	0.154	-0.305	0.031	
Assertiveness problems	Longest Digit Forward	0.114	0.034	0.649	0.002	0.168
	Digit Span Scale Score	-0.312	0.14	-0.435	0.031	
Social competence and teamwork problems	Longest Digit Backward	-0.279	0.11	-0.358	0.015	0.108
Impulse control problems	SDMT Written	-0.052	0.023	-0.307	0.031	0.155
	Longest Digit Backward	-0.6	0.027	-0.3	0.034	
Supervisor evaluation total	Longest Digit Backward	-1.52	0.702	-0.31	0.036	0.076

Table D2: Regression Coefficients for Standard Neuropsychological and Emotional Variables

Outcome	Predictor	B	SE B	β	*p*	Adjusted R^2
Decision making and judgment problems	Perceived Stress	0.531	0.143	0.485	0.001	0.29
	Longest Digit Backward	-0.483	0.147	-0.43	0.002	
Social competence and teamwork problems	Perceived Stress	0.39	0.093	0.514	<.001	0.352
	Longest Digit Backward	-0.364	0.096	-0.468	<.001	
Impulse control problems	Perceived Stress	0.095	0.025	0.486	<.001	0.292
	Longest Digit Backward	-0.084	0.026	-0.418	0.002	
Supervisor evaluation total	Perceived Stress	2.335	0.614	0.488	<.001	0.292
	Longest Digit Backward	-2.033	0.629	-0.415	0.002	

APPENDIX E
Supervisor Survey*

Study Number: ______________________________

Current Rank: ______________________________

Emotional Control and Stress Tolerance Problems								
Anger	How often does this officer get angry at citizens?	1=Very often	2	3	4	5	6	7=Rarely
Utilizing training under stress	Does this officer "revert to training" when under stress?	1=Very often	2	3	4	5	6	7=Rarely
Excessive force	How often does this officer use excessive physical force?	1=Very often	2	3	4	5	6	7=Rarely
Emotional problems	Does this officer show any evidence of emotional problems?	1=Very often	2	3	4	5	6	7=Rarely
Amount of stress	Does this officer show any evidence of being under stress?	1=Very often	2	3	4	5	6	7=Rarely

Routine Task Performance								
Navigation	Does this officer have problems finding his/her way around town?	1=No problems	2	3	4	5	6	7=Has great difficulty
Drawing crime scene/ accident scenes	Is this officer able to make accurate drawings of crime or accident scenes?	1=No problems	2	3	4	5	6	7=Has great difficulty
Following driving directions	Can this officer follow directions to a location	1=Almost always	2	3	4	5	6	7=Rarely
Radio operation	How often does this officer need to have radio communications repeated?	1=Very often	2	3	4	5	6	7=Rarely
Writing/ paperwork	How well does this officer complete his/her paperwork/ report writing?	1=Very well	2	3	4	5	6	7=Very poorly
Directing traffic	How well does this officer direct traffic?	1=Very well	2	3	4	5	6	7=Very poorly
Marksmanship	Rating of marksmanship	1=Excellent	2	3	4	5	6	7=much improvement needed

Decision Making and Judgment Problems								
Disregarding instructions	How often does this officer "conveniently disregard" instructions?	1=Very often	2	3	4	5	6	7=Rarely
Decision making	Does this officer have difficulties in making a decision?	1=Very often	2	3	4	5	6	7=Rarely
Accidental injury	How likely is this officer to be injured while on the job due to mistakes that he/she makes?	1=Very likely	2	3	4	5	6	7=Not very likely
Making charges later dismissed	How often does this officer make charges that are dismissed?	1=Very often	2	3	4	5	6	7=Rarely
Overlooking violations	How often does this officer overlook violations or fail to enforce code sections?	1=Very often	2	3	4	5	6	7=Rarely
Predicting situational outcomes	How often does this officer correctly predict the outcome of a situation?	1=Very often	2	3	4	5	6	7=Rarely
Poor common sense	How often does this officer use common sense in interpreting the law?	1=Very often	2	3	4	5	6	7=Rarely

Problem solving/ judgment	How good is this officer's problem-solving/ judgment in the execution of his/her duties	1=Very good	2	3	4	5	6	7=Very poor
Suspicious personality traits	Does this officer show any "paranoid" traits? (unwarranted, unhealthy suspicions)	1=No evidence of paranoia	2	3	4	5	6	7=Clear evidence of paranoia
Feedback Acceptance Problems								
Feedback acceptance	How well does this officer accept feedback regarding his/her mistakes	1=Very well	2	3	4	5	6	7=Very poorly

Assertiveness Problems								
Assertiveness	Is this officer more assertive or passive?	1=Assertive	2	3	4	5	6	7=Passive

Social Competence and Teamwork Problems								
Social skills – Officers	Does this officer get along with fellow officers?	1=Almost always	2	3	4	5	6	7=Rarely
Social skills – Public	How good are these officers' social skills with the public?	1=Very good	2	3	4	5	6	7=Very poor
Oral communication	How well does this officer communicate orally?	1=Very well	2	3	4	5	6	7=Very poorly
Complains about instructions	How often does this officer complain about his/her instructions?	1=Very often	2	3	4	5	6	7=Rarely
Social skills – Overall	Rating of social skills	1=Excellent	2	3	4	5	6	7=much improvement needed

Integrity Problems								
Trust in officer	Can you trust/depend on what this officer tells you?	1=Always	2	3	4	5	6	7=Rarely
Integrity/ Ethics	Does this officer perform his/her duties in an ethical manner?	1=Always	2	3	4	5	6	7=Rarely

Conscientiousness and Commitment Problems								
Ambition	How ambitious is this officer?	1=Very ambitious	2	3	4	5	6	7=Not at all ambitious
Attention-to-Detail	How much attention does this officer pay to details?	1=A great deal	2	3	4	5	6	7=Very little
Finishing assignments	How often does this officer complete assignments?	1=Very often	2	3	4	5	6	7=Rarely
Initiative	Is this officer's behavior more often self-initiated or is his/her activity directed by others?	1=Self-initiated	2	3	4	5	6	7=Directed by others
Documentation of activities	How often does this officer fail to properly document his/her activities?	1=Very often	2	3	4	5	6	7=Rarely

Substance Use Problems								
Alcohol/ Drugs	To your knowledge, does this officer abuse alcohol/ drugs?	1=Never	2	3	4	5	6	7=Often

Impulse Control Problems

Act without thinking	Does this officer seem to act without thinking things through?	1=Very often	2	3	4	5	6	7=Rarely

Potential For Inappropriate Aggression					
Aggression	Rating of potential for inappropriate aggression	1=Excellent impulse/anger control	2=Average	3=Needs some improvement	4=I worry about this officer's ability to handle situations appropriately
History of inappropriate aggression	History of inappropriate aggressive interactions with others	1 = No history		2 = History	

Global Rating								
Overall performance rating	Overall performance rating	1=Excellent	2	3	4	5	6	7=much improvement needed

*Adapted from: Tarescavage, A. M., Brewster, J., Corey, D. M., & Ben-Porath, Y. S. (2015). Use of prehire Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) police candidate scores to predict supervisor ratings of posthire performance. *Assessment*, *22*(4), 411-428.

REFERENCES

Army Alpha and Beta tests n. in *A Dictionary of Psychology (3).* Retrieved from: http://www.oxfordreference.com/view/10.1093/oi/authority.20110803095424949

Azen, S. P., Snibbe, H. M., & Montgomery, H. R. (1973). A longitudinal predictive study of success and performance of law enforcement officers. *Journal of Applied Psychology, 57*(2), 190-192.

Bartol, C. R. (1982). Psychological characteristics of small-town police officers. *Journal of Police Science & Administration, 10*(1), 58-63.

Beutler, L. E., Storm, A., Kirkish, P., Scogin, F., & Gaines, J. A. (1985). Parameters in the prediction of police officer performance. *Professional Psychology: Research and Practice*, *16*(2), 324-335.

Biddle Consulting Group. (2013). *Content-related and criterion-related validation study of CritiCall: Knowledge, skill, and ability testing software for dispatchers and calltakers of the Florida Highway Patrol.* Retrieved from http://www.criticall911.com/documents/Florida-Highway-Patrol-CritiCall-Validation-Report_9-3-13.pdf

Bittner, E. (1970). *The functions of the police in modern society.* Chevy Chase, MD: U.S. National Institute of Mental Health, Center for Studies of Crime and Delinquency.

Black, J. (2000). Personality testing and police selection: Utility of the "Big Five." *New Zealand Journal of Psychology, 29*(1), 2-9.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24*(3), 383-405.

Cochrane, R. E., Tett, R. P., & Vandecreek, L. (2003). Psychological testing and selection of police officers: A national survey. *Criminal Justice and Behavior*, *30*(5), 511-537.

Colom, R., Martinez-Molina, A., Shih, P. C., & Santacreu, J. (2010). Intelligence, working memory, and multitasking performance. *Intelligence*, *38,* 543-551.

Costa, P. T., Jr., & McRae, R. R. (1992). NEO-PI-R professional manual. Odessa, FL: Psychological Assessment Resources.

CritiCall Test Descriptions. (n.d.). Retrieved June 20, 2017, from http://criticall911.com/images/cctestdescriptions.pdf

Dantzker, M. L. (2011). Psychological preemployment screening for police candidates: Seeking consistency if not standardization. *Professional Psychology: Research and Practice, 42*(3), 276-283.

Dantzker, M. L. & McCoy, J. H. (2006). Psychological screening of police recruits: A Texas perspective. *Journal of Police and Criminal Psychology*, *21*(1), 23-32.

Detrick, P., Chibnall, J. T., Luebbert, M. C. (2004). The revised NEO Personality Inventory as predictor of police academy performance. *Criminal Justice and Behavior, 31*(6), 676-694.

Drews, F.A., Pasupathi, M., & Strayer, D.L. (2008). Passenger and cell phone conversations in simulated driving. *Journal of Experimental Psychology*, *14*(4), 392-400.

Drews, F. A., Yazdani, H., Godfrey, C. N., Cooper, J. M., & Strayer, D. L. (2009). Text messaging during simulated driving. *Human Factors*, *51*(5), 762-770.

Fyfe, J. J. (1999). Good policing. In Stojkovic, S., Klofas, J. and Kalinich, D. (Eds.), *The Administration and Management of Criminal Justice Organizations*, *3rd edition* (pp. 113-133). Prospect Heights, IL: Waveland Press.

Garrison, T. M., Williams, C. C., & Carruth, D. W. (2012, October). *Sources of cognitive load in a simulated law enforcement patrol task.* Paper presented at the Fourth International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Portsmouth, NH.

Gustafson, B. G. (2015). Then & now: Trends in fatal law enforcement traffic collisions. *Policing: An International Journal of Police Strategies & Management, 38,* 407-424.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42,* 377-381.

Heaton, R. K., Chelune, G. J., & Lehman, R. A. W. (1978). Using neuropsychological and personality tests to assess the likelihood of patient employment. *The Journal of Nervous and Mental Disease, 166*(6), 408-416.

Henson, B., Reyns, B. W., Klahm, C. F., & Frank, J. (2010). Do good recruits make good cops? Problems predicting and measuring academy and street-level success. *Police Quarterly, 13*(1), 5-26.

Higgins, D. M., Peterson, J. B., Pihl, R. O., Lee, A. G. M. (2007). Prefrontal cognitive ability, intelligence, Big Five personality, and the prediction of advance academic and workplace performance. *Journal of Personality and Social Psychology, 93*(2), 298-319.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2003). *Impulsive decision making and working memory. Journal of Experimental Psychology, 29*(2), 298-306.

Hosking, S., Young, K., & Regan, M. (2007). The effects of text messaging on young novice driver performance. In: I. J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J. D.

Irwin (Eds). Distracted driving. Sydney, NSW: Australasian College of Road Safety. Pages 155-187.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340-362.

Hunter, J. E. & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72-98.

Inwald, R. E., Knatz, H. F., & Shusman, E. J. (1983). Inwald Personality Inventory Manual. New York: Hilson Research.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, N.J: Prentice-Hall.

Kilday, K. (2003). Pre-employment testing of computerized telecommunication skills. *Law & Order.* Retrieved from http://www.criticall911.com/images/kilday.pdf

King, R. E., Caretta, T. R., Retzlaff, P., Barto, E., Ree, M. J., & Teachout, M. S. (2013). Standard cognitive psychological tests predict military pilot training outcomes. *Aviation Psychology and Applied Human Factors, 3*(1), 28-38.

Kleider, H. M., Parrott, D. J., & King, T. Z. (2009). Shooting behaviour: How working memory and negative emotionality influence police officer shoot decisions. *Applied Cognitive Psychology*, *24*, 707-717.

König, C.J., Bühner, M., & Mürling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance*, *18*(3), 243-266.

Meier, R. D., Farmer, R. E., & Maxwell, D. (1987). Psychological screening of police candidates: Current perspectives. *Journal of Police Science and Administration, 15*(3), 210-215.

Mills, M. C. & Stratton, J. G. (1982). The MMPI and the prediction of police job performance. *FBI Law Enforcement Bulletin,* 10-15.

Mullins, W. C. & McMains, M. (1995). Predicting patrol officer performance from a psychological assessment battery: A predictive validity study. *Journal of Police and Criminal Psychology, 10*(4), 15-25.

Nieuwenhuys, A., Savelsbergh, G. J. P., & Oudejans, R. R. D. (2012). Shoot or don't shoot? Why police officers are more inclined to shoot when they are anxious. *Emotion, 12*(4), 827-833.

Olea, M. M. & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than *g*. *Journal of Applied Psychology, 79*(6), 845-851.

Ophir, E., Nass, C., & Wagner, A. D. (2009). Cognitive control in media multitaskers. *Proceedings of the National Academy of Sciences, 106*(37), 15583-15587.

Paullin, C., Katz, L., Bruskiewicz, K. T., Houston, J. & Damos, D. (2006). Review of aviator selection. United States Army Research Institute for the Behavioral and Social Sciences, i-C-4.

Public-Safety Dispatcher/Calltaker Pre-Employment Testing. (n.d.). Retrieved June 20, 2017, from http://criticall911.com/

Ree, M. J. & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1*(3), 86-89.

Rogers, R. D. & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Saccuzzo, D. P., Higgins, G., & Lewandowski, D. (1974). Program for psychological assessment of law enforcement officers: Initial evaluation. *Psychological Reports, 35,* 651-654.

Sanders, B. A. (2008). Using personality traits to predict police officer performance. *Policing, 31*(1), 129-147.

Saxe, S. J. & Reiser, M. (1976). A comparison of three police applicant groups using the MMPI. *Journal of Police Science and Administration, 4*(4), 419-425.

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.

Sanbonmatsu, D.M., Strayer, D.L., Medeiros-Ward, N., & Watson, J.M. (2013). Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PLoS ONE, 8*(1), 1-8.

Schoenfeld, L. S., Kobos, J. C. & Phinney, I. R. (1980). Screening police applicants: A study of reliability with the MMPI. *Psychological Reports, 47,* 419-425.

Simmers, K. D., Bowers, T. G., & Ruiz, J. M. (2003). Pre-employment psychological testing of police officers: the MMPI and the IPI as predictors of performance. *International Journal of Police Science & Management, 5*(4), 277-294.

Slotkin, J., Nowinski, C., Hays, R., Beaumont, J., Griffith,J., Magasi, S….Gershon, R. (2012). *NIH Toolbox scoring and interpretation guide.*

Smith, A. (1973). *Symbol Digit Modalities Test.* Los Angeles, CA: Western Psychological Services.

Smith, A. (1991). *Symbol Digit Modalities Test.* Los Angeles, CA: Western Psychological Services.

Stern, R. A. & White, T. (2009). *Neuropsychological Assessment Battery Categories professional manual.* Lutz, FL: Psychological Assessment Resources.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary.* New York: Oxford University Press.

Strayer, D.L. & Drews, F.A. (2004). Profiles in driver distraction: Effects of cell phone conversations on younger and older drivers. *Human Factors, 46*(4), 640-649.

Strayer, D. L., Drews, F. A., & Johnston, W. A. (2003). Cell phone-induced failures of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9*(1), 23-32.

Tarescavage, A. M., Brewster, J., Corey, D. M., & Ben-Porath, Y. S. (2015). Use of prehire Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) police candidate scores to predict supervisor ratings of posthire performance. *Assessment, 22*(4), 411-428.

Tarescavage, A. M., Corey, D. M., & Ben-Porath, Y. S. (2015). Minnesota Multiphasic Personality Inventory-2-Restructured Form (MMPI-2-RF) predictors of police officer problem behavior. *Assessment, 22*(1), 116-132.

Terman L. M. (1917). A trial of mental and pedagogical tests in a civil service examination for policemen and firemen. *Journal of Applied Psychology, 1*(1), 17-29.

Thurstone, L. L. (1922). The intelligence of policemen. *Journal of Personnel Research, 1*, 64-74.

Watson, J. M. & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review, 17*(4), 479-485.

Wechsler, D. (2008). WAIS-IV technical and interpretative manual. San Antonio, TX: Psychological Corporation.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(Suppl3), S54-S64.

Weiss, W. U., Serafino, G., Serafino, A., Willson, W., & Knoll, S. (1998). Use of the MMPI-2 to predict the employment continuation and performance ratings of recently hired police officers. *Journal of Police and Criminal Psychology*, *13*(1), 40-44.

White, M. D. (2008). Identifying good cops early. *Police Quarterly*, *11*(1), 27-49.

Yager, C., Dinakar, S., Sanagaram, M., & Ferris, T. K. (2015). Emergency vehicle operator on-board device distractions. *Texas A&M Transportation Institute*, 1-50.