

Florida Institute of Technology

Scholarship Repository @ Florida Tech

Theses and Dissertations

7-2019

Situational Judgement Selection Test Utility: Implications for Police Turnover

Marshall Alan Jones

Follow this and additional works at: <https://repository.fit.edu/etd>



Part of the [Business Administration, Management, and Operations Commons](#)

Situational Judgement Selection Test Utility: Implications for Police Turnover

by

Marshall Alan Jones

A dissertation proposal submitted to the Bisk College of Business at Florida Institute
of Technology
in partial fulfillment of the requirements for the degree of

Doctor of
Business Administration

Melbourne, Florida
July, 2019

© Copyright 2019 Marshall Alan Jones
All Rights Reserved

The author grants permission to make single copies _____

We the undersigned committee hereby recommend that the attached document be
accepted as fulfilling in part of the requirements for the degree of
Doctor of Business Administration.

“Situational Judgement Selection Test Utility: Implications for Police Turnover”
a dissertation by Marshall Alan Jones.

Theodore Richardson, Ed.D.
Professor and Dean
Nathan M. Bisk College of Business
Dissertation Advisor

Darrel Sandall, Ph.D.
Associate Professor
Nathan M. Bisk College of Business

James Glenn, D.B.A.
Adjunct Faculty
Nathan M. Bisk College of Business

Patrick Converse, Ph.D.
Associate Professor
College of Psychology and Liberal Arts

Abstract

Title: Situational Judgement Selection Test Utility: Implications for Police Turnover

Author: Marshall Alan Jones

Advisor: Theodore Richardson, Ed.D.

Police agencies increasingly find attracting, recruiting, selecting, and retaining police officers challenging. Officer candidates selected can have up to a 25 percent fail rate in basic police training, while probationary officers in field training can fail to complete training with rates as high as 40 percent. Police selection batteries often include situational judgment tests (SJTs) designed to help ensure legal and proper hiring standards. The scores of SJTs, and other selection battery scores, are often filed once the hiring process is complete. This study explored the potential SJT selection battery scores may have in the prediction of field training and turnover outcomes. Results of binary logistical regression model testing suggest that reading scores, likely a proxy for cognitive ability, best predicted field training outcomes as well as overall and involuntary turnover. Additional factors with predictive nexuses to training outcomes and retention include selection battery retest, restraint sub-scores, and prior law enforcement experience. Applied recommendations include pre-field training interventions for new officers with low reading scores, expanded use of realistic job previews, and policy reviews related to selection battery retest.

Keywords: situational judgement tests, turnover, voluntary turnover, involuntary turnover, retention, police selection, selection processes, field training, transfer of training, training performance, field training and evaluation, law enforcement turnover model, person-job fit, person-organization fit, shocks, realistic job previews, reading scores

Table of Contents

Abstract	iii
Table of Contents	v
List of Tables and Figures.....	xvii
Acknowledgement.....	xix
Dedication	xxi
Chapter 1 Introduction	1
Overview	1
Background and Rationale of the Study	6
Statement of the Problem	10
Purpose of the Study.....	12
Questions that Guide the Research.....	12
Definition of Terms	12
Significance of the Study.....	16
Organization of the Remainder of the Study	23
Chapter 2 Literature Review	24
Overview	24
Questions that Guided the Research	25
Overarching Theoretical Framework	26

Attraction-selection-attrition (ASA) model.	26
Person job and person organization fit.	28
Impact to assumptions and boundary conditions for this research.	30
Police Recruitment, Selection, and Turnover.....	31
Practitioner response.....	32
Historical context evolving into today’s challenges.	33
Contemporary challenges for police staffing.....	36
Impact to assumptions and boundary conditions for this research.	41
History of Employment Practices and the Legal Environment.	42
Cognitive ability and adverse impact.....	43
The challenge in using cognitive ability for selection.	44
The rise and fall of personality tests in police selection.	45
Alternatives to cognitive ability and personality.	50
Impact to assumptions and boundary conditions for this research.	52
Situational Judgment Tests.....	52
SJT popularity today.....	53
Candidate perceptions of fairness.	53
Organizational advantages.....	54
Previous SJT studies predicting turnover.	54
Previous police SJT and related studies.....	56
SJT format: video-based vs. written scenarios.	58

SJT criterion (job performance) validity.	60
SJT construct validity debate.	60
Current theory and validity efforts.	62
Impact to assumptions and boundary conditions for this research.	67
Faking.....	68
SJT faking.....	70
Faking countermeasures.	71
Impact to assumptions and boundary conditions for this research.	74
Police Field Training and Evaluation Programs (FTEP).....	75
History of police field training and evaluation programs.	76
Today's FTEP.	77
San Jose alternative models.	79
Transition issues from the basic police academy to field training.	80
Legal issues and failure to train.	81
Impact to assumptions and boundary conditions for this research.	82
Transfer of Training	83
Broad overview of transfer of training literature.	84
Relevant transfer or training literature to the study.	86
On-the-job-training and transfer of training.	86
Framing.....	87
Error response, prevention training, and leader-member-exchange.	88

Behavioral modeling.....	90
Training timing: Pre and post training interventions.....	91
Police academy to field training transfer.....	92
FTEP as a learning transfer system.....	93
Impact to assumptions and boundary conditions for this research.....	94
Turnover Research	95
History of turnover research.....	96
Cost of Turnover.....	103
Organizational Performance.....	105
Impact to assumptions and boundary conditions for the research.....	107
Synthesis.....	108
Current retention problem in policing.....	108
Potential applied contribution.....	110
Potential scholarly contribution.....	111
Theoretical Law Enforcement Turnover Model.....	114
Chapter 3 Methodology.....	117
Overview	117
Research Questions and Hypotheses	118
Research Design	119
Overview of Research Approach Used in the Study	120
Population and Sample.....	120

Appropriateness of population.....	121
Procedures	121
Data Analysis	122
Predictor Measures	123
SJT video Score.	123
Writing score.....	124
Reading score.....	124
Criterion Measures	125
Ethical Considerations.....	125
Researcher Positionality	125
Validity and Trustworthiness	126
Assumptions and Boundary Conditions	128
Assumptions.....	128
Boundary Conditions.	132
Chapter 4 Findings	134
Overview	134
Summary of the Research Problem	134
Purpose of the Study.....	136
Research Questions	137
Data Collection, Data Management, and Analysis Implications.....	138
Predictor Measures	139

SJT video score.....	139
Critical thinking.....	139
Investigative.....	140
Team orientation.....	140
Confrontation.....	140
Restraint.....	140
Ethical orientation.....	140
Empathy.....	140
Organization orientation.....	140
Community relations.....	141
SJT sub-score dichotomous recode.....	141
Writing score.....	141
Writing dichotomous recode.....	141
Reading score.....	142
Reading dichotomous recode.....	142
Additional predictor variables.....	143
Dependent Variables	144
Field training and evaluation program (FTEP) passing.....	144
Field training and evaluation program (FTEP) failure.....	144
Involuntary turnover.....	145
Voluntary turnover.....	145

Overall turnover.....	145
Need for remedial training.....	146
Failing remedial training.....	146
Methods and Techniques	146
SPSS analysis and relevant processes.....	148
One-tailed versus two-tailed hypothesis tests.	148
Hosmer and Lemeshow goodness of fit test criticisms.	149
Direct of the relationship (B values).....	150
Manual corrections in dichotomous independent variable odds ratios.....	150
Research Findings	151
Hypothesis 1: SJT Video and Reading Scores Predicting FTEP Passing	151
Hypothesis 1a: Lower Critical Thinking Sub-scores Predicting FTEP Failure .	153
Hypothesis 1b: Lower Reading Scores Predicting FTEP failure.	154
Hypothesis 2: Selection Scores and Sub-scores Scores Predicting Overall	
Turnover	155
Hypothesis 2a: Lower Reading and Critical Thinking Sub-scores Predicting	
Involuntary Turnover	157
Hypothesis 2b: Writing and Select Sub-scores Predicting Voluntary Turnover	159
Supplemental Analyses	160
Reading score less than 90 (Read<90).....	160
Candidate retest (Retest).	161

Candidate retest fail then pass (Retest_FP).	161
Prior military experience (Prior Military).....	161
Prior law enforcement experience (Prior LE).....	161
Writing score 100 percent (Writing 100).....	161
Restraint sub-score (Restraint).....	162
Ethical orientation sub-score (Ethical Orientation).	162
Supplemental Analysis 1: Read<90 Predicting FTEP Performance Failure	162
Supplemental Analysis 2: READ<90 and Retest FP Predicting FTEP Failure .	164
Supplemental Analysis 3: Predictive Model of Overall Turnover.	166
Supplemental Analysis 4: READ<90 and Repeat Predicting Involuntary Turnover	168
Supplemental Analysis 5: Predictive Model of Officers Requiring Remedial Training	170
Supplemental Analysis 6: Predicting Officers Failing Remedial Training.....	173
Synthesis and Summary of Data	175
Specific findings.	176
Context relevant limitations related to findings interpretations.....	176
Field training performance.....	176
Turnover.....	178
Remedial training.....	181
Findings Informing the Problem Statement.	184

Contribution to applied practice.	184
Applied practice contributions.	184
Chapter Summary.	187
Chapter 5 Discussion, Implications, and Recommendations	188
Overview	188
Applied research problem.	188
SJT selection batteries.	190
Study goals toward addressing the problem.	190
Major literature themes.	191
Attraction-selection-attrition model.	191
Job fit.	192
Person-organization fit.	193
Situational judgement tests	194
Transfer of training	195
Turnover.	196
Voluntary versus involuntary turnover.	196
Functional, dysfunctional, and optimal turnover.	197
Unfolding model of turnover (shocks).	198
Law Enforcement Turnover Model.	200
Methods and findings summary.	203
Contributions of the Study	206

Expected contributions.	206
Prediction versus prevention of turnover.....	206
Study contributions.	206
Contributes to limited turnover prediction literature.	207
Temporal order and KSAOs with turnover.....	207
Public sector research.	208
Police training effectiveness.	209
SJTs predicting performance and turnover.	209
Selection battery use toward improved outcomes.	210
Police transfer of training and retention strategies	211
Findings related to the research questions.	212
Findings and theoretical support for the study framework.	213
Selection SJT predicting performance and turnover.....	213
Transfer of training	217
P-O fit and ASA model.....	219
Functional turnover.....	220
Dysfunctional turnover and shocks.....	220
Functional turnover or retention “either” area.....	221
Optimal turnover.....	222
Findings that differ from previously reviewed studies.	223
Difference with study findings	223

Discussions and Implications	224
Key implications informing literature.....	224
Limitations.	224
Methodology.....	225
Sample cohort and size.	225
Source of SJT selection battery	227
Unfolding model.	227
What subconstructs do SJTs tap?.....	228
Boundary Conditions.	229
Sector, industry, and job domain	229
SJT and selection battery	229
Field training and evaluation program.....	230
Turnover Constructs	230
Law Enforcement Turnover Model	230
Recommendations	230
Recommendations for applied utility.....	231
Recommendation for pre-FTEP reading related training	232
Recommendation for improved realistic job preview strategies	233
Recommendation to review selection battery policies on retest.....	235
Recommendations for Future Research.	236
Replication studies.....	236

SJT selection batteries predicting and superior performance	238
Conclusion.....	238
Practitioner and academic partnerships	239
References	241
Appendix	306

List of Tables and Figures

Figure 1. Lievens and Motowidlo determinants and antecedents of SJTs.....	63
Figure 2. The Theoretical Law Enforcement Turnover Model.....	116
Table 1: Logistic Regression: SJT Video and Reading Scores Predicting Training Pass (H1).....	152
Table 2: Logistic Regression: Critical Thinking Predicting Training Performance (H1a).....	153
Table 3: Logistic Regression: Reading Score Predicting Training Performance (H1b).....	155
Table 4: Logistic Regression: Predictors Officer Turnover(H2).....	156
Table 4a: Sub-score Distributions and Range Restrictions H2.....	157
Table 5: Logistic Regression: Reading and Critical Thinking Predicting Involuntary Turnover (H2a).....	158
Table 6: Logistic Regression: Predictors of Voluntary Turnover (H2b).....	159
Table 6a: Sub-score Distributions and Range Restrictions H2b.....	160
Table 7: Logistic Regression: READ<90 Predicting Training Performance (S1)	164
Table 8: Logistic Regression: READ<90 and RETEST_FP Predicting Training Failure (S2).....	166
Table 9: Logistic Regression: Predictors Officer Turnover (S3).....	167
Table 10: Logistic Regression: READ<90 and Retest Predicting Involuntary Turnover (S4).....	170

Table 11: Logistic Regression: Predictors of Remedial Training (S5).....	173
Table 12: Logistic Regression: Predictors of Remedial Training Failure (S6).....	175
Figure 3. The Law Enforcement Turnover Model.....	203

Acknowledgement

I am grateful to so many people who helped me along my path to complete this study. The practitioners, unnamed for privacy, who believed in the potential applied value of this study and who mined the data and information required for analysis. The student researchers who spent countless hours entering daily performance information necessary to actualize this research. My colleagues in the School of Psychology for their open support and encouragement as well the quiet, less obvious, support shown when they picked up administrative tasks to allow me to focus on this research. My DBA faculty and cohort fellows who provided conceptual counterpoints and challenging questions and opinions necessary to scholarly inquiry.

This study would not be possible without my dissertation committee. A special thanks to Dr. Ted Richardson, my dissertation chair and mentor, for encouraging me to “go big or go home” in my study approach. Dr. Patrick Converse, thank you for your constant patience and making time for my methodological questions and to review my analyses. Dr. Darrel Sandall, I appreciate your attention to detail in reviewing my manuscripts and offering alternative methodological perspectives. Dr. James Glenn, thank you for your interest in this topic and thoughtful, but challenging, questions about the research.

I extend my gratitude to the current and past Dean of Psychology. Dean Lisa Steelman, thank you for your support and council for 20 years toward reaching

my goal. Your encouragement and support during the completion of this study was critical in allowing me to focus on the task at hand. To my former dean, Dr. Mary Beth Kenkel, I appreciate your patience during the two decades it took to me to finish what I started when you gave me my opportunity in academia.

I am grateful for the administration and staff of the DBA program for their constant support. To Dr. Emily Vogt, thank you for your openness to ideas and observations related to the program and for insuring that we are prepared to actualize research. Thanks to Taylor Weaver, who constantly went above and beyond to help student in the program be successful. I want to give a special thank you to a former student and friend, Nicole Murray, for her amazing attention-to-detail and patience in copy editing.

Being a late bloomer, I have reflected over the years on how a mediocre student in high school excels in college and, eventually, earns a doctorate. While many teachers, faculty, friends, and colleagues have influenced me, one high school teacher stands out. I want to thank Shelia Byers, my high school math teacher, for her graduation eve lambaste that would have made my Parris Island drill instructors proud. Shelia, thank you for caring enough to share your disappointment that I did not live up to my potential. Your words, “I looked for your name among those graduating with honors to find you lazily not even close” have resonated with me since June 1985. While you may not remember, I sure do. Thank you from the bottom of my heart Shelia.

Dedication

This dissertation is dedicated to my family. To my late father, Arnold, who despite having only attained a sixth-grade education, realized the importance of education and ensured that his children realized opportunities he could not. To my late mother, Charlotte, who quietly worried when I joined the Marines and the police force, and who became my most pragmatic cheerleader during the 20-year journey toward a terminal degree. To my late brother Bill and late Uncle Melvin who poked and prodded just enough over the years to irritate me to forge ahead. I knew full well they were among my biggest supporters, but they knew I needed poking and prodding.

To my Uncle Bobby, whose pragmatic counsel I cherish, for his sincere interest in my research and enthusiasm for my potential to become the first in the family to earn a terminal degree. To my brother Lew and sister Dawn who always made sure I knew the family had my back as I, at times, labored to balance family, work, and school. To all my nephews, nieces, and cousins, especially Niles, who always remembered to ask how the journey was progressing and never doubted the outcome, even when I did.

To my kids, I hope you each follow your heart's path and never quit. I hope I offer you the same unwavering support you have shown me during this process. To my Jenny, who made sure I was allowed the time and ensured that I had the quiet necessary during the weeks of research and writing this study. Your

unwavering understanding, encouragement, and support means more than you can ever know.

Chapter 1

Introduction

Overview

Selecting and retaining employees is well recognized as important to any contemporary organization (Mabon, 1994; Ployhart, 2006; Ployhart & Weekley, 2010; Pulakos & Schmitt, 1996; Rubenstein, Eberly, & Lee, 2015) and can even be more so for governmental organizations (Barrett, Doverspike, & Young, 2010; Selden, Schimmoeller, & Reese, 2013; The Center for State and Local Government Excellence, 2017; Wilson, 2014). There is ample research in the literature, and employment law, on the critical aspects of proper selection criteria and methods (Arnold, 2004; Campbell, McCloy, Oppler, & Sager, 1993; Gutman, 2009; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Pulakos & Schmitt, 1996; Schmidt & Hunter, 1998; Van Iddekinge, Roth, Putka, & Lanivich, 2011). Selection is a cornerstone of human resource management where the focus must balance identifying the most qualified candidates that will produce the best person-organization (P-O) fit (Kristof, 1996) between the employees and the organization in an effort to attract, select, and retain employees (Rubenstein et al., 2015; Schneider, 1987; Selden et al., 2013) while avoiding issues related to adverse impact (Barrick & Zimmerman, 2005; Gottfredson, 1996; Gutman, 2009; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Weekley & Jones, 1999).

An identifiable nexus between selection and turnover is often overlooked

(Hom, Lee, Shaw, & Hausknecht, 2017) and can be attributed to a disconnect between human resources and operations units where training and work is conducted. Human resources must focus on the standards and guidelines of hiring, while operations focus on performance and fit (Ployhart, 2006). This divide can be amplified in organizations where the quest for a diverse force has been a challenge (Barrick & Mount, 2005; Fyfe & Kane, 2006; Ployhart, 2006). Retaining employees is also important, especially in organizations where training time and costs are a significant investment (Hausknecht & Holwerda, 2013; Mabon, 1994; Ployhart & Weekley, 2010; Rubenstein et al., 2015; Seldon et al., 2013).

The cost of turnover is also high in organizations where the financial and operational consequences go beyond simply the cost of selection, hiring, and training (Barrick & Zimmerman, 2009; Hom et al., 2017; Maltarich, Nyberg, & Reilly, 2010; Van Iddekinge et al., 2011), such as first responders in the public sector (Barrett et al., 2010). Organizations must consider the cost of onboarding and training new employees. The average employee, regardless of industry or sector, has a learning curve before the new employee fully contributes to the organization. This learning curve takes even longer in complex roles (Ployhart & Weekley, 2010), such as policing, which serves as the population for this research. Some scholars suggest organizations could screen out candidates with a high potential of leaving with proper data (Hom et al., 2017; Price, 2011).

The U.S recession that began in 2008 led to widespread challenges in

sustaining necessary sworn workforces in policing (Police Executive Research Forum, 2013; Wilson, 2014; Woska, 2016). Woska (2106) revisiting his 2006 review on police recruiting for the International Association of Chiefs of Police (IACP). Woska's 2016 update, focused on police recruiting changes since his 2006 review, cites that the 2008 recession critically impacted local government budgets resulting in unprecedented police layoffs, and ultimately police services themselves. Wilson (2014) reports that budget related reductions, including layoffs, department consolidation, and even agency disbanding further compounded efforts to maintain a proper police workforce. Woska (2016) points to the budget crisis as a catalyst for dramatically decreased average officer to population ratios nationwide. According the U.S. Department of Justice (2011), officer to population ratios fell to a national average of 181 officers per 100,000 residents in 2011 compared to nearly 250 officers per 100,000 reported in the years approaching 2008.

Smith (2016) and Wilson (2014) both argue that post-recession economic conditions have not eliminated the recruiting problem from the turn of the 21st century. They point to the greater competition from other occupations, generational views on policing, and candidate disqualification issues, such as drug use and physical fitness, as having further impacted the volume of qualified applicants.

Woska (2016) and Smith (2016) point to an additional impact to the attraction to policing by candidates resulting from budget-driven reductions in pay and benefits, particularly retirement benefits, and social reactions to policing. The reality for today's

police executives is to necessity to compete for applicants where other sectors and industries have openings and offer more attractive pay, benefits, and social acceptability far exceeding those of policing (Smith, 2016, Wilson, 2014; Woska, 2016). This notion is supported by compounding issues in the rapidly evolving economic and social environment that require innovative solutions (Jordan, Fridell, Faggiani, & Kubu, 2009; Orrick, 2008).

Applied research on the nexus between selection and turnover is limited (Hom, Lee, Shaw, & Hausknecht, 2017). Turnover must be better explored and understood in policing. Most turnover research focuses on *prevention* of turnover (Hom, Lee, Shaw, & Hausknecht, 2017; Rubenstein, Eberly, Lee, & Mitchell, 2018).

Prediction of turnover is rare to turnover literature aside from Barrick and Zimmerman (2009) and Van Iddekinge et al. (2011). Predictive models of *voluntary turnover* (Mobley, Griffeth, Hand, & Meglino, 1979) or *involuntary turnover* (McElroy, Morrow, & Rude, 2001; Rubenstein et al., 2017) are important and offer potentially critical insight to policing at a time when it is much needed.

An example of the critical difference of voluntary from involuntary turnover comes from the Center for State and Local Government Excellence (CSLGE). CSLGE (2016) reports that hiring and turnover were issues during the recession, starting in 2008, on state and local governments which reported 26 percent more employees were laid off or discharged (unintentional turnover) than quit (intentional turnover). The situation has changed in 2016, where 60 percent more employees quit (intentional

turnover) than were laid off or discharged (unintentional turnover). Exploring predictive utility of turnover, and distinguishing voluntary from involuntary, has the potential to offer gateways into effectiveness of police training (National Research Council, 2004).

It is imperative for police organizations to successfully recruit, select, and train new officers based on data driven decisions. The increasingly complex economic, social, and political environments that police practitioners face while safeguarding our communities has life and death consequences and unprecedented scrutiny. Diminished applicant pools, generational attributions, more attractive employment options in other sectors and industries, and increased competition among police organizations make attracting, selecting, and retaining qualified candidates with the potential to become skilled police officers more critical to police organizations than ever before. The fiscal constraints, particularly among small and mid-size organizations, require innovative strategies and initiatives that leverage existing resources, potentially from existing selection process data, to improve retention outcomes of training new officers and to prevent dysfunctional turnover through poor selection. Failures to prevent turnover among new officers exasperates the critical selection, hiring, and training dilemma facing police organizations as each time they must replace a new officer they revisit an increasingly diminished applicant pool where competition, both from other police organizations and other industries, is harsh.

The focus of this research was to explore existing police selection

processes, which include situational judgment tests (SJTs), for potential untapped utility toward predicting candidate attributes of both voluntary (poor selection possibilities) and involuntary (poor on the job training performance) turnover. The applied goal of the research was to provide new predictive insights for decision makers, thereby allowing for improved selection decisions towards decreased new officer turnover.

Background and Rationale of the Study

Person-organization (P-O) fit refers to the compatibility between organizational members, or potential members, and the organization (Kristof, 1996). P-O fit aligns with Schneider's 1987 *attraction-selection-attrition* (ASA) Theoretical Model that posited that turnover, through attrition, can occur when organization members' values, interests, beliefs, or other characteristics fail to align to the job characteristics and social environments of the organization. Ployhart (2006) explored staffing issues for all organizations in the 21st century. He argued that current staffing issues present more complex issues for organizations, noting that challenges, such as increased knowledge, work place diversity, and competition for applicants, requires more targeted recruiting and training initiatives.

Ployhart and Weekley (2010) recognized the changing nature of work and pointed out that trends in longer tenure are shifting given Millennial generational norms toward shorter lengths of employment. Organizations must adapt and maintain critical key concepts of P-O fit and ASA to effectively recruit, select, train, and retain

skilled and valued members. Rubenstein, Eberly, and Lee's (2015) meta-analysis of voluntary turnover discussed the idea of the homophily principle (McPherson, Smith-Lovin, & Cook, 2001) within the notion of P-O fit and an individual's assessment of "fit" socially with the other members of the organization. Turnover can lead to an organizational crisis when the mission cannot be carried out due to workforce shortages or inability.

Mabon (1994) explored turnover and identified this type of organizational dysfunction in selection and retention resulting in heavy cost and loss of performance. Mabon (1994) also explored the costs associated with frequent turnover, from the perspective of employee tenure, and pointed to the seemingly obvious organizational aspect of less turnover leads to less cost related to recruiting, selecting, and training new employees. Maltarich, Nyberg, and Reilly (2010) acknowledged that a great deal of literature on turnover is focused on individual performance but assert that turnover, especially voluntary turnover, is important at organization and team-level performance. Supporting this notion are Hausknecht and Holwerda (2013) who suggested that organizational turnover damages organizational level performance more when remaining members are more novice.

Organizations must consider the cost of onboarding and training new employees. Ployhart and Weekley (2010) discussed the inevitable learning curve, especially in complex roles, necessary before a new employee can fully contribute to the organization. This learning curve takes even longer in complex roles, such as

policing (Ployhart & Weekley, 2010). Pulakos and Schmitt (1996) acknowledged the increased cost in efforts to reduce issues related to adverse impact. Paper and pencil tests are the most affordable with more comprehensive techniques such as high fidelity SJTs being costly.

Multiple studies report results supporting that SJTs have face validity benefits over other selection options (Bauer, Maertz, Dolen, & Campion, 1998; Kluger & Rothstein, 1993; Ployhart & Ryan, 1998; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993) as well as potential for improved candidate reactions to testing (Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; Clevenger, Pereira, Wiechmann, Schmitt, & Schmidt Harvey, 2001).

The quest for the identification of prediction criteria in selection research is established in the literature (Van Iddekinge, Roth, Putka, & Lanivich, 2011). The ability of personnel selection techniques to capitalize on predictor constructs, such as cognitive ability, personality, knowledge, and skills, (Campbell et al., 1993; Schmidt, Hunter, & Outerbridge, 1986) has been studied and assessed to include SJTs in studies such as Schmidt and Hunter's 1998 review of selection methods.

Hogg and Wilson (1995) explored trends in policing at the time and report on a variety of task assessments to evaluate ability, such as assessment centers including job simulations and in-basket exercises. They contend that these high-fidelity exercises are better at predicting performance but the cost, compared to standard cognition tests, can be prohibitive.

The quest for law enforcement agencies to attract, select, and retain the best possible candidates is critical (Cavanaugh, 2013; Haarr, 2001; Macaig, 2018; Smith, 2016; Wilson, 2014; Wolfe & Nix, 2016; Woska, 2016). This can be a real challenge for smaller and mid-sized agencies where salary and benefits are not comparable to larger agencies (Koper, Maguire, & Moore, 2001; Macaig, 2018; Orrick, 2008; Reaves, 2012). Selecting, training, and retaining police officers is a challenge in itself, but this issue can be amplified when it comes to minority recruitment given the resources (Campbell, Christman, & Feigelson, 2000; Jordan, Fridell, Faggiani, & Kubu, 2009; Raganella & White, 2004; Reaves, 2012), such as bonuses and relocation expense reimbursement, that small and mid-sized agencies cannot afford (Campbell et al., 2000; Jordan et al., 2009; Macaig, 2018; Raganella & White, 2004; Reaves, 2012).

Recruiters from smaller agencies must contend with significant pay and benefit differences as compared to the large agencies as well as department budget shortages that equate directly to retention issues (Macaig, 2018; PERF, 2013; Raganella & White, 2004). Recruiting and selecting police candidates is challenging and expensive. It is critical that agencies have a valid and reliable selection process to help ensure and demonstrate that an agency is compliant with employment law and *doing things right* as well as to minimize the frequency of attempting to hire from the diminished pool of applicants. If selection battery data can also help inform retention strategies, then the cost of selection tests may offer added and valuable utility for organizations.

Statement of the Problem

Police organizations today face a long list of increasing dynamic economic, social, and workforce challenges in their quest to recruit, select, and retain a qualified and skilled police force (Cavanaugh, 2013; Haarr, 2001; Jordan et al., 2009; Koper et al., 2001; Maciag, 2018; National Research Council, 2004; Orrick, 2008; PERF, 2013; Selden et al., 2013; CSLGE, 2017; Smith, 2016; Wilson, 2012; Wilson, 2014; Wolfe & Nix, 2016; Woska, 2006; Woska, 2016). These challenges can result in dysfunctional turnover among new officers and exasperate police organizations' ability to effectively select and retain quality applicants. One challenge faced by policing is a dramatically diminishing applicant pool where attracting good candidates, from other industries and among police organizations, is increasingly challenging (Barrett et al., 2010; Cavanagh, 2003; Haarr, 2001; Koper, Maguire, & Moore, 2001; Orrick, 2008; Selden et al., 2013; Smith, 2016; Wilson, 2012; Woska, 2016).

Chan and Schmitt (2002) and Weekley and Ployhart (2005) suggested that SJTs have higher selection validity over the combined validity of cognition, personality, and previous work experience. Whetzel and McDaniel (2009) pointed to the growing popularity in the use of SJTs in the U.S. and Europe. Empirical support for this use of SJTs includes better predictive validity predicting job performance relative to other selection options, such as cognitive ability and personality, while producing less adverse impact (Clevenger, Pereira, Wiechmann,

Schmidt, & Harvey, 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Weekley & Jones, 1999). Lievens and Motowidlo (2016) provided evidence supporting situational judgment tests (SJTs) as predictors of job performance by tapping specific job relevant content as well as general domain knowledge.

Hom, Lee, Shaw, and Hausknecht (2017) reviewed a century of turnover research and suggest that applied research linking selection processes to turnover is severely limited. Rubenstein et al.'s (2015) suggestions for future turnover research included exploring actual turnover, compared to turnover intentions, and to dichotomize turnover as *voluntary* or *involuntary* turnover.

A research gap exists on exploring the potential utility of SJT, and other selection tools, toward field training and turnover outcomes in high consequence professions, such as policing, since these selection data often never leave human resources. There may exist valuable utility in SJTs, and other selection assessments, if these data are used predictively to better inform decisions toward selection, potentially impacting intentional turnover (quitting), and field training strategies, potentially impacting unintentional turnover (termination due to poor field training performance) directly related to retention. The study explored the potential utility to inform potential applied strategies for improved selection and field training performance outcomes and less turnover, both intentional and unintentional.

Purpose of the Study

This research explored the existing gap in literature regarding SJT's potential predictive nexus with turnover (Ployhart, 2006; Ryan & Ployhart, 2013; Price, 2011). This study sought to better understand how SJTs, along with other selection instrument scores, may impact actual voluntary, involuntary, and overall turnover. This study has potential applied utility regarding predictive pathways toward potential development of proactive measures, treatments, or initiatives to increase officer retention. The following overarching research questions will be explored with several hypotheses.

Questions that Guide the Research

This study explored the potential predictive nexus between SJTs, along with other selection instrument scores, to officer voluntary (choice), involuntary (performance), and overall (collective) turnover. The following overarching research questions were explored with several hypotheses.

RQ1. Do Situational Judgement Tests (SJTs) offer predictive utility toward field training performance.

RQ2. Do SJTs offer predictive utility for turnover outcomes.

Definition of Terms

Adverse Impact: occurs when selection rates in employment decisions may indicate disadvantage to members based on race, gender, or ethnic group (EEOC, 1978).

Attraction-Selection-Attrition (ASA) Model: explains the selection cycle where members are attracted, selected, and eventually leave organizations (Schnieder, 1987).

Daily Observation Report (DOR): is a “standardized form completed by the FTO which reflects his or her observation of a trainee’s performance in areas determined to be critical in an agency’s law enforcement duty” (FDLE, 2011, p. 2).

Disparate Treatment: occurs when candidates are denied employment opportunities based on membership in a class based on race, gender, or ethnic origin (EEOC, 1978).

Dominant Analytical Mindset (DAM): is the result of overgeneralizing research findings due to a lack of clearly defined contextual and operational definitions (Allen, Hancock, Vardaman, & McKee, 2014).

Dysfunctional Turnover: is a condition of malfunction resulting from the loss of a valued employee detracting from organizational effectiveness (Abelson & Baysinger, 1984).

Faking: is defined as applicant’s conscious distortion of responses on employment tests with the intention of scoring most favorably and to increase their perceived chance of selection (McFarland & Ryan, 2000).

Fidelity: refers to “the extent to which the assessment task and context mirror those actually present on the job” (Lievens & Sackett, 2012, p. 462).

Field Training and Evaluation Program (FTEP): also referred to as *field training program*, encompasses the entire 14-16 week field training experience including the written directives, procedures, and activities related to training newly hired police officers (FDLE, 2011).

Field Training Officer (FTO): A non-probationary officer who performs normally assigned patrol functions plus the added duty of trainer, role model, supervisor, counselor, and evaluator when assigned to a new officer (FDLE, 2011).

Ferguson Effect: is a concept that anti-cop rhetoric fueled by media attention that is based on incomplete information leads to increased, often unfair, politicization of events (Wolfe & Nix, 2016).

Goal Setting Theory: asserts “task performance is regulated directly by the conscious goals that individuals are trying for on a task” (Locke & Latham, 1990, p. 240).

High Stakes Professions: for the context of this research are public sector positions, such as police, fire, and medical personnel, where consequences for proper selection are high (Barrett, Doverspike, & Young, 2010).

High Stakes Testing: are testing occurrences where one’s score determines admittance (i.e. college), selection for a high stakes job (i.e. police, military, government), or other significant high stakes impacts (Lievens, Buyse, & Sackett, 2006).

Involuntary Turnover: considered as avoidable and a function of poor individual performance or insubordination resulting in dismissal (McElroy, Morrow, & Rude, 2001).

Knowledge, Skill, Ability, or Other Characteristics (KSAO): refers to the knowledge, skills, abilities, and other characteristics necessary for proper workplace performance (Bernardin & Russell, 2013).

Optimal Turnover: a level of turnover with positive organization consequences, such as reduced labor costs, better P-O fit combinations, or infusion of new diverse ideas (Abelson & Baysinger, 1984).

Person-Job (P-J) Fit: is defined as “congruence, match, or similarity between the person’s...abilities and demands of the environment.” (Edwards, 2008, p. 168).

Person-Organization (P-O) Fit: in a general sense, is defined as *compatibility* (supplementary, complementary, needs-supplies, or demands-abilities) between the person and the organization (Piasentin & Chapman, 2006).

Poaching: considered a shock and defines interorganizational competition for experienced officers due to the tight labor market (Agarwal, Ganco, & Ziedonis, 2009; Gardner, 2005; Home et al., 2017).

Policing: is the term operationalized in this research to define local, county, and state law enforcement in the United States.

Remedial Training: is “a plan of action developed to assist a trainee in achieving

acceptable performance level in areas of weakness” (FDLE, 2011, p.5).

Retention: is the opposite of turnover.

Standardized Evaluation Guidelines (SEG): “is an acknowledged measurement system used to compare the trainee’s performance as a basis for assigning ratings” (FDLE, 2011, p. 2).

Self-Efficacy: refers to "how well one can execute courses of action required to deal with prospective situations" (Bandura, 1982, p. 122).

Shocks: are related to turnover where external, to the organization, events can *push* or *pull* employees to leave (Burton, Holtom, Sablinski, Mitchell, & Lee, 2010).

Situational Judgment Test (SJT): tests designed to assess an applicant's judgment regarding a situation encountered in the work place (Weekley & Ployhart, 2006).

Transfer of Training: is the generalization, retention, or enhancement of job relevant KSAOs acquired as a function of training (Baldwin & Ford, 1998).

Qualified Applicants: Applicants possessing the requisite KSAOs for the essential duties of a particular job (Gutman, 2000).

Voluntary Turnover: employee self-initiated separation from the organization (Mobley, Griffeth, Hand, & Meglino, 1979).

Significance of the Study

Ployhart (2006) pondered the neglected questions in recruiting. He further

pondered why managers fail to heed the results of research and science, wondering if the results are too convoluted or vague for application on organizations. How do we get decision makers in organizations to apply research?

This research added to the existing literature with potential applied utility for police managers. First, the majority of turnover research focuses on prevention of turnover opposed to prediction and applied research on turnover is very limited (Hom et al., 2017). The nexus between SJTs and tenure is virtually unexplored apart from a 1994 study looking at potential post-selection SJT prediction of insurance agent turnover by Dalessio (2014). Barrick and Zimmerman (2009) and Whetzel and McDaniel (2009) pointed to the void of research between SJTs and turnover and potential applied utility. This research was an opportunity to actualize the Rubenstein et al. (2018) suggestion that future turnover research included exploring actual voluntary turnover as opposed to *turnover intentions* as well as distinguishing *voluntary* from *involuntary* turnover. This aligns with the quest to identify prediction criteria in selection research (Van Iddekinge et al., 2011) while addressing Hom et al.'s (2017) identified need for applied turnover research.

Second, this research added to the literature exploring the ability of personal selection tools to capitalize on predictor constructs, such as cognitive ability, personality, knowledge and skills, (Campbell et al., 1993; Schmidt et al., 1986) that have been studied and assessed to include SJTs (Schmidt & Hunter, 1998). This study added to the limited existing turnover prediction from Barrick et al. (2009) and Van

Iddekinge et al. (2011).

Third, employing actual daily performance evaluations (DORs) offers measurement opportunities discussed in the literature as lacking. Rubenstein et al. (2018) called for research to explore the temporal order of turnover as opposed to a predictor at one point in time. Utilizing DORs offered an opportunity to increase the frequency of measurement points temporally as suggested by Rubenstein et al. (2017). This nexus between knowledge, skills, abilities, and other characteristics (KSAOs) of policing and measuring actual field performance has been a critical missing link in exploring police officer performance (National Research Council, 2004). Rubenstein et al. (2015) also suggested measuring specific events as part of potential future predictive analysis of turnover. Exploring the performance measures of the DORs offered potential in filling this gap in the literature.

Fourth, Selden et al. (2013) examined new employee turnover in state governments exploring turnover and high-performance work systems (HPWS) framework. Their findings suggested that three aspects of HR practices are associated with predicting new hire turnover. The areas were recruiting and selection, compensation, and training and development. Ployhart (2006) directly queried if government sector selection research, commonly based on job analysis, is generalizable to the private sector where the pace of change and need for innovation prevail. He noted that the practice is to assume that utility exists without exploration. The research explored government employee turnover while exploring

selection and training outcomes associated with new hire turnover potentially adding to Selden et al.'s (2013) work specifically with public sector new hires. It also paved a pathway for replication with private sector organizations that deploy similar SJT and selection instrumentation to evaluate the utility question posed by Ployhart in 2006.

This unique population provided an opportunity to discuss the nexus between business and management to the government sector as it relates to the research. The Association to Advance Collegiate Schools of Business (AACSB) defines organizations in The Business School Conceptual Framework as business, non-profit, and government (AACSB, 2010). The interdisciplinary activities promoted by AACSB (2010) are geared to the outcomes including improving organizational performance by means of applied research with a goal of creating intellectual capital. This research project was an applied research endeavor with strong potential service utility, not only toward the government organizations, specifically the law enforcement profession, but for organizations in general. The Academy of Management (AOM) is a leading professional association for management and organizational scholars (AOM, 2018). The Public and Nonprofit Division of the Academy of Management (PNP) is one of the 25 management disciplines of specific scholarly interest with focus on aspects of public organizations such as human resource management and decision making (PNP, 2018).

Fifth, The National Research Council (2004) reported a significant unresolved question regarding effectiveness of police training and a lack or rigorous research necessary to help shape training, especially with problem-solving. They contend that until research yields better details on training, beyond simple correlational designs, little progress is likely. For example, Geller and Toch (1995) reviewed model policies and best practices regarding training and officer use of force. They concluded that despite literature recommending specifics of such training, there were no empirical validations of the claims. This research offered predictive analysis (logistic regression) of selection data on police performance during field training by use of DORs. This study offered insight into the National Research Council's (2004) assertion that research exploring the nexus of KSAOs would be valuable to practitioners and police managers by helping inform potential improvements to police training. Caro (2011) specifically discussed the advantage of exploring police performance during field training, contending that it is a substantial method of measuring performance that is lost after field training. The agency explored utilizes the San Jose model, thereby providing a systematic system and solid framework to explore performance related constructs.

Sixth, Ryan and Ployhart (2014) discussed the continuing evolution of SJT research. They contended that future research needs to include SJT predicting performance in unexplored domains including turnover. Ployhart and Weekley (2006) offered suggestions for future science and practice in SJT utility. They

acknowledged that most SJT research are field studies but have been limited to correlational design. They suggested more experimental approaches to explore SJTs measure related to decision making and performance. This research deployed predictive analysis, well beyond correlational designs, to specifically address SJT prediction on turnover outcomes thereby addressing gaps in the literature as identified by Ployhart and Weekley (2006), Ryan and Ployhart (2014) and The National Research Council (2004).

Seventh, Ployhart (2006) noted that every organization uses some form of selection process without guarantee of effective, appropriate, or optimal use. He further contended that unless research gaps are addressed this trend will continue. Hausknecht and Holwerda (2013) suggested that personnel tracking approaches with turnover details may yield opportunities for organizations to design strategic interventions. This research attempted to specifically explore selection processes for additional utility toward optimal use suggested by Ployhart (2006) while searching for potential foundational predictive research useful for innovative interventions as encouraged by Hausknecht and Holwerda (2013). Hom et al. (2017) and Price (2011) also suggested that data-driven initiatives may result in organizations gaining an ability to screen out those with a high chance of leaving. Potential applied utility may inform both improved transfer of training toward better retention while providing a glimpse into how valid and reliable data can help organizations select out likely leavers.

Eighth, Baldwin, Ford, and Blume (2009) discussed the importance of transfer of training research to practitioners charged to train newly selected members and organizations depending on the trainings impact on job performance. Baldwin et al. (2009) contended that evidence-based interventions from literature have the potential to inform application. They also call for future research to explore potential improvement for training designs, trainee readiness, and supervisor support related to training transfer in applied domains. This echoes the call from Hom et al. (2017) urging empirical observation of management strategies and practices and how they may impact turnover outcomes across various industries. This study explored police strategies and practices on turnover. This research also explored aspects related to the transfer of training within the context of policing and was motivated by the potential to inform police organizations, even in a small way, toward potential interventions that can improve transfer of training, thereby resulting in less turnover.

Lastly, several recognized aspects for improvement and future recommendations in turnover research may be explored. Allen, Hancock, Vardaman, and McKee (2014) reviewed 447 empirical turnover studies and contended that any literature growth and field utility from turnover research requires that there is clear articulation of both the context and operational definitions. The context is critical to scope and frame potential turnover antecedents (Allen et al., 2014; Hausknecht & Trevor, 2011; Hom et al., 2017;

Nyberg & Ployhart, 2013). Allen et al. (2014) further contended that the lack of a context focus has created a dominant analytical mindset (DAM) that creates assumptions that likely hamper advances in turnover research, especially toward utility in applied settings. To combat DAM in turnover research, Allen et al. (2014) recommended that future research include context specific observations of voluntary and involuntary turnover using measures beyond surveys of intentions to quit. This research was context specific and used actual turnover data compared to common surveys measuring intentions in keeping with the recommendations of current research (Allen et al., 2014; Hausknecht & Trevor, 2011; Hom et al., 2017; Nyberg & Ployhart, 2013).

Organization of the Remainder of the Study

Chapter 1 introduces the background of the problem, study purpose, research questions, operational definitions, and scope of the study. Chapter 2 reviews literature organized by key constructs relevant to this study. Key literature includes overarching theoretic framework, police recruiting, history of employment practices and legal requirements, situational judgment tests, faking, police field training, transfer of training, and turnover. Chapter 3 contains the research methodology including research design, population and sample, procedures, data collection, analysis, assumptions, and boundary conditions.

Chapter 2

Literature Review

This literature review was conducted to gain insight into the depth and scope of existing research in the research topic in order to provide adequate knowledge to inform this research. This review of literature was intended to establish a foundational overview of the key constructs associated with the study. This research sought to explore potential predictive utility of selection processes, specifically situational judgement tasks, in this case, in policing. Exploring these various potential factors, such as fit and transfer of training, and how it leads to voluntary and involuntary turnover informs an existing gap in the literature with applied implications in reducing turnover for police organizations.

Overview

This chapter offers a review of existing literature to provide an overview of the various research, theories, frameworks, themes, as well as a historical background on relevant employment law and police training practices. This review began with identifying an overarching theoretical framework for the research. Next, literature related specifically to the applied problem and consequences of high police turnover was reviewed. Following this, a historical overview of the evolution of employment practices and the legal environment was discussed. This overview explains the conditions, within employment selection, that explains the current use and utility of *situational judgement tests (SJTs)* by many organizations. While this

research did not explore the validity or constructs of SJTs it is important to explore SJTs, along with relevant *faking* research, to properly ground this critical aspect of this research. The following section provides a historical review and practice details of *field training and evaluation programs* (FTEP) used in police organizations. This is followed by an exploration of research and practical implications of *transfer of training* relevant to the context of this study. The final literature review explored the dependent variable, *turnover*, from a historical and contextually relevant viewpoint.

This chapter concludes with a discussion on the synthesis of the foundational literature, frameworks, and constructs that serve as the historical perspective and inform the theoretical assumptions and boundary conditions for this study. The potential applied utility of this research is later discussed in light of the reviewed literature. Aspects regarding the potential nexus of this research with the discovered future directions, recommendations, and applied potential was discussed.

Questions that Guided the Research

This study explored the existing gap in literature regarding SJT's potential predictive nexus with turnover (Ployhart, 2006; Ryan & Ployhart, 2013; Price, 2011). This study has potential applied research implications in seeking to better understand how SJTs, along with other selection instrument scores, may predict actual turnover, and how it may impact actual voluntary (choice) and involuntary

(performance) turnover. The following overarching research questions were explored with several hypotheses.

RQ1. Do Situational Judgement Tests (SJTs) offer predictive utility toward field training performance.

RQ2. Do SJTs offer predictive utility for turnover outcomes.

The potential impact of exploring these constructs included findings to inform police organizations of a proactive opportunity to develop and provide early interventions toward improved success in the transfer of training related to retention.

Overarching Theoretical Framework

The relevant models, theories, and frameworks of this research rest on a basic assumption that organizations must intentionally select organizational members that are likely to facilitate the organization's goals, vision, and mission. The overarching theoretical framework is important as it establishes the foundational assumptions for a theoretical foundation for the study. The framework for this study was based on Schneider's (1987) attraction-selection-attrition (ASA) model with additional foundational support from person-job fit (Edwards, 1991) and person-organization fit (Judge & Ferris, 1992).

Attraction-selection-attrition (ASA) model. The overarching theoretical framework for this study was Schneider's (1987) attraction-selection-attrition (ASA) model. At ASA's core is Kurt Lewin's 1936 theory about behavior. Lewin's

(1936) behavioral equation $B = f(P, E)$ theorized that behavior is a function of the person and the environment. Schneider (1987) added to Lewin's notion contending that environments are functions of the individuals (actors) behaving within them. Schneider's (1987) theory contends people are not randomly attracted or selected into an organization. They are selected based on criteria established by existing members.

The ASA model describes processes of attraction of potential members to an organization, selection of new members, and attrition. This ASA cycle informs understanding into the processes of organizational dynamics associated with organizational culture. As people are attracted, selected, and remain in an organization they then become the determinants of the environment (Schaubroeck, Ganster, & Jones, 1998). Simply put, people are attracted to others with similar interests, attitudes, and values (Schaubroeck, Ganster, & Jones, 1998).

Organizational actors select new members based on the perception of the potential new member's fit within the organization and the existing members. Poor matches, such as bad fits, lead to attrition (De Cooman, Gieter, Pepermans, Hermans, Bois, Caers, & Jegers, 2009; O'Reilly, Chatman, & Caldwell, 1991; Schneider, et al., 1998).

Organizations understanding this concept can better adapt and manage recruiting and selection strategies. However, there can be negative consequences associated with ASA if not managed strategically. For instance, ASA and fit can

lead to organizational homogeneity in some facets (Schaubroeck, Ganster, & Jones, 1998). Homogeneity among organizational members, for instance, can have negative consequences for the organization in terms of adaptation to change, groupthink, and hampered innovation (Astakhova, Beal, Schriesheim, & Camp, 2015; Schneider, Smith, Taylor, & Fleenor, 1998).

Person job and person organization fit. The *fit* between individuals, professions, and organizations is linked to this study's dependent variable of turnover (Ployhard, 2006). Some turnover is known to be attributed to failures in the transfer training resulting in performance-based dismissals. Some of those who leave for non-performance related issues are theorized to have either person-job (P-J) or person organization (P-O) fit related reasons. It is therefore important to provide an overview to both P-J and P-O fit.

Person-job fit. P-J fit (Edwards, 1991) refers to the fit between individual members' KSAOs and the job, or industry, as opposed to a specific organization (P-O fit) within an industry. P-J fit constructs inform job analysis processes and impact selection, training, and promotions in organizations (Breugh & Starke, 2000; Kristof, 1996; Kristof-Brown, 2002; Werbel & Gulliland, 1999).

Policing is a dangerous and scrutinized profession. P-J fit is critical in public safety positions (Lough & Von Treuer, 2013). Selection batteries in policing that are examining for P-J fit are commonplace (Guller, 1994, 2004; Ho, 2001; White, 2008). Consequences of poor P-J fit include burnout (Maslach & Leiter,

1997) as well as decreased organizational commitment, job satisfaction, and job performance (Kristof-Brown, Zimmerman, & Johnson, 2005).

Person-organization fit. P-O fit garners theoretical support from Schneider's (1987) ASA model. P-O fit represents compatibility between the person and the organization (Piasentin & Chapman, 2006). P-O fit is theoretically rooted to how individuals are accepted and liked by other organizational members (Judge & Ferris, 1992) based on shared characteristics (Kristof, 1996; Verquer, Beehr, & Wagner, 2003). Piasentin and Chapman's (2006) review of P-O fit literature identified four common P-O fit operationalizations including *supplementary*, person has similar characteristics to organizations, *complementary*, where a person fills a vacant need in the organization, *needs-supplies*, where the organization fulfills a person's needs, and *demands-abilities*, where the person's abilities meet the demands of the organization.

The ASA cycle can lead to similarity among members. This *homophily*, where group, cultural, or behavioral characteristics are similar, can lead to a homophily bias if not recognized (McPherson, Smith-Lovin, & Cook, 2001). These group similarities can evolve from ASA cycles and may lead to what Schneider, Smith, and Goldstein (2000) call the "dark side" of good fit. These can include organization-wide diminished innovation, groupthink, and resistance to inclusion and diversity (Jones, McChrystal, Kung, & Griffith, 2002).

P-O fit has empirical support suggesting it has interaction impacts on other

organizational behavioral constructs such as social support (Chao, O’Leary-Kelly, Wolf, Klein, & Gardner, 1994), job satisfaction, (Dawis & Lofquist, 1984; Edwards, 2008; Pindar, 2008; Schaubroeck, Ganster, & Jones, 1998), organizational commitment (Astakhova, 2016), assimilation into the organization’s culture (Chatman, 1989; Kristof-Brown, 2000), tenure (Bretz & Judge, 1994; Schneider et al., 2001), and turnover (Arthur, Bell, Villado, & Doverspike, 2006).

P-O fit has demonstrated nexus to turnover (e.g., Arthur, Bell, Villado, & Doverspike, 2006; Holland, 1993; Schneider, 1987). According to Holland (1993), congruence exists between employee interest and how those interests are reflected in the work environment (1997). Alignment between employee interest, P-O fit, and organization support leads to better retention (e.g., Arthur et al, 1994; Edwards, 2008; Mitchel, Holton, Lee, Sablinski, & Erez, 2001; Schleicher et al., 2011). A meta-analysis review of P-O fit and turnover by Artur et al. (2006) points to attitude constructs linking P-O and turnover.

Impact to assumptions and boundary conditions for this research.

Both P-J and P-O fit have their place in shaping good selection processes given the complex selection criteria for today’s organizations is a combination of both (Bowen, Ledford, & Nathan, 1991; Rynes & Gerhart, 1990; Werbel & Gilliliand; 1999). In a recent review of validity and utility of selection methods over the past 100 years, Schmidt, Oh, and Schaffer (2016) explored meta-analyses of constructs related to job performance. Of specific note, they found no meta-

analytic studies related to P-J fit, P-O fit, or knowledge based SJTs and job performance. ASA provides a theoretical foundation for a predictive examination of interest and attitude constructs (specifically P-J and P-O fit) with actual turnover (Arthur et al., 2006; Holland, 1993; Griffith, et al., 2000; Schneider, 1987; Tett & Meyer, 1993).

Police Recruitment, Selection, and Turnover

The challenge facing police organizations to successfully recruit, select, train, and retain officers was introduced in chapter one. The importance of selection and retention, in any organization, was supported in literature (i.e. Mabon, 1994; Ployhart, 2006; Ployhart & Weekley, 2010; Pulakos & Schmitt, 1996; Rubenstein et al., 2015) and can even be more of a concern for governmental organizations (Barrett et al., 2010; Selden et al., 2013; The Center for State and Local Government Excellence, 2017; Wilson, 2014). The critical legal, process, and pragmatic aspects of proper selection methods were introduced (i.e. Arnold, 2004; Barrick & Zimmerman, 2005; Campbell et al., 1993; Gottfredson, 1996; Gutman, 2009; McDaniel et al., 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Schmidt & Hunter, 1998; Van Iddekinge et al., 2011; Weekley & Jones, 1999).

Hom et al. (2017) contended that a nexus between selection and turnover exists and contend that it is often overlooked in research. They further suggested that this missing nexus is a result of the common disconnect between human resource departments, focused on hiring regulations, and the operational departments where

work is carried out (Ployhart, 2006). This divide, as articulated by Mom et al. (2017), can be amplified in organizations experiencing challenges in diversifying their work force (Barrick & Mount, 2005; Fyfe & Kane, 2006; Ployhart, 2006). Retention of employees can be even more critical to organizations since the costs of training, in time and resources, are significant (Hausknecht & Holwerda, 2013; Mabon, 1994; Ployhart & Weekley, 2010; Rubenstein et al., 2015; Seldon et al., 2013).

Practitioner response. Today, police executives face organizational challenges that are posing a substantial threat to maintaining a stable police force that is necessary to maintain quality of life for citizens. The International Association of Chiefs of Police (IACP) (Smith, 2016; Woska, 2006; Woska, 2016), the Law Enforcement Executive Forum (Wilson, 2014), the Center for State and Local Government Excellence (2017), and the Police Executive Research Forum (PERF) (2014) are among policing and related professional practitioner associations and organizations expressing this concern. Academic support for the increasingly significant erosion toward maintaining a stable police workforce is growing from challenges in effectively recruiting and retaining police officers (i.e. Cavanagh, 2003; Haarr, 2001; Koper, Maguire, & Moore, 2001; Orrick, 2008; Selden et al., 2013; Wilson, 2012).

The IACP's Police Chief Magazine, a premier practitioner publication for police executives, began focusing on the industry recruiting challenges in the early 2000's. Woska's (2006) IACP review of recruiting challenges for policing

included the significance of baby-boomer retirements and the much smaller pool of potential replacements in the current labor pool. Beyond the pure numbers, Woska (2006) discussed competition for an applicant pool from the military, due to personnel ramp-ups and evolving opportunities in the high-tech sector (also mentioned by Smith, 2016). Negative publicity of high profile, often racially charged, incidents regarding use-of-force also have a discouraging effect on applicants toward the profession of policing in general. Woska (2006) adds that eighty percent of the nation's (17,000) police agencies have unfilled positions with as few as five percent of police applicants passing standard law enforcement employment screenings. Compounding policing's challenge for maintaining staffing is new officer training failures, reported as high as 25 percent (Woska, 2006).

Historical context evolving into today's challenges. Today's complex police staffing issues, such as recruiting, selection, and turnover (Ployhart, 2006; Selden et al., 2013) did not arrive unannounced. Ployhart (2006) explored anticipations for 21st century staffing issues related to increasingly complex workforce issues for organizations. Ployhart (2006) suggested that organizations would need to spend more capital and resources on targeted recruiting and training initiatives to compete in the increased knowledge-based workforce. Issues related to evolving work place diversity, with ever increasing competition for applicants, and compounded by generational differences and population dynamics, would

require organizations to expend much more on recruiting and selection as well as effort to retain employees. To best understand the context of the current issues in police staffing we must look at the issues at the turn of the century, the shifts to the police staffing challenges sparked by the recession of 2008, and today's ever evolving and nuances challenges.

Police workforce challenges turn of the 21st century. Police organizations in the U.S. experienced a shortage of qualified applicants in the late 1990's. For example, The Florida Police Chiefs Association (FPCA) commissioned a statewide study to explore issues in recruiting and retention in a search for best practices and solutions (Jones, McChrystal, & Kung, 2002). The FPCA study reported that thirty percent of Florida police and sheriff departments reported difficulties recruiting and half reported challenges recruiting minority applicants (Jones, McChrystal, Kung, & Griffith, 2002). Additional issues emerged related to *poaching* of smaller department officers by large police and sheriff departments leading to interagency distrust (Jones et al, 2002). Generational factors impacting recruiting and retention included questioning authority, organizational commitment, and a lack of mentoring (Griffith, Jones, McChrystal, & Kung, 2002).

A host of best practices and initiatives, such as including police recruiter training, leadership training to improve leader and subordinate relations, marketing police professions, and innovative realistic job previews were being developed (Griffith et al, 2002; Jones et al., 2002; Scott and Jones, 2003). The urgency to

follow through on the state-wide initiative lost momentum as the recession, which began with the stock market losses post September 11, 2001 (Makinen, 2002), led to higher unemployment which, in turn, pushed more qualified candidates to the public sector. Police candidate selection ratios improved, at least for a few years.

Housing market crash and 2008 recession. Police recruiting and retention stabilized in terms of qualified applicants and retention. However, the 2008 U.S recession, sparked in large part by the housing market collapse, is credited with sparking a new challenge in sustaining ample police forces (Police Executive Research Forum, 2013; Wilson, 2014; Woska, 2016). The issues of enough qualified applicants from the turn of the 21st century is replaced with the loss of tax revenues at the state and local level, resulting in various reductions on force issues sparked by drastic budget cuts across government entities.

Wilson (2014) points out that these budget related reductions, such as layoffs, early retirement buy-outs, and elimination of some police agencies, made it difficult maintain proper police staffing levels. Woska (2016) suggests that budget crisis was evident in the decreased average police officer to population ratios nationwide. According the U.S. Department of Justice (2011), officer to population ratios fell from 250 per 100,000 residents in 2008 to a national average of 181 officers per 100,000 residents in 2011. That equated to a 37.6 percent reduction in force.

Woska (2016) warned in 2006 that budget issues were of grave concern to police administrators in his review on police recruiting for the IACP. Woska's 2016

update, focused on police recruiting changes since his previous 2006 review, posits that the 2008 recession critically impacted local government budgets resulting in unprecedented police layoffs, and ultimately police services themselves.

The Police Executive Research Forum (2013) reviewed additional impacts resulting from the budget constraints and shrinking workforces. key issues included eliminating special units, reducing overtime, layoffs, and use of non-sworn personnel and volunteers for functions traditionally performed by officers. Police Executive Research Forum (2013) reports these issues have a direct nexus with officer job satisfaction and direct implications in preserving critical patrol functions for policing.

Contemporary challenges for police staffing. Police executives once had to compete with other police agencies for interested candidates. The challenge then shifted to significant budget constraints leading to downsizing, layoffs, pay freezes, and other fiscal challenges. As the economy rebuilt after the 2008 recession a different set of challenges to maintaining stable police workforces emerged.

Smith (2016) and Wilson (2014) both argue that post-recession economic conditions did not eliminate recruiting and retention problems. They point to the increased competition from other occupations, generational views on policing, and candidate disqualification issues (such as drug use and physical fitness) presenting significant challenges to police organizations. With unemployment low and increasingly complex challenges to police organizations today, police executives face

increasingly complex challenges to navigate. These challenges include negative public perception, job attractiveness, generational issues, and cost of turnover.

Negative public perception. The disparity in finding candidates for public safety may be attributed, in part, to negative police publicity, as first discussed as early as 2006. The notion of potential candidate hesitation for policing as a career gained credibility in the wake of, what St. Louis Police Chief Sam Dotson coined, *the Ferguson Effect* (Gold, 2015). The Ferguson Effect is the effect of anti-cop rhetoric, often fueled by media attention. This public attention that can be based, at times, on incomplete information leads to increased politicization of police-citizen encounters.

Results of the Ferguson effect include negative effects on incumbent job satisfaction (Wolfe & Nix, 2016) as well as perceptions further eroding the attractiveness of policing for the potential applicant pool (Smith, 2016; Woska, 2016). Barrett et al. (2010) recognize the additional potential poor public attention from poor selection. They point to the public sector, particularly police and fire, where the consequences of poor selection can be witnessed on the evening news. These high stakes professions result in the highest level of complaints, criticism, and legal action in the public sector (Barrett et al., 2010) and likely make well qualified candidates ponder other less scrutinized and dangerous professions.

Job attractiveness. Public perception concerns only add to a reduced level of attraction for potential candidates to policing. Challenges posed from private

sector jobs include differentials in pay, benefits, and social acceptability where policing continually falls short (Jordan et al., 2009; Orrick, 2008; Smith, 2016, Wilson, 2014; Woska, 2016). Woska (2016) and Smith (2016) contend reductions in pay and benefits, particularly retirement benefits, as a result of the 2008 recession, have left a significant challenge for police organizations in closing the pay gap with private sector competitors for similar candidate pools. The pay challenges related to attractiveness of policing may be further fueled by the technology boom. Cavanagh (2003) pointed to applicant pool dilution issues facing police recruiting, such as the attractiveness of more lucrative careers and negative images of policing, long before the 2008 recession and Ferguson effect phenomena. The question police executives and recruiters must face is to convince prospective candidates that high-risk and high-profile careers in policing are rewarding in ways beyond comparable pay, benefits, quality of life, and stress from private sector opportunities.

Generational Issues. Ployhart and Weekley (2010) recognize an additional emerging challenge to organizations related to a shift in generational norms. They point to specific challenges related to generational perspectives favoring shorter lengths of employment which lead to the prospect of more turnover. Law enforcement, and public safety professional in general, have long enjoyed good employee benefits.

One such benefit to attracting and retaining officers has been a defined-benefit retirement plan. These plans are an advantage over traditional plans as

retirement, usually around age 55, is guaranteed for life. Traditionally, officers can retire with 20-25 years of service at 60 to 90 percent of their salaries. This benefit has long been an advantage to police recruiters. Many agencies across the country have experienced benefit reductions in the wake of the 2008 recession. Newer officers have a higher chance of being in a contributory plan similar to other local government employees. These retirement plans were once a motivator for retention, without them retention is more challenging (Wilson, 2012).

Generational differences are magnified in smaller police organizations which are more susceptible to workforce shortages. Maciag (2018) discusses small agencies' recognition that officers are being lured away by larger agencies with increasing pay, more attractive job options, and other benefits. Maciag (2018) reports small agency loss of officers as a nation-wide frustration point for small agencies. Agencies do recognize that officers with good performance records and experience are valuable and marketable assets in the labor market (Maciag, 2018). Research by Koper, Maguire, and Moore (2001) extend this concern for qualified applicants with finding that one-half of all small U.S. police departments surveyed (n=409) reported difficulty recruiting qualified candidates. Large departments, serving populations of 50,000 or more residents, (n=538) reported higher difficulty (65%) in finding qualified applicants. It only makes sense that larger departments would consider luring experienced officers.

Smith (2016) points to issues such as a shift from the Boomer's "living to

work” to Millennial’s “working to live” attitudes (Espinoza, Ukleja, & Rusch, 2010). Wilson (2014) suggests that dynamic economic and social constructs related to the profession of policing are challenges in themselves but suggests that the profession of policing’s inability to address generational differences may be the most significant core challenge. Smith (2016) contends that the changing generational attitudes along with relaxing social standards on issues such as drug experimentation and driving histories compound police organization efforts to attract, select, and retain good officers. The National Research Council (2004) and Kaminski (1993) have conducted research that suggests non-whites, particularly African-Americans, have more negative views on police work and thus lower interest in the profession.

Cost of turnover. The cost of turnover to organizations can extend beyond simply finances and time (Barrick & Zimmerman, 2009; Hom et al., 2017; Maltarich, Nyberg, & Reilly, 2010; Van Iddekinge et al., 2011). Police organizations must consider all costs associated with selection and training new employees. The average employee, in any industry or sector, has a learning curve before they can contribute effectively to the organization. This learning curve takes even longer in complex roles (Ployhart & Weekley, 2010), such as policing.

The pragmatic cost of turnover results from poor selection where a new officer is unsuccessful in training and the organization must seek another candidate. Woska’s (2006) finding that eighty percent of U.S. police agencies have unfilled positions

where as few as five percent pass entrance testing, and 25 percent fail during training suggests that police agencies need over dozens of candidates to have a reasonable chance to select, train, and retain one officer.

As an example, Fry (2017), reported on police recruiting and selection records from San Diego Police Department (SDPD). In 2013, SDPD reported 200 vacancies of it 2,000 sworn positions. Only four percent of the 4,439 applicants passed the written exam, necessary background checks, and completed the police academy. Fry did not report the number of academy cadets that made it through training but using Waska's (2006) 25 percent that would leave the section ratio at 3/100. In the case for SDPD, they are losing about 150 officers per year due to attrition including retirement. According to Fry, 2017, SDPD reports 25 percent less applicants between 2015 and 2016. To sustain a stable police force, SDPD must either find around 20,000 qualified applicants to test and process or poach experienced certified officers from other agencies. Even if partially successful at each, replacing 150 officers, on average per year, appears unsustainable.

Impact to assumptions and boundary conditions for this research.

The assumptions for this research included a general need for improved recruiting, selection, training, and retention for U.S. police organizations. Additionally, this research assumed that there is a level of disconnect between human resources and police operations offering an opportunity to share data toward improved retention outcomes. Boundary conditions of this research included that it

is a mid-sized police agency in Florida.

History of Employment Practices and the Legal Environment.

Organizations rely on members. Proper recruiting, selection, training, and retention strategies result in positive employee performance (Ployhart, Nyberg, Reilly, & Maltarich, 2014). Popovich (1998) argues that employee productivity continually improves as they receive training and gain experience, thus facilitating organizational goals. These organizational goals are, in a general sense, broken down into articulable tasks (Lunenburg, 2012) used to identify necessary knowledge, skills, abilities, and other attributes (KSAOs) for adequate performance of tasks and duties of the job (Gutman, Koppes, & Vodanovich, 2011; Liu, Combs, Ketchen Jr., & Ireland, 2007; Murphy, 2012).

Research has long recognized that different attributes, such as KSAOs, exist between people and can be measured (Sharp, 1899; Thorndike, 1914; Wallin, 1911). Metcalf (1912) recognized early in the 20th century the value of measured differences for employment decisions. These notions hold true in employment decisions today (Ackerman & Humphreys, 1990; Arnold, 2004; Gutman et al., 2011, Jepsen and Dilley, 1974). Proper hiring decisions are facilitated through the use of valid employment assessments of these KSAOs and result in much better performance over simple random selection (Guion & Highhouse, 2004; Gutman et al., 2011; Schmidt & Hunter, 1998).

The Society for Industrial & Organizational Psychology (SIOP), 2003,

discusses employment testing and selection measures most commonly used. SIOP (2003) was established to help standardize the manner of testing related to psychometrics of quantifying the job as well as individual and performance characteristics. Organizations must be mindful for the practical and legal aspects of proper KSAO identification and suitable selection processes to fairly assess candidate characteristics (Guion & Highhouse, 2004; Gutman et al., 2011; SIOP, 2003).

Cognitive ability and adverse impact. Use of poorly developed or improper selection procedures have legal implications for organizations. The main considerations for organizations relate to the concepts of disparate treatment and adverse impact. The Equal Employment Opportunity Commission (EEOC), 1978, states *disparate treatment* results when candidates are denied employment opportunities based on membership in a class based on race, gender, or ethnic origin (EEOC, 1978, Gutman et al., 2011).

Adverse Impact, also known as disparate impact, is a selection process concept that examines the proportion of qualified candidates from a protected class compared to the overall pool of qualified candidates (EEOC, 1978; Guion, 1991; Gutman et al., 2011). The EEOC (1978) operationalizes uniform guidelines related to adverse impact as the “four-fifth’s rule” (Gutman, 2009). This scenario considers the potential adverse impact to protected class members, when qualified applicants from that protected class (race, gender, ethnic origin) have a selection rate less than

four-fifths of the selection rate for non-protected class members (EEOC, 1978; Gutman, 2009; Gutman et al., 2011).

Cognitive ability tests have strong empirical support as perhaps the current best psychometric predictor of job performance (i.e. Chan, Schmitt, DeShon, Clause, & Delbridge, 1997; Kuncel & Hezlett, 2010; Maltarich & Reilly, 2010; Roth, Bevier, Bobko, Switzer III, & Tyler, 2001; Schmidt & Hunter, 1998). In fact, Schmidt, Oh, and Shaffer's (2016) recent review of research related to selection led to their contention that researchers have historically *underestimated* the effects of cognition related to performance. The Schmidt et al. (2006) assertion is based on their recognition on the recent advances in measurers and analytical techniques. These statistical and analytical methods, according to Schmidt et al. (2006), yield much more accurate results due to these more advanced methods in correcting for range restriction.

The challenge in using cognitive ability for selection. The historical and legal issues with the use of cognitive ability tests are they have been shown to produce the largest score differences in racial and ethnic origin scores leading to higher possibilities of adverse impact (Breaugh, 2003; Cohen, 1988; Gaugler, Rosenthal, Thornton, & Benson, 1987; Gutman et al., 2011; Hough, Oswald, & Ployhart, 2001; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001; Roth, Bobko, & Buster, 2013; Roth, Bevier, Bobko, Switzer & Tyler, 2001; Sackett & Ellingson, 1997; Weekley & Jones, 1997; Weekley, Ployhart, & Harold, 2003;

Whetzel, McDaniel, & Nguyen, 2008).

The law does not consider adverse impact, in itself, as discriminatory (EEOC, 2010; Gutman et al., 2011; SIOP, 2003). Violations of the four-fifth's rule can initiate an EEOC complaint and potential costly litigation (EEOC, 2010; Gutman et al., 2011; SIOP, 2003). Instruments producing adverse impact are not automatically deemed improper or illegal. They can be used with proper validation (Hough, Oswald, & Ployhart, 2001) but it could negatively impact workplace diversity initiatives (Lievens & Sackett, 2007). The U.S. Supreme Court has explored discrimination claims with rulings that prefer alternative predictors to cognitive ability that produce lower levels of adverse impact (i.e. *Albemarle Paper Company v. Moody*, 1975; *Griggs v. Duke Power Co.*, 1971). The EEOC followed these rulings with their 1978 uniform guidelines (EEOC, 1978) stressing alternative predictor options. These court rulings lead to a shift in selection practices from cognitive ability towards other tests such as personality measures (i.e. Janik, 1994; Gutman et al., 2011; Weiss & Inwald, 2018; White, 2008).

The rise and fall of personality tests in police selection. In the mid-1970's states started conforming to the increased scrutiny, which generated more oversight and government interventions in employment testing in policing (Janik, 1994; Landy, Farr, Saal, & Freytag, 1976; Weiss & Inwald, 2018; White, 2008; Walker & Katz, 2002). Two federal reports in 1967 sparked action. The first, The Presidential Commission on Law Enforcement and the Administration of Justice recognized the

need for agencies to screen for emotional stability. The other, The National Advisory Commission on Annual Justice Standards and Goals recommended that all police agencies utilize psychological testing professionals to improve police selection. (Janik, 1994; Landy, et al., 1976; Weiss & Inwald, 2018; White, 2008; Walker & Katz, 2002).

Early state legislative mandates called for pre-employment screenings for significant psychopathology as a screen-out mechanism (Landry et al., 1976; Weiss & Inwald, 2018). This police initiative gained momentum in 1976 with funding for the National Working Conference on the Selection of Law Enforcement Officers funded by the Department of Justice. This conference was the first of many to advance the use of various personality assessments in policing (Weiss & Inwald, 2018). As research evolved, the nexus between personality test variables and police work emerged. Inwald and Shusman (1980) compiled data on over 1,000 pre-employment assessments and completed the Inwald Personality Inventory (IPI) designed and validated specifically for policing.

A 1981 lawsuit verdict in New York (*Bonsignore v. The City of N.Y., 1981*) reinforced to police administrators and government officials that the need, and expectation, for psychological tests for police officers was critical. The case involved a NYPD officer who used his service weapon to shoot his wife and then commit suicide. His wife survived and filed suit in federal court, claiming that the department had not done its due diligence in screening. The lower and appellate

courts agreed that the NYPD had not implemented sufficiently effective psychological screening processes (*Bonsignore v The City of N.Y.*, 1981). This served as a landmark case and a wakeup for police agencies ignoring psychological screening processes (Weiss & Inwald, 2018).

Psychological testing in the 1980's experienced dramatic increases. Not only did selection processes include personality tests but expanded use to fitness for duty evaluations for existing officers (Inwald, Kaufman, & Roberts, 1991). By the end of the 1980's, roughly half of police departments used some sort of psychological screening, signifying a huge advance in applying psychology in the field (Weiss & Inwald, 2018). However, validity issues with selection personality tests began to emerge (Gutman, 2009; Kleiman & Gordon, 1986).

Of particular note, generalizability of personality tests across industries and specific jobs, in a selection context, drew criticism (Lough, Wald, Byrne, & Walker, 2007) although some studies, with correctional officer samples for example, demonstrated partial generalization to police populations (Barret, Polomsky, & McDaniel, 1999; Brough, 2005). Despite the IPI's validation specifically using police performance evaluations, many other common instruments were scrutinized for validation with police academy cadets as opposed to tenured officers' job performance (Burkhart, 1980; Gonder and Gilmore, 2004; Hargrave, Hiatt, & Gaffney 1986; Inwald & Shusman, 1984; Weiss & Inwald, 2018). Given more recent research suggesting that only as much as ten percent of academy

performance accounts for field performance (Buerger, 1998; Caro, 2011) in policing this was a valid concern. Studies (i.e. Bartol, Bergen, Volckens, & Knoras, 1992; Burbeck & Furnham, 1985; Daniels & King, 2002) had similar conclusions determining if psychological tests, specifically the commonly deployed Minnesota Multiphasic Personality Inventory (MMPI), could discriminate for police selection purposes predictively for performance. Finally, three studies (Daniels & King, 2002; Winters, 1990; Wright, Dai, & Greenbeck, 1990) failed to find prediction between MMPI test scores and police performance.

Americans with Disabilities Act (ADA). Weiss and Inwald's (2018) historical review of personality testing in police employment pinpoints that the landscape for psychological testing for pre-employment screening in policing came to a quick decline in 1990 with the passing of the Americans with Disabilities Act (ADA). The ADA effectively eliminated the use of psychological testing for pre-employment screening. Police psychologists, at the time, lobbied for continued use, with empirical evidence, to not avail prior to the 1992 ADA implementation (Inwald, Kaufman, & Roberts, 1991). This effectively abolished the use of personality-based testing for pre-conditional police selection (Weiss & Inwald, 2018).

The ADA only allowed for psychological screening in post-conditional situations where failing a battery would then eliminate a candidate from a position they have been conditionally offered (Gutman, 2009; Weiss & Inwald, 2018). The

loss of pre-employment utility was significant, but fitness for duty, as sparked by the *Bonsignore* decision in 1981, maintained utility (Weiss & Inwald, 2018). The value in the IPA and MMPI remains a potential screen-out, although post conditional, police selection tool (Weiss & Inwald, 2018).

The Civil Rights Act of 1991 and Ricci. The Civil Rights Act of 1991 (CRA 1991) further eroded the conditions where psychological testing, and other common police selection methods, could be used in employment (Gutman, 2009; Gutman et al., 2011; Weiss & Inwald, 2018). In essence, the CRA 1991 mandated that selection had to be made based on the *best qualified* candidate. No longer could organizations consider race or gender for preferential hiring (Gutman, 2009; Gutman et al., 2011; Weiss & Inwald, 2018). The legal requirements established by the ADA and CRA 1991 for police, as well as other industries, mandated a search for alternative employment selection instruments leading to the current trend and use of measures such as situational judgment tests (SJTs).

The United States Supreme Court (USSC) reinforced the “best qualified” notion of the CRA 1991 in the firefighter promotion case *Ricci v. DeStefano, 2009*. In this case, the City of New Haven, Connecticut, (city) failed to certify the results of a lieutenant promotional exam for 118 candidates because the city feared a disparate impact suit from African-American candidates (Campbell, 2010; Dickerson, 2009; Mitchell, 2013). The white and Hispanic candidates filed suit. The USSC found that the city had conducted a proper job analysis to identify

KSAOs for fire lieutenant and, based on the job analysis, developed a written examination to measure the job-related knowledge (Mitchell, 2013). The city then compiled training manuals and study materials and provided a three-month study period (Mitchell, 2013). Outside assessors were used. All held higher rank than lieutenant, with sixty percent of assessors being from a minority (Mitchell, 2013). The USSC ruled in *Ricci* that a city's fear of potential litigation from a protected group is not reason to ignore results of a valid test (*Ricci v. DeStefano*, 2009).

Legal observers (i.e. Biddle & Biddle, 2013; Campbell, 2010; D'Aquila, 2010) contend that public sector organizations face an impossible chore of perceptions of discrimination from any group not proportionally represented in a promotional context. Biddle and Biddle (2013) suggest organizations comply with proper and legal selection guidelines with the goal of facilitating that the best qualified candidates emerge, without regards to class, and let the legal process resolve issues that may arise. Assessment centers for promotion (Biddle & Biddle, 2013; Mitchell, 2013) and SJTs for selection (Barrett, Doverspike, & Young, 2010; Chan & Schmidt, 2002; Legree & Psotsha, 2006; Lievens, Buyse & Sackett, 2005; Hausknecht & Langevin, 2010; McDaniel et al., 2001; Hanson & Ramos, 1996; Hunter, 2003; Peeters & Lievens, 2007; Weekley & Ployhart, 2006) have emerged as contemporary, but costly, options.

Alternatives to cognitive ability and personality. The ADA and CRA 1991 lead to a quest for alternatives to cognitive ability and personality in police

employment testing (Weiss & Inwald, 2018). SJTs have become increasingly popular in selection (Barrett et al., 2010; Chan & Schmidt, 2002; Legree & Psotsha, 2006; Lievens et al., 2005; Hausknecht & Langevin, 2010; McDaniel et al., 2001; Hanson & Ramos, 1996; Hunter, 2003; Peeters & Lievens, 2007; Weekley & Ployhart, 2006). SJTs present job candidates with job-relevant situations and ask them to choose among a list of possible actions in response (Whetzel & McDaniel, 2009).

Meta-analysis by Whetzel and McDaniel (2009) point to the popularity of SJTs in the U.S. and Europe and cite several reasons. They discuss a renewed research interest and allow deployment of more sophisticated data collection and analysis, such as meta-analytic studies. These meta-analyses have offered empirical support for SJT which includes better validity predicting job performance relative to other selection options, such as cognitive ability and personality (i.e. Christian, Edwards, & Bradley, 2010; Whetzel & McDaniel, 2009). Research supports that SJTs produce less adverse impacts than other common selection testing methods, thus allowing organizations to deploy selection methods in keeping with case law and EEOC guidelines (Barrett et al., 2010; Callanin & Robertson, 2000; Chan & Schmidt, 1997; Motowidlo & Tippins, 1993; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Weekley & Jones, 1997; Whetzel, McDaniel, & Nguyen, 2008).

Impact to assumptions and boundary conditions for this research. The primary applied purpose of this research was to help law enforcement determine if selection battery components, particularly SJTs, offer any additional utility toward predicting officer retention. Given the legal landscape where SJTs are preferred in light of the limitations of cognitive ability and personality batteries, the potential impacts in the field supported the importance of the nexus between SJTs and KSAOs that are related to officer performance through transfer of training.

The key assumption of this section related to the research was that the selection batteries deployed by the population agency are valid and reliable measures of initial police performance and that they conform to legal guidelines of selection. Given that the focus is on potential expanded utility beyond selection, this assumption served as a boundary condition for the study.

Situational Judgment Tests

McDaniel, Hartman, Whetzel, and Grubb III (2007) describe situational judgment tests (SJTs) as selection instruments that typically tap into behavioral tendencies or knowledge in work-related scenarios. Candidates are presented a list of plausible actions related to the scenario and are evaluated on how they would perform or how effective their selected action (option) would be in the job context (Whetzel & McDaniel, 2009).

Whetzel and McDaniel's (2009) article discussing the history of SJTs tracked their use back to the 1870s with selection for the U.S. Patent Office. Early

Benet children's intelligence tests in 1905 (Whetzel & McDaniel, 2009) included situational judgment items. McDaniel and Nguyen (2001) discuss the use of SJTs specifically to improve selection outcomes since the 1950s. Throughout the early 20th century practical judgment tests were used for selection including for the U.S. Office of Personnel Management, which continues today.

SJT popularity today. Meta-analysis by Whetzel and McDaniel (2009) pointed to the popularity of SJTs in the U.S. and Europe and cited several reasons. They discussed a renewed research interest and allowed deployment of more sophisticated data collection and analysis, such as meta-analytic studies. These meta-analyses have offered empirical support for SJTs which includes better validity with predicting job performance relative to other selection options, such as cognitive ability and personality measures (i.e. Christian, Edwards, & Bradley, 2010; Whetzel & McDaniel, 2009). Research supports that SJTs produce less adverse impact than other common selection testing methods, thus allowing organizations to deploy selection methods in keeping with case law and EEOC guidelines (Barrett et al., 2010; Callanin & Robertson, 2000; Chan & Schmidt, 1997; Motowidlo & Tippins, 1993; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Weekley & Jones, 1997; Whetzel, McDaniel, & Nguyen, 2008).

Candidate perceptions of fairness. SJTs offer advantages in selection due to their face and content validity related to the work scenarios, leaving candidates more satisfied with the perceived fairness of the instruments (Rynes & Connerly,

1993; Whetzel & McDaniel, 2009). Hausknecht, Day, and Thomas (2004) conducted a meta-analysis and found that job applicant perceptions of fairness are high when tests are face valid and have perceived predictive validity.

Organizational advantages. Organizations are also attracted to SJTs for many reasons, including economy of development, administration, and scoring (Lievens & De Soete, 2012), with fewer items compared to cognitive ability tests (Hanson & Ramos, 1996), while also offering improved face validity (Clevenger, et al., 2001; Hausknecht et al., 2004; Ployhart & Ryan, 1998; Rynes & Connerly, 1993; Whetzel & McDaniel, 2009). Face valid instruments and assessment are associated with more positive candidate reaction thus reducing negative perceptions that can impact future behaviors of employees (Anderson, 2011; Anderson, Salgado, & Hülshager, 2010; Hülshager & Anderson, 2009; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). These SJT advantages collectively reduce organizational chances of litigation claims associated with selection and later performance (Barrett et al., 2010; Kaminski, 2002; Hoover, 2002; Murphy, 2012; Walker, 2005).

Previous SJT studies predicting turnover. There is a recognized void in literature exploring SJTs and turnover (Barrick & Zimmerman, 2009; Whetzel & McDaniel, 2009). Rubenstein et al. (2018) suggested that future turnover research explore actual turnover as opposed to *turnover intentions*. Support exists for more applied turnover research to inform both practitioners as well as to validate prediction

criteria for selection research (Hom et al., 2017; Van Iddekinge et al., 2011).

Literature searches revealed only one article exploring potential SJT prediction of turnover. Dalessio (1994) explored video-based SJTs and insurance agent turnover. He was exploring the potential predictive utility of SJTs and retention of agents at one year. The insurance industry experienced high turnover (average 44 percent annually) and Dalessio's (1994) article was an effort to potentially inform the insurance industry toward hiring retention. The applied goal of the project was to explore potential use of SJTs as a prediction of retention during hiring processes (Dalessio, 1994).

Dalessio's 1994 study population consisted of newly contracted agents (N=677) across 14 insurance companies. The SJT was not part of a selection battery, as in the current project, but part of post-selection training. This SJT training was conducted within the first eight months of employment with most, 90 percent, of new agents completing the SJT within their three months on the job. The dependent variable was continued activity with the company after one year (Dalessio, 1994).

Dalessio (1994) reported predictive findings consistent with those of realistic job previews (.09) and job enrichment (.17) on turnover (McEvoy & Cascio, 1985). Dalessio (1994) contended that his study likely underestimated the relationship between the SJT and retention, citing both statistical issues related to calculating dichotomous criterion (Hunter & Schmidt, 1990) and a retention rate of

the population of 67 percent, 10% higher than the industry average.

Dallessio (1994) suggested that his research results support the use of SJT in selection decisions, as opposed to post-selection training. He argued SJTs may be a proxy for realistic job previews, but additional research is necessary. He also recognized that the timing of the SJT test delivery in his study, at least three months after selection, potentially contaminates the subjects with training and some level of experience, which may have reduced his predictive analysis.

An aspect of Dallessio's (1994) study, shared with this study industry, is the high cost of turnover experienced by insurance agencies in early 1990's and policing today. In both, identifying behavioral patterns predictive of turnover, even small ones, can be helpful in better retention and significant cost savings (Dallessio, 1994).

Previous police SJT and related studies. While research related to predictive nature of SJTs is lacking, research on SJTs and other measures in the police context exist in the literature. A review of these studies helps inform the current study and offers some comparison for context and potential performance constructs.

Swedish police and cognition, personality, and fitness. Anell, Lindfors, and Sverke (2015) studied 750 Swedish police recruits and explored the relationship between selection methods of cognitive ability tests, personality inventories, and fitness tests and how they may relate to job performance,

satisfaction, retention, and health. This longitudinal study examined the measures at three points in time: first, at the end of the police academy, then after six months of field training, and finally after the first year of employment.

Annell et al.'s (2015) findings included that verbally loaded cognitive tests predicted academic and work performance while general cognition scores negatively predicted only satisfaction over the three periods in the longitudinal study. Personality was predictive except for job performance, which is a common criticism of using personality measures in selection processes. This study informs the Annell et al.'s (2015) previous finding related to verbally-loaded cognitive ability tests have prediction ability for performance. Cognitive ability is an indirectly measured sub-score in the SJT section battery for this study's population.

Netherlands police multimedia SJT. Lievens, De Corte, and Westerveld (2015) explored 208 police applicants in the Netherlands where multimedia SJTs were deployed. Behaviorally constructed responses, where candidates act out their responses as opposed to writing them down, were found to be better predictors of performance after one year, but the authors did note a small effect size. Lievens and colleagues (2015) viewed this type of test as a hybrid between assessment centers and SJTs but do note the time and cost being high as well as potential issues with interrater reliability.

SJT versus assessment centers. Lievens and De Soete (2012) further argued that SJTs are more practical and affordable to agencies than high-fidelity, time,

resource dependent, and expensive simulations such as assessment centers. Barrett et al.'s. (2010) review of police selection points to increasing use of video-based SJTs as alternatives to the cost, in money and time, to assessment centers.

Advantages of SJTs over assessment centers is more than cost, as research demonstrates similar validity coefficients to performance (Lievens, Buyse, & Sackett, 2005; Ployhart & Holtz, 2008) and productivity (Baker & Gebhardt, 2012).

Singapore civil service SJT. Chan and Schmitt (2009) examined situational judgment tests' ability to predict job performance of 160 Singapore civil service employees. This study examined overall job performance along with subdomains of technical proficiency, job dedication, and intrapersonal facilitation. Their research was consistent with a body of growing research that suggests situational judgment tests have higher validity over the combined validity of cognition, personality, and experience as previously discussed in this chapter.

SJT format: video-based vs. written scenarios. SJTs originally were developed in paper and pencil formats but adapted in recent years to video and multimedia formats (Brown, Jones, Serfass, & Sherman, 2016; Dalessio, 1994; Funke & Schuler, 1998; Lievens & Sackett, 2006; Weekley & Jones, 1997). Lievens and Sackett (2006) specifically studied the impact of transition from video-based SJTs to paper and pencil tests. This situation was possible due to a transition from video-based to written scenarios at a Belgian medical school related to admissions. The demographics of the populations (over 1,100 subjects in each

group) were similar and the results were examined. The video-based SJTs were better predictors for intrapersonal oriented criteria than the written scenario-based paper and pencil, which was more closely correlated to cognitive ability. This is an interesting finding given cognitive ability tests have more issues related to adverse impact, supporting the video-based mode.

Cucina, Su, Busciglio, Thomas, and Peyton (2015) looked at video-based testing and found a predictive nexus to training outcomes and job performance. They concluded that video-based testing also assisted in rater reliability. They suggested the use of video-based testing where applied social and interpersonal skills are job-related. Law enforcement is certainly one such area.

Brown et al. (2016) added clarity to reasons why video-based SJTs may be better than written scenarios. They discussed that scenario situations align more with real life situations, which they denote help frame the situation which then offers better prediction in determining actual human behavior. These sets of cues, characteristics, and classes of situations and events associated with life provoke psychologically based behavioral responses. Brown et al. (2016) did discuss and acknowledge that clearly articulating and operationalizing the situations from an academic construct perspective is a dilemma but also did suggest that SJTs offer rich and face valid workplace related situations to assess how someone may behave at work.

Naemi, Martin-Raugh, and Kelly (2016) further expanded on video-based

and virtual SJTs as adding richness over written situations in predicting behaviors such as interpersonal skills. They point to media richness theory that posits the expression of multiple cues and multiple methods improves message delivery. In the case of video-based SJTs these include nonverbal cues to include subtle behavioral nuances, such as body language. Naemi et al. (2016) also asserted that these aspects act as a measure of general domain knowledge.

SJT criterion (job performance) validity. McDaniel, Hartman, Whetzel, and Grubb III (2007) used a meta-analysis (sample of over 19,000 tests with nearly 30,000 criterion validity sources) and collected data from not only researchers, but also intentionally included data from consultants that routinely use SJTs in their work. One of their key findings was the incremental validity of SJTs over cognitive ability tests, the Big Five personality traits, and a composite of both. This meta-analysis is consistent with a large body of research suggesting that SJTs have better predictive ability than cognition and personality tests. The meta-analysis adds strength to the argument by looking at such a large and diverse sample.

Calvasina and Calvasina (2016) point out that the U.S. law pertaining to employment does not prohibit any kind of test, provided that it is a valid and reliable predictor of appropriate job-related performance. McDaniel et al.'s (2007) meta-analysis does offer convincing evidence, compared to cognition and personality, that SJTs are reasonably predicting appropriate job performance.

SJT construct validity debate. SJTs research provides arguable evidence

of criterion-related validity (Campion, Ployhart & MacKensie Jr., 2014; Christian, Edwards, & Bradley, 2010; McDaniel, et al., 2001) as they relate to the criteria of work performance, but the constructs as to *why* are still elusive. While the use of SJTs is widely accepted and has demonstrable validity advantages over cognitive and personality measures, there still lacks a depth of operationalization and theoretical explanations for the exact *constructs* being measured by the SJTs (Motowidlo, Dunnette, & Carter, 1990). The discussion and debate of exactly *what* SJTs tap into is garnering attention.

Meta-analytic support for SJT nexus with constructs of leadership, interpersonal, and team skills. Christian, Edwards, and Bradley (2010) conducted a meta-analytic review of SJTs to possibly address the lacking literature on construct measures associated with SJTs in a quest to establish theory-driven frameworks. Their research resulted in support for SJTs assessing leadership skills, interpersonal skills, and where applied, teamwork skills. Their findings included that matching predictors to criterion constructs improves criterion validity. They also suggested that video-based SJTs possess stronger criterion validity than written scenarios.

Police SJT and integrity construct. A SJT designed to tap police integrity was conducted by de Meijer, Born, van Zielst, and van der Molen (2010) where they examined a multi-ethnic section of Dutch police. They found that when SJTs were used to measure one construct, in this case integrity, it had construct validity

across several ethnic populations. Husbands, Rodgers, Dowell, and Patterson (2015) also found psychometric robustness for integrity SJTs among medical student applicants.

Current theory and validity efforts. Lievens and Motowidlo (2016) are sparking debate and academic engagement toward clarifying constructs and theory development of SJTs. Their goal is to better inform *why* SJTs work and *what* they measure. This effort cumulated in a dedicated issue of *Industrial and Organizational Psychology* in March 2016. Lievens and Motowidlo (2016) offered a general domain knowledge model as the special edition focal paper. Twelve additional papers from SJT researchers were included offering a variety of validity and construction critiques.

General domain model. Lievens and Motowidlo (2016) offer a model proposing SJTs as a measure of *general domain knowledge* not contingent on context. They define general domain knowledge as, “knowledge about the utility or importance of traits such as these for effectiveness in a job that actually requires expressions of these traits for effective performance” (p. 4).

Lievens and Motowidlo (2016) contend that SJTs tap a variety of job-relevant content domains, as depicted in Figure 1, of which many pertain to interpersonal relations. This notion is supported by previous research by Lievens, Buyse, and Sackett published in 2006 that examined over 7000 Belgian medical school applicants. This high stakes testing situation found that SJTs were better

predictors of grade point average than cognition in curricula related aspects of medical school performance related specifically to interpersonal behaviors.

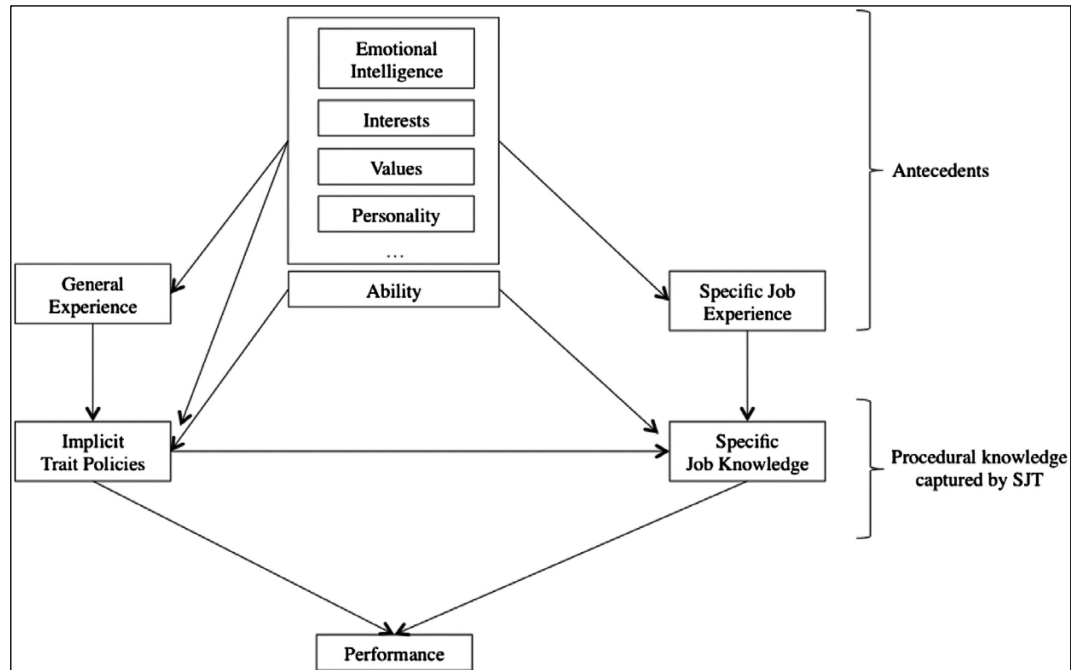


Figure 1. Lievens and Motowidlo determinants and antecedents of SJTs.

Adapted from “Situational judgement tests: From measures of situational judgement to measures of general domain knowledge” by F. Lievens and S.J. Motowidlo, 2016, *Industrial and Organizational Psychology*, 9. P. 4. Copyright 2015 by Society of Industrial and Organizational Psychology.

The general domain knowledge (Lievens & Motowidlo, 2016) argument has some support from meta-analysis. Clevenger, Pereira, Wiechmann, Schmitt, and Veronica (2001) examined federal investigative officers, customer service personnel, and manufacturing engineers and found low correlations and no

statistical significance between SJT scores and previous job experience.

Interestingly however, the various SJTs among the three populations studied by Clevenger et al. (2001) did produce mixed results as related to tapping into cognitive ability. This suggests that SJTs have some variance of validity depending on type of industry or job domain.

Dissenting views to general domain knowledge. The general domain knowledge theory (Figure 1), as proposed by Lievens and Motowidlo's (2016) has critics. Fan, Stuhlmán, Chen, and Weng (2016) argue that while they agree with the general domain notions supported by Lievens and Motowidlo, they believe the context is critical to providing important nuances necessary to SJTs. Harris, Siedor, Fan, Listyg, and Carter (2016) support Lievens and Motowidlo's (2016) general domain knowledge approach toward a quest for more generalizable and non-context specific SJT validity. They do, however, argue that the general domain knowledge and situational context are both critical toward meaningful SJTs.

Crook (2016) argues that discounting context and displaying SJT situations in a general, non-specific, context can have unintended consequences, such as reduced face validity and thus lower applicant perceptions of fairness. Crook (2016) contends that SJTs could lose utility on section contexts. Harvey (2016) adds to the concerns for unintended consequences in off-the-shelf generic SJTs, as suggested by Lievens and Motowidlo (2016). Harvey (2016) contends that scoring SJTs requires context for raters. For instance, interpersonal communication would

rate much differently for a customer service candidate versus a police candidate.

Whetzel and Reeder (2016) add to the academic dissent about a general, context neutral, approach to SJTs. They point to literature that explores failure of success of SJTs to predict performance. In cases where SJTs fail at predicting performance, Whetzel and Reeder (2016) point to issues in test construction and interpretation. Interpretation issues can emerge at any level of SJT development, deployment, or interpretation. For example, subject matter experts developing item response options, test takers, and rater perspectives can vary due to context specific aspects of a particular job. Context lessens the possible impact of ambiguity to the situation and thus the behavioral construct being explored.

Practical intelligence and context specific constructs. Chan and Schmitt (2009) offer a slightly different operationalization of SJTs and contend that SJTs are multidimensional tests of individual differences. In this context the testing scenario is designed to predict performance from a work context, rather than traditional cognitive and personality tests, that may signal work-related knowledge, skills, abilities, and other attributes (KSAOs). The results include support for SJTs tapping into practical intelligence as well as adaptive performance related to job situations. Additionally, Chan and Schmitt's (2009) study suggests SJTs measure individual differences not associated with previous experience and that are distinct from cognitive and personality tests, consistent with Clevenger et al. (2001).

Contextual framing viewpoint. Brown et al. (2016) challenge SJT theorists

to examine the role of *situation* in SJTs. McDaniel, List, and Kepes (2016) address this notion with their SJT construct validity theory. McDaniel et al. (2016) discuss that the situational scenarios may be viewed as simply a better way to frame questions in a manner that reduces ambiguity for test takers and in item responses. Melchers and Kleinmann (2016) point to the *situational judgement* part of SJT arguing that a general domain approach all but eliminated the *situation* from valuable contextual framing.

McDaniel et al. (2016) assert that SJTs may allow for both reflection of job-related knowledge and general knowledge domains. They discuss the challenge of attempting to identify specific traits associated with an item response on a SJT item when the item may be tapping into many traits at once (McDaniel et al., 2016). Chen, Fan, Zheng, and Hack (2016) echo Brown et al. (2016) and McDaniel et al. (2016) in calling for more focus on specifications of the situation to build context.

Tacit knowledge argument. Torres and Beier (2016) also make a case that SJTs may also tap tacit knowledge. Tacit knowledge is defined by Sternberg and Wagner in 1993 as “practical know-how one needs for success on the job” (as cited in Torres and Beir, 2016). While it is hard to argue the importance of tacit knowledge in any work situation, it is challenging as a clear and defined construct.

As the construct debate continues it is clear that operationalizing various theories related to why and what SJTs measure is critical (Borneman, 2016, Chen et al., 2016; Crook, 2016; Lievens & Motowidlo, 2016; Torres & Beier, 2016).

Various theoretic views, such as general domain knowledge critical (Lievens & Motowidlo, 2016), practical intelligence (Chan & Schmitt, 2009), contextual framing (McDaniel et al., 2016), and tacit knowledge (Torres & Beir, 2016) are attempting to tap into similar behavioral aspects related to performance that are currently challenging to precisely define and operationalize.

Lievens (2017) supports the agenda among SJT researchers to explore constructs related to prediction. While criterion validity, such as overall performance, is established (i.e. Calvasina & Calvasina, 2016; Campion et al., 2014; Christian, et al., 2010; McDaniel, et al., 2001; McDaniel et al., 2007) the real potential advancement is to establish a body of research robust enough to serve as basis for meta-analytic studies of the long sought constructs associated with SJTs (Lievens & Motowidlo, 2016; Schmidt, Oh, & Schaffer, 2016).

Impact to assumptions and boundary conditions for this research. This research did not examine construct or criterion validity of SJTs. The key assumption, for this proposal, was that SJTs are reliable and valid selections for the population of the study. It is also an assumption that the existing literature supporting the prediction of performance by SJTs was relevant to the population. Boundary conditions related to SJTs include the domain of law enforcement, specifics of the SJT and sub-scales associated with the test are generalizable to the industry, as used by the Florida mid-sized police organization.

Faking

The history of employment testing has evolved, due in large part to the advances in scientific observation adding degrees of precision between true test scores related to actual work performance (Schmidt & Hunter, 1998; Schmidt, Oh, & Shaffer, 2016). Schmidt, Oh, and Shaffer (2016) reviewed the last 100 years of selection research and reported significant advances over the past 20 years. One of the reported advancements is related to statistical procedures correcting for range restrictions (Hunter, Schmidt, & Le, 2006). Another important finding by the Schmidt et al., (2016) review was the assertion that historic validity measures and analytical techniques *underestimated* the effects of cognitive ability related constructs. This is important to all aspects of selection literature given the recognition of cognitive ability and performance (Schmidt & Hunter, 1998; Schmidt, et al., 2016). In the context of faking, cognition is important in areas of potential faking countermeasures in item response development where higher cognitive load items with low transparency are conceptualized to mitigate faking opportunity (McDaniel & Nguyen, 2001).

It is psychometrically difficult to measure latent constructs, such as cognition, personality, integrity, or conscientiousness, without recognizing and accepting some level of unsystematic variance (Kiefer & Benit, 2016; Miller, 2015; Schmidt, Oh, & Shaffer, 2016). Research confounds not related to measurement error, such as testing fatigue or boredom, can be statistically controlled (e.g.

McFarland & Ryan, 2000; Miller, 2015; Schmidt et al., 2016). However, potential serious threats to the validity and reliability of any instrument come from test takers intentionally faking responses (e.g. Converse, Peterson, & Griffith, 2009; Griffith & Peterson, 2011; McFarland & Ryan, 2000; Rosse, Stecher, Miller, & Levin, 1998). McFarland and Ryan (2000) define faking as a conscious distortion of applicant responses on employment tests to score favorably and increase their chance of selection.

To what extent do applicant's fake? This is a very legitimate concern (Converse et al., 2009; Ones, Viswesvaran, & Korb, 1995) with some researchers contending test takers fake with the potential to impact selection (e.g. Birkeland, Manson, Kisamore, Brannick, & Smith, 2006; Donovan, 1999; Hough, Eaton, Dunnette, Kamp, & McCloy 1990). Research on faking during selection has focused heavily on non-cognitive self-report measures, such as personality tests, biodata inventories, and integrity tests (Alliger & Dwight, 2000; Dalen, Stanton, & Roberts, 2001; Graham, McDaniel, Douglas, & Snell, 2002; Martin, Bowen, & Hunt, 2002; McFarland, & Ryan, 2000; McFarland, Ryan, & Aleksander, 2002; Ones & Viswesvaran, 1998). Another potential confound in the exploration for a generalizable understanding in faking research recently gaining attention is faking related to culture. Fell and König (2016) explored faking across 43 countries to examine cross-cultural effects on faking. Their results suggest that faking is related to culture in constructs including uncertainty avoidance, collectivism, and

assertiveness (Fell & König, 2016). This suggests that not all candidates have a propensity to fake at the same rates, which is one of many variables in seeking solutions, such as countermeasures, to faking in test design.

Instruments using self-ratings are opportunities for deliberate distortions (Birkeland et al., 2006; Hough et al., 1990; Griffith, Chmielowshi, & Yoshita, 2007). In studies where subjects are instructed to fake good (e.g. Viswesvaran & Ones, 1999) they can successfully do so. Meta-analytic findings by Ones, Viswesvaran, and Korb (1995) suggest that as much as one-half standard deviation in scoring can be manipulated by faking. This is enough to impact who may be selected. What remains unknown is the to extent that actual applicants, as opposed to lab-controlled subjects, actually engage in faking and what, if any, impacts results in distortions to validity that could actually impact selection decisions (Ones et al., 1995; Rosse et al., 1998; Ziegler, MacCann, & Roberts, 2011). There is a growing body of literature that estimates at least, and upwards, of 30 percent of actual job applicants fake (e.g. Converse et al., 2009; Griffith et al., 2007).

SJT faking. While faking research for personality, biodata, and integrity tests have extensive review, the research on SJT faking is limited (Peeters & Lievens, 2005). There have been studies focused on intentional distortion on SJT faking (Juraska & Drasgow, 2001; Lievens & Peeters, 2008; Schmidt & Wolf, 2003) with mixed results. The critical concern for potential faking on SJTs is the

potential impact on criterion validity that may impact actual selection (e.g. Birkeland et al., 2006; Donovan, 1999; Hough et al., 1990; Ones & Viswesvaran, 1998). Lab studies (e.g. Hooper, Jackson, & Motowidlo, 2004; Nguyen, McDaniel, & Biderman, 2002; Peeters & Lievens, 2005) and applied studies (e.g. Reynolds, Winter, & Scott, 1999; Schmitt & Wolf, 2003; Weekly, Ployhart, & Harold, 2003) examine faking and SJT scores providing mixed results about the potential impact, if any, from faking in SJT scores.

Peeters and Lieven's (2005) research findings suggest that SJTs are susceptible to faking. Their subjects were college students (N=293) who were separated into *honest* and *intentionally fake* groups. They used a SJT based on situations around various student activities, such as studying, social responsibility, and integrity, and reported significance between faking and SJT results (Peeters & Lievens, 2005). Studies, such as Peeters and Lievens (2005) that manipulate actual faking versus coached honest responses are helpful to inform research and literature but may not be generalizable to applied settings with actual applicants, especially in high stakes settings (e.g. Bing, Kluemper, Davison, Taylor, & Novicevic, 2011; Ellingson, Sackett, & Connelly, 2007; McDaniel & Nguyen, 2001; McFarland & Ryan, 2000).

Faking countermeasures. Researchers have recognized faking as a threat to validity, especially in high stakes testing, and require efforts to reduce faking's impact (e.g. Bing et al., 2011; Ellingson et al., 2007; McDaniel & Nguyen, 2001;

McDaniel & Nguyen, 2001; Peeters & Lievens, 2005). Literature related to countermeasure includes behavioral uniform response options that present behavioral item responses on a continuum varying slightly between response selections (Ployhart & Ryan, 2000; Weekley, Ployhart, & Holtz, 2006); response time variations in faked responses (Holden, Kroner, Fekken, & Popham, 1992); item responses design utilizing higher cognitively loaded items with low transparency (McDaniel & Nguyen, 2001); and false consensus SJT item construction (Oostrom, Köbis, Ronay, & Cremers, 2017).

Of particular potential relevance to this study are response instructions. Ployhart and Ehrhart (2003) demonstrated SJT construct validity can be impacted by the type of instructions. There are essentially two common methods deployed in SJTs (Oostrom et al. (2107). The first are the behavioral tendency “would do” instructions and the second is the knowledge-based “should do” schema (McDaniel & Nguyen, 2001).

McDaniel, Hartman, Whetzel and Grubb (2007) argue that faking can be controlled, to some degree, with careful response instruction design. In the McDaniel et al. (2007) study, response instructions were classified into the groupings of knowledge-based and behavioral tendency formats. Their study utilized response design compared knowledge-based formats, “what is the best response” versus behavioral tendency, such as “what would you do?” McDaniel et al.’s (2007) findings support use of the knowledge-based design which has been

supported by other studies (e.g. Lievens, Sackett, & Buyse, 2009; McDaniel & Nguyen, 2001; Nguyen, Biderman, & McDaniel, 2005).

Recent research by Oostrom et al. (2017) also explores alternative SJT item response instructions to explore the potential of faking prevention. Oostrom et al. (2017) used new instructions based on the false consensus (FC) effect (Ross, Green, & House, 1977) that incorporates response items related to what “other” people should do as opposed to what would “you” do in situational work-related dilemmas. Oostrom et al.’s (2107) experiment adds support to McDaniel et al.’s (2007) assertion that “should-do” instructions tap knowledge opposed to the “would do” instructions tapping of attitudes and personality.

The “would do” responses relate to work by Ones and Viswesvaran (1998) exploring social desirability. They contend that social desirability faking is not a danger to criterion related prediction in personnel selection. Ones and Viswesvaran (1998) contend that, for selection, the candidate knows the “right” or “acceptable” social or behavioral response regardless if the candidate would actually behave in the performance environment or not. An assumption of the Oostrom et al. (2107) FC SJT is this style of item response explores if a candidate knows the appropriate course of action in the work dilemma. Oostrom et al. (2107) contend that their FC SJT response model may be particularly useful when exploring constructs susceptible for socially acceptable faking, such as integrity or dark-side traits (Alliger & Dwight, 2000; Becker, 1998) by making the item more complex in what

an “other” should do as opposed to oneself. According to Alliger and Dwight (2000) and Becker (1998) the FC approach offers an alternative to SJT instruction especially in testing situations where candidates perceive an advantage in responding with socially desirable responses.

Impact to assumptions and boundary conditions for this research. The study agency utilizes a video-based SJT where items are based on police situations and a behavioral “would do” response schema from an industry vendor. While “should do” responses may be better responses for knowledge-based exams (McDaniel et al., 2007) in reducing the likelihood of faking, the reality is the police organizations must use “should do” items due to the combination of applicants with and without any police training or experience. Deploying a knowledge-based item response would significantly disadvantage applicants who have yet to attend the police academy let alone have any policing experience.

The assumption for this research, related to the prevalence of faking, was that any faking falls within the concept of the social desirability type of potential candidate manipulation. It was further assumed that faking resulted in more positive scores indicating that the candidate recognizes the correct “would do” response in the scenarios regardless if their actual behavior and performance would conform, in keeping with assumptions of Ones and Viswesvaran (1998) and Oostrom et al. (2107).

Police Field Training and Evaluation Programs (FTEP).

The evolution of psychological services for policing, including assessment, can be traced to psychology's expansion beyond its original origins, in order to help inform selection decisions, including intelligence officers and spies, during World War II (Janik, 1994; Weiss & Inwald, 2018). Social unrest in the 1960's sparked events where police response and reaction were critically scrutinized by the media, government, and the courts (Weiss & Inwald, 2018). Two federal reports in 1967 initiated psychology's role in police selection (Janik, 1994). The first, the Presidential commission on Law Enforcement and the Administration of Justice, recognized the need for agencies to screen for emotional stability. Second, The National Advisory Council on Criminal Justice Standards and Goals recommended that all police agencies utilize psychological testing professionals to disqualify police candidates detrimental to policing by 1975 (Janik, 1994; Weiss & Inwald, 2018; White, 2008; Walker & Katz, 2002).

The profession accepted these recommendations with skepticism but there were some successes, such as the adoption of in-house police psychologists in places such as King County (WA) Sheriff's Office and the Los Angeles Police Department (Walker & Katz, 2002; Weiss & Inwald, 2018). To aid the adoption of professional psychological interventions, the U.S. Department of Justice founded a new initiative in the Law Enforcement Assistance Administration (LEAA) in 1968. The LEAA was instrumental to fund programs in police psychological testing

and grants exploring predictive police behavior to inform psychological constructs related to policing (Walker & Katz, 2002; Weiss & Inwald, 2018).

History of police field training and evaluation programs. There were no standards for police training after candidates met their state's basic training requirements prior to the 1970's (Walker, 2005). Recent academy graduates, or newly hired officers, in states with no formal basic training, faced sink or swim situations. Common practices were basic skills training, skill proficiency checklists, and performance review, whether formal or informal, with simple outcomes of either termination or continued employment (Fischer, 2015; Walker & Katz, 2002).

Prior to 1971 there was no formal field training for new police officers, until The San Jose model was developed by San Jose police lieutenant, Robert Allen, after officer involved incidents highlighted the need for new officer training. (Walker & Katz, 2002). The San Jose police chief had recently hired a director of psychological services, Dr. Michael Roberts, to assist in the formal development of the police selection and assessment programs (Weiss & Inwald, 2018). Dr. Roberts and a team of officers, some with advanced degrees, worked in the early years to identify specific behavioral traits necessary for success in policing. The San Jose Police Department initiative resulted in a set of 31 behavioral traits that lead to the daily observation reports (DORs) still in use today (Walker, 2005), as well as the first standardized field training program. The San Jose Model was recognized in 1973 by the International Association of Chiefs of Police (IACP) with the

California state legislature adopting the San Jose Model as the standard for state police field training (Fischer, 2015; Walker, 2005).

The field training and evaluation programs (FTEP) are where new officers attain on-the-job training with a field training officer (FTO). Haberfield (2002) discusses the FTEP experience as resembling an apprenticeship period. Practitioner and academic support for the theory, practice, and necessity of this critical training process is widespread (FDLE, 2011; Fischer, 2015; Lord, 1998; Sanders, 2003; Tyler & Makenzie, 2014; Walker, 2005; Walker & Kat, 2002; Wright, Dai, & Greenbeck, 2011).

Today's FTEP. Today, The San Jose model of police training is most popular in the U.S. (Caro, 2011; Chappell, 2007; COPS Office, 2001; Fischer, 2015; Kaminski, 2002; Scott, 2000; Scott, 2010; Walker, 2005; Walker & Katz, 2002). The standard Jan Jose model is a 14-16 week (MPCTC, 2017) training program (Fischer, 2015; Walker & Katz, 2002) that includes daily training, observation, and evaluation through the use of the daily observation report (DOR). Field training officers (FTOs) are officers with the collateral duty to supervise trainees, as well as having the unspoken role as key transmitters of the values of the organization (Chappell, 2007; Goldstein, 1987; Haberfield, 2002; Sander, 2003; Walker & Katz, 2002; White, 2008). FTOs are trained to facilitate training and to complete the DOR which contain both rating and narrative scales (Haberfield, 2002; Hoover, 2002)

The Maryland Police and Correctional Training Commission's (MPCTC) 2017 field training guide and the Florida Department of Law Enforcement's (FDLE) 2011 field training officer lesson plan augment the academic literature to inform this review. The rating scales vary somewhat, as the San Jose model is designed for customization (FDLE, 2011; Haberfield, 2002; MPCTC, 2017). Most DOR's are behaviorally anchored scales (Landy et al., 1976) that allow field training officers (FTOs) to rate critical police performance (Coutts & Schneider, 2004) using a 7-point Likert-style scale ranging from unacceptable (1) to superior (7) in an effort to improve FTO ratings across trainers and trainees. (San Jose, 2018; Walker, 2005). Some San Jose modelled programs anchor the 7-point Likert scale from "4" being the standard of a solo level officer (Caro, 2011; FDLE, 2011; MPCTC, 2017).

The length and structure of the San Jose FTEP model is broken down into three four-week phases (1-3) concluding with a two-week checkoff phase (FDLE, 2011; Haberfield, 2002; MPCTC, 2017). FTOs rate performance on the KSAOs in the DOR in both scale and narrative formats (Chappell, 2007) and make a final recommendation if the trainee should advance to the next phase of training. New officers are expected to handle 10 percent of the overall police work in phase one, fifty percent in phase two, and 90 percent in phase three. New officers are supervised by a different FTO in each of the first three phases. This rotation is intended to offer trainees different insights and various styles of policing (Cox,

1996; Tyler & Makenzie, 2014; Walker, 2005). The final two-week checkoff phase returns the trainee with their phase one FTO. This provides the phase 1/check-off FTO the perspective of skill evolution from phase one, where the trainee was accountable for only 10 percent of the field work, to checkoff where the trainee is expected to work at a “solo level” handling 100 percent of the duties (FDLE, 2011). FTOs in checkoff will often wear civilian clothes, as opposed to a police uniform, during this final phase and only observe and report unless a call requires the FTO to assist or intervene (McC Campbell, 1986).

When performance is not in keeping with the expected levels at each phase, the FTO can opt to do individualized feedback and training to remedy the deficiency (FDLE, 2011; MPCTC, 2017). If a trainee is not responding to training (NRT) then training is paused, usually up to two weeks, for more formal efforts in a remedial training effort (FDLE, 2011; MPCTC, 2017). Trainees positively responding to remedial training with improved and acceptable performance return to field training. Those who are unable to correct performance deficiencies are unusually terminated, based in failure to complete probation, once they have exhausted the typical two-week remedial training policy (FDLE, 2011; MPCTC, 2017).

San Jose alternative models. Public sector selection, hiring, and training processes lag well behind their private sector counterparts (Jordan, Lindsay, & Schraeder, 2012). Burkhart (1980) discussed the need, at that time, for research and

evolution of police job performance criteria. More contemporary literature echoes this need (Caro, 2011; Walker, 2005; Fischer, 2015; Walker, 2005). These sentiments have led to alternative, or derivative, models to the San Jose FTEP. While the study agency utilizes the San Jose model, it is relevant to briefly discuss the alternatives. These include the Reno model, the Mesa model, and the Houston model of field training (MPCTC, 2017).

The most promoted alternative to the San Jose FTEP model today is the Reno model. In the late 1990's the U.S. Department of Justice was focused on promoting the paradigm shift in U.S. policing from proactive policing to community-oriented policing (e.g. Walker, 2015; Warner, 2010) Through their COPS office, grant funding was provided to the Reno (NV) Police Department to develop a new field training program as an alternative model utilizing problem-based learning (PBL) (MPCTC, 2017; Walker, 2005; Warner, 2010). PBL is a learning focused based method of instruction originating in medical schools (Walker, 2005). It is described as a learning model where students confront contextualized and ill-structured problems with a goal to arrive at meaningful solutions (Rhem, 1998).

Transition issues from the basic police academy to field training. Positive effects of quality basic academy training can wane once officers are exposed to the realities of field work (Harr, 2001; Mastrofski & Ritti, 1995). Harr (2001) researched new officers over four periods extending past one year with findings

that suggest the first-year shapes officer attitudes. Chappell (2007) argues FTEP is critical for shaping new officer attitudes. Myers (2004) considers the FTEP period the most profound learning experience for new officers. Adding that not only are the critical skills are transferred from the academics of the police academy, but that the new officer is exposed to the organizational culture which is critical in order to understand how each agency does things (Myers, 2004).

Caro (2011) examined police academy performance compared to field performance, finding that the academy only accounts for more than 10 percent of the variance in field performance. This finding did not surprise Caro (2011), pointing to literature that discusses the intent of the basic academy to not only provide initial basic orientation and training of the job but to also disqualify those lacking P-J fit (Buerger, 2008; Lough & Von Treuer, 2013; White, 2008) and to start the orientation into the culture of policing (Walker & Katz, 2002).

Legal issues and failure to train. Agencies must be mindful of their responsibility to train (Hoover, 2002; McCampbell, 1987; Walker, 2005). Failure to train litigation historically comes from Title 42 U.S.C. 1983 which provides remedies for violations of civil rights. Failure to train is a legal concept supported by *Monell v. New York City Department of Social Services* (1978). Clarity was added, for both plaintiffs and cities, with the ruling from *Canton v. Harris* (1989) where the failure-to-train amounted to deliberate indifference to the rights of an individual. Plaintiffs seeking to use Title 42 U.S.C. 1983, must demonstrate that the

training programs are inadequate for officer's tasks, that the inadequacy resulted in the city's deliberate indifference, and that the inadequacy is closely related or the cause of the plaintiff's injury (*Johnson v. Cincinnati*, 1999).

The courts do not expect perfection in training outcomes. In fact, the majority of the Section 1983 cases successfully won involved multiple incidents. Kaminski (2002) explored liability related to police field training programs and found less than four percent were successful. A review of Texas cases surveying 849 agencies resulted in a report of 576 lawsuits, with only 34 directly related to improper training (Vaughn, Cooper, & Del Carmen, 2001). Hoover (2002) contends it is not the model an agency chooses that matters but rather the implementation and oversight of the training.

Impact to assumptions and boundary conditions for this research. The assumptions for this research were associated with and were supported from the section related to field training and evaluation, which includes that the organization in the study deploys common theoretical and technical frameworks, in the form of the San Jose FTEP model (FDLE, 2011). All Florida agencies share state guidelines related to field training that allow for generalization in aspects related to field training policies and practice. The study agency's field training program uses the popular San Jose model (Caro, 2011; Chappell, 2007; COPS Office, 2001; Fischer, 2015; Kaminski, 2002; Scott, 2000; Scott, 2010; Walker, 2005; Walker & Katz, 2002).

Transfer of Training

Salas, Tannenbaum, Kraiger, and Smith-Jentsch (2012) contend organizations may be wasting resources on training. Their key argument is that the billions of dollars spent on training annually in the U.S. do not equal a return on investment due, in large part, to training ineffectiveness. Salas et al. (2012) argue training ineffectiveness is rooted in the gap between science and application. This gap results in organizations missing out on potentially valuable data-derived methods in the design, delivery, and implementation of training (Baldwin & Ford, 1988; Blume, Ford, Baldwin, & Huang, 2010; Buerger, 2010; Bradley & Nixon, 2009; Campbell, 1971; Chappell, 2007; Goldstein & Ford, 2002; Grossman Burke-Smalley, 2018; Salas et al., 2012; Sleezer, Russ-Eft, & Gupta, 2015; Walker, 2005; Wexley & Baldwin, 1986).

Baldwin, Ford, and Blume (2017) have reviewed 30 years of research related to the positive transfer of training. They point out most research has been geared at exploring transfer resulting in individual performance. Goldstein and Ford (2002) stress the nexus between the transfer of training into actualized performance is the *paramount goal* of any organization's training function. However, Burke and Baldwin (1999) suggest that a large part of organizational training budgets do not lead to positive transfer of knowledge and, therefore, lower job performance.

Blume et al.'s (2010) metaanalysis on transfer of training explored a host of

predictive factors, concluding that the identification of a “magic bullet” from research to inform practitioners remains elusive. Research on transfer is extensive and has produced some significant predictors across studies, such as cognitive ability, conscientiousness, motivation, and supervisor support, but few prove consistently strong across job domains and organizations (Grossman & Salas, 2011; Saks, Salas, & Lewis, 2014).

Organizations are becoming more attuned to training outcomes with higher expectations on the return on investment (Aguinis & Kraiger, 2009; Becker & Huselid, 2006; Chuang & Liao, 2010) and the focus is expanding from individual performance to include the impact on overall organizational performance (Arthur, Bennett, Edens, & Bell, 2003; Birdi, Clegg, Patterson, Robinson, Stride, Wall, & Wood, 2008; Blume et al., 2010; Taylor, Russ-Eft, & Chan, 2005; Tharenou, Saks, & Moore, 2007).

Broad overview of transfer of training literature. Transfer of training can only occur after a learning experience (Baldwin, Ford, & Blume, 2009). In the case of the study population, new police officers have completed the state (Florida) basic academy training and passed the state certification exam (FDLE, 2011). New officers, considered trainees, then transition to their organizations and, after some level of organizational orientation, begin on-the-job-training in the form of the FTEP.

Facteau, Dobbins, Russell, Ladd, and Kudisch (1995) offer a good

foundation toward understanding transfer of training with their assertion about the necessary requirements. According to Fecteau et al. (1995) transfer of training only occurs when trainees first believe they have the capability to learn, followed with an expectation that the effort expended will impact performance, and lastly, that the results of the effort in training have valued outcomes. The Fecteau et al. (1995) transfer of training requirements are also consistent with Vroom's (1965) *expectancy theory* that individuals are more likely to engage in activities they perceive as achievable and the fulfill an individual need (Bretz & Judge, 1994; Nadler & Lawler, 1977; Van Eerde & Thierry, 1996).

These notions align with Bandura's (1986) *social cognitive theory* that contends self-efficacy is critical for trainees to effectively apply acquired skills, from training initiatives, to job performance. Bandura (1995) expanded his discussion to *self-efficacy theory* that asserts trainees must make a choice to expend the energy required to overcome task obstacles effectively. Self-efficacy beliefs are critical in how trainees perceive their ability, motivations, and behaviors (i.e. Bandura, 1995; Johnsen, Espevik, Saus, Sanden, Olsen, & Hystad, 2017; Stajcovic & Luthans, 1998; Tai, 2006; Williams, 2010).

Baldwin and Ford's (1988) comprehensive review of research and literature on transfer of training, at that time, recognized the research themes in the literature regarding challenges of organizations toward transfer of training into job performance. Even then, arguments suggested a significant research and

practitioner disconnect toward scientifically supported initiatives and interventions. (Campbell, 1971; Hatch & Dyer, 2004; Goldstein & Ford, 2002; Wexley & Baldwin, 1986).

Baldwin, Ford, and Blume's 2009 follow-up to the 1988 review included 140 (of about 300) interdisciplinary journal articles on transfer of training. Their review decisions were more precise being based on *intervention-based studies* and sorted by the *research design*. Their findings recognized advances in research focus on complex and authentic training contexts going well beyond earlier research focused on simple motor skills training. They also noticed research was exploring *interventions* in training for better transfer. Their third identified trend was a new research focus on *pre and post training* influences of transfer. Lastly, they found an increasing variety of research designs and measures regarding the time-intervals of observation toward evaluation of transfer.

Relevant transfer or training literature to study. Literature related to transfer of training is both broad and deep. Nuances based on type, timing, complexity, and organizational complexities make generalization challenging. There exist, however, specific themes in the transfer of training literature that may specifically inform this research and offer potential fodder for prospective future research and application based on the results of the research.

On-the-job-training and transfer of training. Saks and Burke-Smalley (2014) explore transfer of training and firm level (macro) performance. They

surveyed training practitioners about their methods (on-the-job-training, classroom, or computer), finding empirical support that on-the-job-training is most strongly related to organization performance. Salas et al. (2012) adds that informal training, such as the on-the-job-training in the police population, as in this study, builds on tacit knowledge. Blume et al.'s (2010) contention that original learning, such as the basic police academy in this research population case, is rarely enough to render training effective.

Framing. Quinones (1995) explored potential perceptions of training based on labels of *remedial* or *advanced* training. Their findings suggest that the framing of training can be important to trainees. Specifically, how training was framed impacted participant perceptions related to fairness and willingness to learn (Quinones, 1995). Tai (2006) studied supervisor frame training and found that framing predicts trainee motivation and self-efficacy. Tai's (2006) findings also suggest that subordinate motivation to learn is also positively impacted by supervisor's positive framing. Motivation toward training has strong empirical support, demonstrating positive effects on learning and training transfer (Bhatti & Kaur, 2010; Colquitt, LePine, & Noe, 2000; Grohmann, Beller, & Kauffeld, 2014; Huang, Blume, & Ford, 2015; Mathieu, Tannenbaum, & Salas, 1992; Shantz & Latham, 2012; Warr, Allan, & Birdi, 1999; Weissbein, Huang, Ford, & Schmidt, 2011). The notion that framing and the potential impact of improved trainee motivation may have a nexus relevant for the policing environment context.

Specifically, trainee errors in the field are considered critical failures. These trainee errors result in halting of on-the-job-training in FTEP and are hence labelled as *remedial training* with immediately recognizable consequences to trainees. FTO framing may play part in trainee motivation.

Error response, prevention training, and leader-member-exchange.

Training interventions regarding errors are important considerations in the context of the research as it is highly relevant to correction or prevention of critical errors (Baldwin, Ford, & Blume, 2009) common to the job domain of policing, in general, let alone during high stakes field training. Heimbeck, Frese, Sonnentag, & Keith (2003) explored potential impacts on transfer of training by allowing trainees to make errors versus receiving training aimed at preventing errors. Results suggest that allowing trainees the opportunity to fail, labeled as error, followed by a training intervention facilitated better transfer. Additionally, a training program that includes trainee devised error management plans, along with allowing for the pre-training error, resulted in longer lasting and more impact outcomes.

Gully, Payne, Koles, and Whiteman (2002) explored error training effectiveness in decision-making scenarios. Gully et al. (2002) when controlling for cognitive ability, conscientiousness, and openness to experience found that error training effectiveness is dependent on both cognitive ability and disposition. Trainees scoring higher in cognitive ability and openness to experience benefit

most from post error training. Those higher in conscientiousness did struggle somewhat, as the encouragement by supervisors to make errors negatively impacted self-efficacy (Gully et al., 2002).

New police officer training is very dynamic and often relies on real-time events as they unfold. The quality of the trainee/supervisor relationship cannot be ignored as it has direct training transfer implications (Bates 2003; Fagan, 2017; Ng, 2014; Tziner, Haccoun, & Kadish, 1991; Lewin, Eddy, & Tannenbaum, 2003) and is consistent with leadership and organizational behavior theories of leader-member-exchange (LMX) theory. Various meta-analytic studies have explored various antecedents (expectations, perceived justice, personality traits, perceived similarity, and trust) and consequences (such as commitment, performance, organization citizenship behavior, turnover, satisfaction, and role conflict) of LMX (Wayne, Shore, & Lien, 1997; Wayne, Shore, Bommer, & Tetrick, 2002; Dulebohn, Bommer, Liden, Brouer, & Ferris, 2012). Brunetto, Teo, Shackloack, Farr-Whetton, and Shriberg (2017) explored the relationship of LMX and police training finding support for the LMX accounting for one-fifth of training outcomes and one-third of office training satisfaction.

Police trainee supervisors must balance safety and training efficacy based on the potential consequence of allowing a trainee to make an error. Routine administrative tasks, such as report writing, radio etiquette, or timeliness of non-critical knowledge recall, allow for errors and mistakes to spark an individualized,

or group, lesson. On-the-job training performance related to officer safety, arrest, and other first response life and death situations require the training supervisor to make immediate judgments and corrections (FDLE, 2011).

This critical training supervisor role involving the timeliness and detail of feedback and coaching also have support in literature as significant predictors of training transfer (Burke & Baldwin, 1999; Tannenbaum, Mathieu, Salas, & Cannon-Bowers, 1991, Smith-Jentsch, Salas, & Brannick, 2001; Xiao, 1996)

Behavioral modeling. An aspect of the on-the-job nature of police FTEP is the direct and constant opportunity for trainees to observe behaviors modelled by their supervisor and other officers they work alongside. Taylor, Russ-Eft, and Chan's (2005) meta-analysis on behavioral modeling training and the impact on job behavior suggests that the combination of both positive and negative modeling best facilitates training transfer. Their finding also supports the idea of allowing trainees to develop training scenarios, thus capturing problems or situations they perceive as challenging. Trainee derived scenarios were more impactful when the training had specific goal settings (Wexley & Baldwin, 1986) for the training outcome, trainee supervisors participated in training, and there was immediacy related to rewards or punishments which occurred in work environment (Taylor et al., 2005). It is therefore wise that police training supervisors and organization training units recognize the potential value in capitalizing on debriefs from on-the-job-training. Additionally, many police organizations have specific non-field

training days where scenarios are often used to facilitate training, especially high-liability training such as use of force, police-citizen conflict, and active shooter training.

Baldwin's (1992) modeling research offers additional research with potential utility in this research domain. Baldwin (1992) examined negative and positive modeling in a behavioral-based assertive communication training program. Findings support modeling, in training scenarios, that combines positive and negative examples, improved training retention, and better generalization of the new skill. This notion is consistent with research by Hitt, Bierman, Shimizu, and Kochhar (2001) who suggests that trainees provided opportunities to problem solve experience better learning outcomes.

Training timing: Pre and post training interventions. Saks and Belcourt (2006) investigated the timing of training activities and transfer of training. They explore training activities that occurred before, during, and after the training and found support suggesting stronger transfer in the work environment occurring before or after the training. This has implications for the research environment as the nature of field training provides both pre and post micro-training around work specific events.

Post-training interventions have also shown promise (Baldwin, Ford, & Blume, 2009; Gaudine & Saks, 2004; Gist, Stevens, & Bavetta, 1991; May & Kahnweiler, 2000; Noe, Sears, & Fullenkamp, 1990; Wexley & Baldwin, 1986).

Gist et al. (1991) explored goal setting and complex interpersonal skill training. Goal setting was explored as either traditional goal setting (Locke & Latham, 1990) or goal setting with a self-management intervention plan. The findings support added value from inclusion of the self-management aspects resulting in improved skill generalization and long-term performance (Gist et al., 1991).

Walker (2005) discusses the promise of daily activity journaling incorporated into the Reno field training model. Shantz and Latham (2012) add to the notion of self-managed interventions in suggesting trainees who write self-guides, as a prescribed training tool, demonstrate improved trainee reported self-efficacy. Grossman and Burke-Smalley (2018) challenge future researchers to further investigate the fledgling aspects of such self-management tools as potential robust strategies toward more effective transfer, especially for complex skill training.

Police academy to field training transfer. Caro (2011) explored basic police academy performance and field performance during FTEP for new state police officers (N=187). His findings suggest that basic academy performance explains only up to 10 percent of the variance in field performance. Caro (2011) asserts his findings align with Buerger's (1998) contention that the police academy training focuses on the basic policing concepts and skills, with a disconnect to actually implementing these various skills in applied settings. Caro's (2011) finding is also consistent with controversial estimates by Georgenson (1982) that

10 percent of training is transferred. Georgenson's (1982) proposition to the limit to transfer was discussed and advanced by others (i.e. Baldwin & Ford, 1988; Clark, 2010; Fitzpatrick, 2001).

Farrington (2011) took specific issue with the Georgenson's 10 percent citing research has supported transfer of training estimates at various rates (Anderson, Reder, & Simon, 1996; Bassok & Holyoak, 1989; Singley & Anderson, 1989), but adds that a generalizable transfer of training estimate, across all industries and sectors, is unlikely, a notion with additional academic support (Blume, Ford, Baldwin, & Huang, 2010; Taylor, Russ-Eft, & Taylor, 2009). While there is ample debate to the topic, Caro's (2011) study offers support for Georgenson's (1982) proposition, at least for Caro's sample of police officers.

FTEP as a learning transfer system. The applied relevance of transfer of training is that trainees actualize training efforts into the pro-work behaviors associated with both individual and work performance. Hatch and Dyer (2004) discussed that new organizational members are key assets, in any organization, toward meeting the organizational mission. They assert that it is critical for organizations to facilitate self-efficacy toward autonomous deployment of the critical KSAOs to contribute to the organization.

Holton, Bates, and Ruona (2000) explored the training program and environmental design that impact the transfer of training. Their research suggests the actual training transfer system must be recognized and exploited to facilitate

better outcomes. They further articulate a transfer system as factors, including individuals, training events, and organizational attributes, that influence training transfer to actual performance. Caro (2011) contends that police organizations must deliberately concern themselves with creating a supportive environment for FTEP.

This study capitalized on the Holton et al (2000) framework should results better identify data of training transfer failures. These data provided diagnostic information with the potential to inform trainers and the organization where, and when, potential interventions may positively impact transfer.

Impact to assumptions and boundary conditions for this research. The assumptions for this research were supported from the transfer of training literature review are important to inform how selection scores may predict involuntary turnover. Transfer of training in this context is that nexus as it potentially mediates the relationship between selection and turnover. Additionally, this part of the literature suggested that new officers, fresh from the academy, begin field training with much to learn about applying the skills in the field that were acquired in the sterile context of the academy (Buerger, 1998; Harr, 2001; Mastrofski & Ritti, 1995). The assumption is the focus of research by Caro (2011) study that only 10 percent of field training variance is explained by police academy performance. Caro (2011) found theoretical support from these literature (e.g. Baldwin & Ford, 1988; Clark, 2010; Fitzpatrick, 2001; Georgenson, 1982). Another assumption

informing this research came from research that links cognitive ability to transfer of training (Grossman & Salas, 2011; Gully, et al., 2002; Maltarich et al., 2010; Saks et al. 2014; Van Iddekinge et al., 2011).

Turnover Research

Hom, Lee, Shaw, and Hausknecht (2017) reviewed a century of turnover research and stressed that research with the potential to inform practitioners is severely limited. A growing body of literature supports the notion of the potential utility of good turnover data to screen out candidates that can be predicted to have short tenure (Hutchinson, Villalobos, & Beruvides, 1997; Hom et al., 2017; Mabon, 1994; Ployhart & Weekley, 2010). Ployhart (2006) ponders the neglected question of turnover's relationship with recruiting. Ployhart (2006) also poses questions related to management failures to deploy research science, wondering if the results are too convoluted or vague for application in organizations. In the case of selection and turnover, the answer is simply that the literature is barely existent (Hom et al., 2017).

Part of the applied research void is discussed by Hom et al. (2017) where they postulate that turnover research, at the individual level, likely cannot account for the collective selection processes to help inform collective turnover. Additionally, most turnover research focuses on prevention of turnover (Hom et al. 2017) as opposed to actual turnover. Prediction of turnover is scarce in the literature aside from three identified studies (Hom et al., 2017). The first by

Barrick and Zimmerman (2009) studied selection tests and demonstrated a nexus with predicting turnover. Their findings suggest that personality traits, as measured during selection batteries, did have prediction for turnover. The second study is by Maltarich, Nyberg, and Reilly (2010), whose findings support a relationship between cognitive ability and turnover. Maltarich et al. (2010) found that cognitive ability impacts P-O fit in cognitively demanding jobs, thus mediating turnover. Lastly, Van Iddekinge, Roth, Putka, and Lanivich (2011) conducted a meta-analysis and found that construct-focused interest scales linked employee interest and organizational characteristics, and thus produced improved tenure when interests and organizational characteristics aligned.

History of turnover research. Hom et al.'s (2017)'s review of turnover research attributes the first published article to Marion A. Bills (1925) where he explored potential predictive explanations on why clerical staff was leaving. Later, the 1958 model by March and Simon initially conceptualized movement desirability and ease as the cornerstone for their turnover model (Hom et al., 2017). Mobley, Griffeth, Hand, and Meglino (1979) are credited with launching the content model to inform broad conceptual notions as to why people quit. Mobley et al.'s (1979) model informed new models and methods for turnover research significantly (Hom et al. (2017). Mobley's (1977) process model is based on the notion that dissatisfaction leads to turnover. The path theorized by Mobley is sequential, starting with dissatisfaction, leading to thinking about quitting,

evaluating one's subjective expected utility (SEU) related to consequences of quitting, followed by search intentions, alternative job evaluation, comparisons to present job, intentions to quit, and finally leaving (Mobley et al., 1979; Mom et al., 2017).

Voluntary and involuntary turnover. Turnover is traditionally categorized as voluntary or involuntary (Price, 1977). Mobley et al. (1979) discuss voluntary turnover as employee self-initiated separation from the organization. McElroy, Morrow, and Rude (2001) consider involuntary turnover as organizationally avoidable, and a function of poor individual performance or insubordination resulting in dismissal. Dalton et al. (1982) assessed that attempts at reducing turnover focus mainly on voluntary turnover, which is considered preventable.

Porter contributions to turnover theory. Lyman Porter is considered a seminal contributor to turnover research (Hom et al., 2017). Porter and Steers (1973) extended turnover research by exploring the nexus with Vroom's (1965) expectancy theory, thus shaping theory in job satisfaction (Hom et al., 2017). Lyman was later instrumental in multiple contributions, including the notion of functional turnover. He asserted that the cost of extra, underperforming, or costly employees can actually benefit organizations (Dalton, Krackhardt, & Porter, 1981). Porter was a pioneering force in organizational commitment (Porter, Crampon, & Smith, 1976). Krackhardt and Porter (1985, 1986) explored social networks to explore how interconnected networks of employees can impact

turnover.

Function, dysfunctional, and optimal turnover. The notion that not all turnover is dysfunctional (Dalton, Tudor, & Krackhardt, 1982) has informed additional research. Dalton et al. (1982) further operationalize dysfunctional turnover as occurring when an employee wants to leave the organization, but the organization would prefer they stay. On the other hand, functional turnover occurs when the employee wants to leave the organization and the organization does not mind them leaving (Dalton et al., 1982).

Dalton et al (1982) also added a dichotomy by adding the notion of functional turnover to dysfunctional turnover. Functional turnover can occur when a marginally performing or disgruntled employee leaves. This can actually have positive impacts on the remaining members. Functional turnover can reduce homogeneity among members (Schaubroeck et al., 1998), which can result in negative organizational performance in areas such as adaptation to change, groupthink, and innovation (Astakhova et al., 2015; Schneider et al., 1998).

Abelson and Baysinger (1984) further refine the notion of dysfunctional turnover. They contend that the term “dysfunctional” implies a condition of organization malfunction. This may not be the case. Optimal turnover, according to Abelson and Baysinger (1984), is organization turnover that balances organizational costs of turnover with the cost associated with reducing it in an economic construct or return on investment perspective. Thus, a certain level of

turnover, in terms of economy and performance, has optimum outcome potential. Employees that voluntarily separate when the organization would prefer to retain them is still dysfunctional in this optimum turnover model, but it does recognize organizational realities and value in natural turnover of members. After all, everyone leaves at some point (Ployhart & Weekely, 2010).

Dalessio (1994) explored insurance agent turnover and asserts that 74% of agents (N=677) not retained after one year was *functional turnover*, poor performers leaving, as opposed to *dysfunction turnover*, such as high-performing agents leaving. Dalessio (1994) further contends that any level of reductions of dysfunctional turnover can save organizational capital in time and financial costs.

The cost aspect related to Abelson and Baysinger's (1984) model is further advanced by research such as Williams and Livingstone (1994). They contend that the cost of turnover, such as separation pay, the cost to recruit replacements, and training, are not the only costs that organizations must consider. The true cost of turnover cannot be realistically assessed without exploring the impact of turnover on performance (Boudreau & Berger, 1985; Cascio, 1982; Williams & Livingstone, 1994).

Avoidable and unavoidable turnover. Dalton et al. (1981) also presented another taxonomy of turnover that considered it avoidable or unavoidable. Abelson researched the avoidable and unavoidable notion in his 1987 research. Abelson (1987) excluded employees leaving involuntarily and focused on those

who stay or leave for involuntary reasons. Involuntary reasons may include conditions of layoff, retirement, severe illness or injury, or death. Findings suggest employees who stay and those who leave involuntarily are not different. In one particular study, Barrick and Zimmer (2005) explored avoidable voluntary turnover. Their findings suggest that attitude and intentions, such as organizational commitment, predict turnover. Implications to this study are that voluntary turnover must be further considered beyond simply P-J or P-O fit.

Organizational commitment and turnover. Lyman Porter's influence is also visible in the construct development of organizational commitment. Porter, Crampon, and Smith (1976) contend that organizational commitment explains more turnover than job satisfaction. Hom et al. (2017) credit Porter and his colleagues for advancing organizational commitment well beyond its relationship to turnover. Porter's influence did not stop with organizational commitment (Hom et al., 2017).

Methodologic advances and distractions. Advances in the late 1980s and 1990's in methodological design, such as structural equation modeling, allowed for more robust and complicated models of turnover related constructs (Hom et al., 2017). The unintended consequence of the advance, according to Hom et al. (2007) was a shift away from seeking explanations of turnover toward explaining covariances among constructs that may lead to turnover. This did not help inform the field or advance precise research on turnover related constructs (Schmidt &

Hunter, 1998; Schmidt, Oh, & Shaffer, 2016).

The unfolding model of turnover. The incremental theoretical refinement in the 1990's eventually led to discussion and renewed innovative models such as Lee and Mitchell's (1994) unfolding model of turnover. This paradigm shift, from March and Simon's (1958) model, contends that job dissatisfaction is a paramount turnover cause, causing employees to seek and perhaps leave for perceived better job options, and that potential leavers compare potential jobs to their current job based on Mobley's (1977) notion of SEUs (subjective expected utility).

Lee and Mitchell (1994) introduced the concept of shock toward their broader model of turnover. Shocks are external jarring events that push or pull employees to contemplate leaving, driving alternative paths to more contemporary turnover models (Burton, Holtom, Sablinski, Mitchell, & Lee, 2010). Lee and Mitchell's (1994) unfolding model specifies four distinct paths toward turnover. Path one involves a shock activity activating a preexisting plan to leave. This would include the potential of unwanted turnover resulting from an employee's decision to stay at home and raise children.

Path two involves negative job shocks that violate the employee's values or goals that impact organizational commitment (Lee & Mitchell, 1994). This may take the form of a supervisor falsely accusing an employee of a prohibited act or fairness perceptions related to being passed over for a promotion. The third path involves the shock of unsolicited job offers where employees must weigh the

potential advantages of the new offer over the existing job regardless of their current satisfaction or commitment with their current position (Hom et al., 2017; Lee & Mitchell, 1994). Lastly, path four relates to the conventional models where dissatisfied employees leave once they secure a better job (Hom & Griffith, 1991; Hom et al., 2017; Lee & Mitchell, 1994).

Lee and Mitchell's (1994) unifying theory is the dominant theory informing research and practice today (Holstrom, Mitchell, Lee, and Eberly, 2008; Hom, 2011; Hom et al., 2017). Qualitative methods have been used to validate unifying theory models (Lee, Mitchell, Holtom, McDaniel, & Hill, 1999; Lee, Mitchell, Wise, & Fireman, 1996). Lee et al. (1996; 1999) have interviewed leavers to ascertain if any of their four models explained the decision to leave and found that the majority of leavers' decisions could be explained with one of the paths. Later research supports findings consistent with the qualitative investigations and suggest the shocks drive more turnover than dissatisfaction (Holstrom et al., 2008; Kammeyer-Mueller, Wanberg, Glomb, & Ahlburg, 2005). While the unfolding model is considered a significant advancement in turnover research (Hom et al., 2017) it is not without limitations. For example, the theory rests on the qualitative research, leaving it lacking robust testing with predictive designs (Hom, 2011).

Unfolding model with job embeddedness and turnover. Current unfolding, model-based scholarship continues today (Hom et al., 2017). Mitchell and Lee

(2001) added job embeddedness theory to their model. Their additive model is based on the notion that embeddedness factors can mitigate the effect of shocks by exerting a pull force for someone to stay with the organization (Burton et al., 2010). Mitchell et al.'s (2001) job embeddedness inclusion has, in part, reframed turnover from why people leave to why they stay. Research has suggested staying constructs including training opportunities (Hom et al., 2017), kinship (Price & Mueller, 1986), job sacrifices (Meyers & Allen, 1997), nonwork influences (Mobley et al., 1979), community embeddedness (Feldman, Ng, & Vogel, 2012), family embeddedness (Feldman et al., 2012; Mitchell et al., 2001), and occupational embeddedness (Feldman & Ng, 2007).

Maertz and Campion (2004) added to the notions of the unfolding model by adding elements related to how and why people leave. They identified four leaver decision types to explain motivations to leave. They labeled them as the *impulsive quitting* due to insufficient attachment, *comparison quitting* such as quitting for another job, *preplanned quitting* from an advance intention of “when” a trigger occurs, and *conditional quitting* to quit from an “if” an event happens (Maertz & Campion, 2004). The Maerta and Campion (2004) decision types do correspond with turnover paths but they are not identical (Hom et al., 2017).

Cost of Turnover. Barrick and Zimmer (2005) are among scholars that recognize the immense applied potential in better understanding the dynamic effects involved in employee turnover. Ployhart (2006) explores staffing issues for

organizations in the 21st century. He argues that current staffing issues present more complex issues for organizations, noting that challenges such as increased knowledge work, work place diversity, and competition for applicants, require more targeted recruiting and training initiatives.

Organizations must consider the cost of onboarding and training new employees. The average employee, regardless of industry or sector, has a learning curve before the new employee fully contributes to the organization. This learning curve takes even longer in complex roles, such as policing (Ployhart & Weekley, 2010).

Mabon (1994) also explores the costs associated with frequent turnover, from a tenure perspective, and points to the seemingly obvious organizational aspect of less turnover leading to fewer costs related to recruiting, selecting, and training new employees. He urges that tenure and lack of turnover, should be intentional goals of organizations' recruiting plans. The basic assumption posited by Mabon (1994) is simply that less turnover reduces the organizational cost associated with selecting replacements.

There is growing recognition in the literature that changing employment conditions related to issues such as labor market conditions, generational differences, norms related to shorter lengths of employment, and job mobility call for more deliberate research in understanding and predicting turnover (Chen & Chang, 2010; Dess & Shaw, 2001; Hom et al., 2017; Kacmar et al., 2006; Mabon,

1994; Ployhart, 2006; Ployhart & Weekley, 2010). Ployhart's (2006) exploration of staffing issues for organizations in the 21st century suggests that staffing issues present increasingly complex issues for organizations. Ployhart noted, in 2006, that challenges, such as increased knowledge work, work place diversity, and competition for applicants, would require more targeted recruiting, training, and retention initiatives. In the case of policing today, Mabon's (1994) cost of replacing employees assertion takes on added importance to police organizations where applicant pools are ever-shrinking among the host of previously discussed challenges (Cavanagh, 2003; Haarr, 2001; Koper, Maguire, & Moore, 2001; Orrick, 2008; Selden et al., 2013; Smith, 2016; Wilson, 2012; Woska, 2016).

Organizational Performance. Organization performance consequences to turnover have been recognized for decades (e.g. Abelson & Baysinger, 1984; Hom et al., 2017; Hutchinson et al., 1997; Maltarich, Nyberg, & Reilly, 2010; McElroy, Morrow, & Role, 2001; Mowday et al., 1982; Price, 1977; Shaw, 2011). Pierce (2013) states that historically studies focused on individual level effects such as who stayed, who left, and why good performers leave (Hom et al., 2017). Maltarich, Nyberg, and Reilly (2010) add that a great deal of literature on turnover is focused on individual performance but contend that voluntary turnover is important when looking at firm-level performance.

Meta-analysis by Hancock, Allen, Bosco, McDaniel, & Pierce (2013) suggests that various turnover effects associated with performance include

performance efficacies and costs, quality of service, and safety outcomes. Hancock et al.'s (2013) review also indicated various organizational aspects such as industry, type of turnover, and organizational size, moderates the effect. A research review by Hausknecht and Holwerda's (2013) of organizational turnover effects suggest the effect depends on who remains as well as those who leave.

More recent meta-analytic studies (e.g. Hancock, 2013; Heavey et al., 2013; Park & Shaw, 2013) found associations between turnover and organization performance but, according to Hom et al. (2017), lacked operational definitions and theoretical frameworks to draw generalizable conclusions. This situation has limited new theory development related to tenure. Organizational threats to performance from turnover, beyond the cost of replacement, have been explored in the literature (Hom et al., 2017).

The literature provides seemingly generalizable studies including work by Kacmar, Andrews, van Rooy, Steilberg, and Cerrone (2006). The explored supervisor and employee turnover in a fast-food setting discovered that turnover resulted in longer customer wait times, thus reducing customer satisfaction and profit. Call, Nyberg, Ployhart, and Weekley (2015) examined retail performance finding lower performing stores had higher turnover. Both of these examples seem to make sense. New employees do not have the skills and experience of those with more tenure. Teams can be impacted by tenure.

The impact of turnover on team performance can get complicated. For

example, Shaw, Duffy, Johnson, and Lockhart (2005) examined turnover among members who are central to workplace communication networks. Their findings suggest that when a key networking member leaves it disrupts not only communication but reduces overall organizational performance. Hausknecht et al. (2009) adds the additional team-level nuances related to turnover and performance. Hausknecht et al. (2009) explored customer service quality in smaller work units and found that tenure was related to positive customer service ratings. Additionally, these units could sustain performance as long as the proportion of new team members was small.

Impact to assumptions and boundary conditions for this research.

Ployhart (2006) noted that every organization uses some form of selection process without guarantee of effective, appropriate, or optimal use. He further contended that unless research gaps are addressed this trend will continue. The nexus between SJTs and tenure is seemingly unexplored. This research was a modest attempt to explore this potential nexus between selection and turnover suggested to exist in the literature (Artur et al., 2006, Barrick & Zimmerman, 2009; Huffcutt, 2011; Whetzel & McDaniel, 2009). Ryan and Ployhart (2013) have suggested research explore SJTs to P-O fit and turnover. This research attempted to explore this nexus.

This literature review informed this research by proving the background, rationale, and operational definitions informing the dependent variable of turnover.

For example, a key assumption, discussed fully in chapter three, was that both P-J and P-O fit (Maltarich et al., 2010) are antecedents to both voluntary and involuntary turnover.

Hom et al. (2017) suggested strengthened turnover literature when researchers better capture the context of studies exploring potential antecedents to turnover. Allen, Hancock, Vardaman, and McKee (2014) argued that research related to turnover should capture the context and boundary conditions for the antecedent, in the case selection scores, nexus to turnover. The boundary conditions related to turnover for this study included one southeast mid-sized police department with less than 250 sworn officers deploying a selection battery, including a SJT, and conforming to Florida standards of field training and evaluation. Full boundary conditions will be further outlined in chapter three.

Synthesis

The sections of this chapter provide a relevant review of the theoretical frameworks, problems in police turnover, relevant employment law history, police selection, SJTs, faking, transfer of training, and turnover. This study was supported by a literature review consisting in excess of 440 references including peer reviewed journals, books, practitioner publications, and government sources. Literature is synthesized in summary to provide concluding arguments on the merits of this research.

Current retention problem in policing. The importance of selection and

retention, in any organization, is well supported in literature (i.e. Mabon, 1994; Ployhart, 2006; Ployhart et al., 1996; Rubenstein et al., 2015). Recruiting and retaining employees is arguably more critical for governmental organizations charged with public safety (Barrett et al., 2010; Selden et al., 2013; The Center for State and Local Government Excellence, 2017; Wilson, 2014). Legal issues in selection, including avoiding concerns with adverse impact, make proper selection critical and necessary (Arnold, 2004; Barrick & Zimmerman, 2005; Campbell et al., 1993; Gottfredson, 1996; Gutman, 2009; McDaniel et al., 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Schmidt & Hunter, 1998; Van Iddekinge et al., 2011; Weekley & Jones, 1999).

Police agencies today face challenges stemming from financial challenges resulting from the 2008 recession (Smith, 2016; Wilson, 2014), reduced public perception (Barrett et al., 2010; Gold, 2015; Wolfe & Nix, 2016), job attractiveness disadvantages (Jordan et al., 2009; Kaminski, 1993; Orrick, 2008; Smith, 2016, Wilson, 2014; Woska, 2016), and generational differences (Maciag, 2018; Espinoza et al., 2010; Ployhart & Weekley, 2010). These are among current challenges facing police administrators in maintaining a competent police force charged with public safety and order.

Hom et al. (2017) contend that the nexus between selection and turnover is often overlooked. Ployhart (2006) contends this nexus is overlooked as a result of a common disconnect between human resources departments and operational

units. Human resources focus on the standards and guidelines of hiring, while operations focus on performance and fit (Ployhart, 2006). This divide can be amplified in organizations where the quest for a diverse force has been a challenge (Barrick & Mount, 2005; Fyfe & Kane, 2006; Ployhart, 2006). Training time and costs are also a significant investment (Hausknecht & Holwerda, 2013; Mabon, 1994; Ployhart & Weekley, 2010; Rubenstein et al., 2015; Seldon et al., 2013). The challenges that police agencies face resulted from an accumulation effect from diminished candidate pools, interagency competition, benefit and pay disparities to other professions, and complex social reactions and expectations, all of which are of significant concern to practitioners.

The cost of turnover in policing extends beyond simply the cost of selection, hiring, and training (Barrick & Zimmerman, 2009; Hom et al., 2017; Maltarich et al., 2010; Van Iddekinge et al., 2011). Not only are onboarding and training costs high in policing, the prospects of replacing someone due to turnover is ever challenging (Ployhart & Weekley, 2010).

Potential applied contribution. The current applied challenge of maintaining a stable and skilled police force is recognized as critical to community safety and quality of life. This research has the potential to inform practitioners on potential valuable data available in current selection processes that may have predictive utility for turnover. These existing data have the potential to identify both functional and dysfunctional turnover (Dalton et al., 1982) models. The

models that may segregate functional turnover, such as failure to transfer training or poor performance, from the dysfunctional turnover, such as a good officer leaving for another agency, may help inform selection processes that mitigate and reduce dysfunctional turnover as well as reduce functional turnover with informed prediction of behavioral characteristics that may be present in selection (Barrick & Zimmer, 2005; Maybon, 1994).

The prospect of using good turnover data to screen out candidates that can be predicted to have short tenure is encouraged in literature (i.e. Hutchinson et al., 1997; Hom et al., 2017; Mabon, 1994; Ployhart & Weekley, 2010). Hom et al. (2017) and Price (2011) suggest that data-driven initiatives may yield turnover data informing organizations as to candidates with a high chance of leaving.

Potential scholarly contribution. Aside from the 1994 Dalessio study exploring post-selection use of SJT and turnover with insurance agents, this literature review did not identify research exploring the potential nexus between selection SJTs and turnover. There is, however, literature suggesting that the research explore this potential nexus. Barrick and Zimmerman (2009) and Whetzel and McDaniel (2009) identified a lack of research exploring SJTs and turnover and pointed to the potential applied utility. Rubenstein et al. (2018) suggested that future turnover research include actual turnover as opposed to the commonly used *turnover intentions* resulting from surveys. This project also seeks to identify prediction criteria in selection research as posited by Van Iddekinge et al. (2011).

Turnover temporal order potential. The use of actual daily performance measurement, through DORs, offered measurement opportunities that Rubenstein et al. (2018) contended are lacking in turnover research. Rubenstein et al. (2018) suggested that turnover research is lacking exploration of temporal orders related to turnover as opposed to a single predictor at one point in time. Utilizing DORs offered an increase in the frequency of measurement to potential antecedent constructs related to turnover outcomes offering a linear temporal observation related to behaviors associated with turnover.

Transfer of training effectiveness. The National Research Council (2004) contended that a significant unresolved question in policing is the effectiveness of police training. They contended that rigorous research is necessary to help shape training, especially with problem-solving. Geller and Toch (1995) reviewed model policies and best practices regarding training and officer use of force and found, that despite literature recommending specifics of such training, there were no empirical validations of the claims. This research offered predictive analysis (logistic regression) of selection data on daily police performance during field training. This offered potential insight into the National Research Council's (2004) assertion that research exploring the nexus of KSAOs would be valuable to practitioners and police managers by helping inform improvements to police training. Caro (2011) specifically discussed the advantage of exploring police performance during field training, contending that it is a substantial method of

measuring performance that is lost after field training.

SJT and turnover nexus. Ryan and Ployhart (2014) contended that future SJT research should include SJT prediction into unexplored domains including turnover. Ployhart and Weekley (2006) indicated that most SJT research using field studies have been limited to correlational design. They suggested more experimental approaches to explore SJTs measure related to decision making and performance. This research deployed predictive analysis, well beyond correlational designs, to specifically address SJTs' prediction on turnover outcomes. This study demonstrated potential in addressing gaps as identified by Ployhart and Weekley (2006), Ryan and Ployhart (2014) and The National Research Council (2004).

Strategic intervention for turnover. Ployhart (2006) noted that every organization uses some form of selection process without guarantee of effective, appropriate, or even optimal utility. Ployhart (2006) argued that the trend will continue unless gaps are informed by research. Hausknecht and Holwerda (2013) suggested that turnover research deploying personnel tracking approaches may yield opportunities for interventions. This research attempted to specifically explore selection processes for additional utility toward optimal utility suggested by Ployhart (2006) while searching for potential foundational predictive research useful for innovative interventions as encouraged by Hausknecht and Holwerda (2013).

Micro level transfer of training insights. Baldwin et al. (2009) discussed the importance of transfer of training research to practitioners charged to train new members. Baldwin et al. (2009) contended that evidence-based training interventions from literature have the potential to inform application. Baldwin et al. (2009) and Hom et al. (2017) suggested that future research should deploy empirical observations of training designs, trainee readiness, and supervisor support related to training transfer in applied domains. This study explored police training strategies, including remedial training, and trainee turnover outcomes.

Based on the reviewed literature this research has potential to inform both applied and scholarly audiences. The applied problem is often exasperated, especially in smaller agencies, by a lack of resources. Exploring already existing selection data to potentially inform recruiting, selection, and training practices makes good sense. The literature on turnover, transfer of training, and SJTs is ripe with suggestions for more applied and context specific research of actual turnover. The potential for this study to inform some of these calls for research is exciting.

Theoretical Law Enforcement Turnover Model. The law enforcement turnover model (Figure 2) is conceptualized for the public sector industry of law enforcement to capture the robust and unique initial field training provided by the standardized FTEP. This model begins with the traditional notions of organizations selecting employees from an applicant pool by testing for job specific knowledge, skills, abilities, and other attributes. The “funnel” in the model visualizes the SJT

selection battery deployed to assess these KSAOs that are expected to predict performance. The selection process is reflected as a “funnel” in the law enforcement turnover model (Figure 2) recognizing these other factors as “?”.

Once selected, newly hired officers complete a department orientation period of a month before starting a fourteen-week field training and evaluation program (FTEP). Training of new probationary officers is conducted in the FTEP by state trained and certified field training officers (FTOs) who provide training and officer guidance (Fischer, 2015; Walker & Katz, 2002). Daily performance is observed, evaluated, and documented by FTOs through the use of daily observation reports (DORs), which are anchored scale performance ratings based on the expectations of a solo officer working alone. Figure 2 visualizes the transfer of training as documented in these DORs leading to various turnover and retention outcomes (FDLE, 2011; Haberfield, 2002; MPCTC, 2017).

The model also recognizes aspects related to P-J fit (Edwards, 1991) and P-O fit (Judge & Ferris, 1992) related to transfer of training (Baldwin & Ford, 1998) as well as shocks (Lee & Mitchell, 1994) accountable for turnover outcomes. These conceptual pathways lead to optimal turnover, functional turnover, and an “either-or” turnover that may fall between functional and dysfunctional turnover.

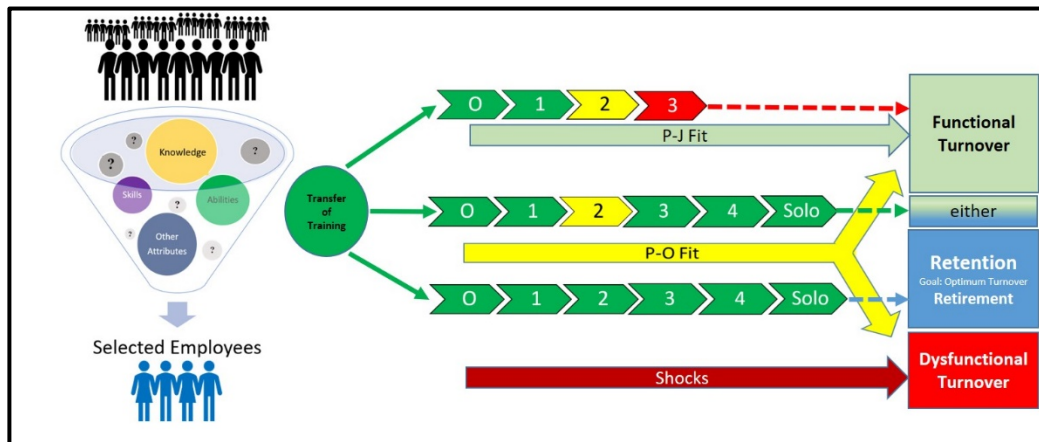


Figure 2. The Theoretical Law Enforcement Turnover Model. This figure illustrates the theory related to the impact of transfer of training on turnover or retention outcomes in police agencies.

Chapter 3

Methodology

Overview

Allen, Hancock, Vardaman, and McKee's (2014) review of 447 empirical studies related to turnover led them to criticize the lack of clear context in the studies. Allen et al.'s (2014) recommendations include arguments that future turnover related research better articulate and discuss the context of the study as well as better operationalize constructs and definitions of the study. This critical context added significantly to the literature and supported applied utility by examining potential antecedents to turnover (Hausknecht & Trevor, 2011; Hom et al., 2017; Nyberg & Ployhart, 2013). Allen et al. (2014) further argued that context specific research can help reshape the DAM that they argue hampers turnover research today. To overcome the DAM, Allen et al. (2014) recommend that exploration be inclusive, including at least voluntary and involuntary turnover operationalizations, with measurement beyond surveys, explore actual turnover as opposed to intentions, and include more than one organization.

This study explored the relationship of SJTs and initial job training performance associated with voluntary and involuntary turnover in a mid-sized police organization in Florida. The primary purpose of this research was to explore any potential predictive nexus between selection battery scores, including SJT, and turnover. The potential applied utility includes the potential to inform law enforcement managers if selection battery scores offer any potential predictive utility, thereby allowing interventions to potentially increase officer

retention. This chapter will review the research methodology suggested for this study.

Additionally, this chapter discusses the research questions, research design, methodology, population sample, data acquisition, analysis, assumptions, and boundary conditions.

Research Questions and Hypotheses

This study explored the existing gap in literature regarding SJT's potential predictive nexus with turnover (Ployhart, 2006; Ryan & Ployhart, 2013; Price, 2011). This study has potential applied research implications in seeking to better understand how SJTs, along with other selection instrument scores, may predict actual turnover, and how it may impact actual voluntary (choice) and involuntary (performance) turnover. The following overarching research questions were explored with several hypotheses.

RQ1. Do Situational Judgement Tests (SJTs) offer predictive utility toward field training performance.

RQ2. Do SJTs offer predictive utility for turnover outcomes.

The potential impact of exploring these constructs includes potential findings to inform police organizations of a proactive opportunity to develop and provide early interventions toward improved success in the transfer of training related to retention. Additionally, findings potentially may offer valid prediction data of leaving, an avoidable form of voluntary turnover. Hom et al. (2017) and Price (2011) suggested that selection research with the potential to inform organization's selection decisions to assist in screening out candidates with high

propensity to leave should be explored. Based on the review of literature and specific knowledge of the job domain of law enforcement the following hypotheses were posited:

Hypothesis 1 (H1): Scores on the video-based SJT (video score) and reading scores will offer prediction in field training performance.

Hypothesis 1a (H1a): Lower SJT sub scores in critical thinking will predict field training performance.

Hypothesis 1b (H1b): Lower scores in the reading assessment will predict field training performance.

Hypothesis 2 (H2): SJT sub-scores will add valid prediction to the three main scores (video, writing, and reading) toward overall turnover.

Hypothesis 2a (H2a): Lower SJT sub scores in critical thinking and reading will predict involuntary turnover.

Hypothesis 2b (H2b): Higher SJT sub scores in writing and critical thinking with lower sub scores in confrontation and organizational commitment will predict voluntary turnover.

Research Design

The dependent variables (DVs) include officer retention as measured simply by the fact that the officer has not resigned or been terminated and thus is still employed by the agency. Turnover is defined as intentional (i.e. officer leaves) or unintentional (i.e. officer is terminated). This assumes that retention, as opposed to

turnover, implies that at least the minimally accepted performance is being maintained. Additional DVs are daily observation performance evaluations.

The independent variables (IVs) are the selection instruments scores consisting of the video-based SJT with sub-scores, video-based situational report writing exercise, and reading instrument. The SJT sub-scores are broken down into critical thinking, communication, team orientation, confrontation, restraint, ethical orientation, and empathy scores, respectively.

Overview of Research Approach Used in the Study

This study used archival data from a southern mid-sized law enforcement agency. The research purpose is of interest to the agency administration in their quest to facilitate improved training and retention of new officers through their post-academy field training programs. The potential impact of this study includes offering a proactive opportunity to provide early interventions to improve success in core training and retention outcomes.

Population and Sample

Archival data came from official records of new police officers hired since 2012 from a southern mid-sized police agency (150-200 officers) who are utilizing video-based SJTs as part of their selection process. Other selection data was collected from each subject as well as daily observation reports (performance evaluations) and retention outcome data as retained, terminated (unintentional turnover), or resignation (intentional turnover).

Appropriateness of population. The law enforcement population is appropriate due to the common use in the industry of SJTs in selection processes and the unique nature of daily performance evaluations during field training. Newly hired officers are put through a department orientation period of a month before starting a fourteen-week field training and evaluation program (FTEP) where state certified and trained field training officers (FTOs) provide training and guidance in the field for probationary police officers (PPOs).

Their daily observations reports (DORs) are anchored scale performance ratings based on the expectations of a solo officer working alone. Law enforcement offers a unique and rich opportunity to explore this post-hiring on-the-job training performance from these daily ratings. The potential to identify prospective problem areas in training and address these potential performance deficiencies prior to the start of training has promise to improve successful training completion, which has positive aspects to agencies and new hires alike.

Procedures

The law enforcement agency provided the employment testing component scores of the video-based SJT as well as the sub-scores, the video-based situational report writing exercise, and the reading instrument. Each officer's scores were coded with a subject identifier to protect the identity of each officer. The agency also provided data regarding the officers' retention status as retained, terminated (unintentional turnover), or resignation (intentional turnover). Finally, the agency

provided electronic copies of all DORs. The DORs were coded into a Microsoft Excel file with officer names replaced by their appropriate subject identifier. The data was subsequently loaded into SPSS statistical software for analysis.

Data Analysis

This study used logistic regression for data analysis. Logistic regression allows for the prediction of group membership (Tabachnick & Fidell, 2007), and in this case, turnover. This predictive modeling technique is used when the categorical dependent variable is dichotomous (allowing only one outcome) and the predictor variables are continuous, discrete, or dichotomous (Pallant, 2013; Tabachnick & Fidell, 2007).

According to Tabachnick and Fidell (2007) the limitations of logistic regression are limited mostly from the discrete requirement for the outcome dependent variable. Strengths include allowing the exploration of interactions among predictors for possible multicollinearity where potential predictors are covariates in prediction related to the dependent variable (Tabachnick & Fidell, 2007).

Considerations in choosing logistic regression for this study include logistic regression's assumption that the predictors have a linear relationship to the dependent variable (Tabachnick & Fidell, 2007). In this case, various selection assessment outcomes toward turnover. Additionally, Tabachnick and Fidell (2007) stress the importance of theory driven selection of predictors for this type of

modeling. This research theoretically predicts involuntary turnover resulting from, in part, a transfer of training issue related to standards of performance. Voluntary turnover is theorized to occur, in some part, due to P-J fit, P-O fit, or shocks.

Predictor Measures

Successfully hired officer candidates have to complete a multiple hurdle selection process at the population agency. One of these hurdles involves a test day where candidates take a three-part test. The exam is provided by a vendor so proprietary information or data on the exam was not accessible, but the vendor has validated the test and it is used throughout the U.S.

SJT video Score. One part of the exam involves a series of law enforcement workplace video scenarios lasting one minute or less followed by a short period for candidates to select the best option of response, multiple-choice, to the scenario. This score is reported as a percentage with the passing cut score of 65 percent.

The SJT video scores has sub-scores reported in terms of applicant suitability. These are ranked as *high risk* (1), *some problems* (2), *normal* (3), or *good* (4). The sub-domains are:

Critical Thinking Ability. Accurately analyze immediate field situations in timely manner.

Investigative Communication. Effectively and analytically gaining information from pertinent parties.

Team Orientation. Collaborates effectively on field situations.

Confrontation. Initiative without delay or avoidance to confront situations.

Restraint. Uses only necessary and appropriate levels of force to resolve law enforcement situations.

Ethical Orientation. Makes ethical decisions regardless of sacrifice.

Empathy. Exhibits appropriate understanding and approaching emotional and challenging human situations.

Organization Orientation. Contributes to organizational initiatives and supervisor efforts.

Community Relations. Courteous and helpful to community policing goals.

Writing score. Candidates watch a video-based scenario common to a citizen-police encounter where candidates must then write a structured written report. While the underlying constructs are not available from the vendor, the exercise is face valid and a realistic representation of a work product. This exercise is easily related to listening skills and situational awareness as well as demonstrative of writing ability and sense-making. This score is reported as a percentage with a passing cut score of 70 percent.

Reading score. A third test is administered where work-related situations are presented in a timed reading format and comprehension is tested. This score is reported as a percentage with a passing cut score of 70 percent.

Criterion Measures

The criterion measures are related to turnover. These are field training performance, overall turnover, involuntary turnover, and voluntary turnover.

Various predictor relationships to criterion measures were explored using a direct logistical regression. The direct logistical regression allowed for testing of a predictive model, from hypotheses, with statistical explanations of the criterion as a whole, i.e. turnover, along with predicting those who were retained or not.

Ethical Considerations

The primary ethical consideration is to ensure subject anonymity. All data is archival, but did have individual identifiers with names and department identification numbers. Once the data was transposed into Microsoft Excel per subject, all personal identifiers were replaced with a subject number. The Florida Institute of Technology Institutional Review Board (IRB) was approved and ensured that the human subjects were appropriately protected.

Researcher Positionality

England (1994) discussed research as a *process* as well as a *product*. Bourke (2014) discussed researcher positionality, specifically researcher personality and *insider* or *outsider* status, in the context of researcher reflection and acknowledging potential bias. In this light, it is important to acknowledge the researcher as a law enforcement practitioner *insider* (Bourke, 2014) having retired from law enforcement with FTEP experience, being a police trainer, as well as a

supervisor. The researcher consults regularly with law enforcement agencies since retirement as well as conducts applied research in areas that include recruiting, training, retention, and promotional processes.

The researcher acknowledges that being a subject matter expert is an advantage, but that it can also be a source of bias and must be acknowledged. Bias is possible given previous practitioner experiences and consulting. The researcher will endeavor to utilize extensive literature review and work closely with dissertation committee members to mitigate potential recognized areas of bias and conduct research objectively.

Validity and Trustworthiness

SJTs have considerable evidence of criterion-related validity as they relate to the criteria of work performance, but the constructs as to why are still elusive. While the use of SJTs is widely accepted and has demonstrable validity advantages over cognitive and personality measures, there still lacks a depth of operationalization and theoretic explanations for the exact constructs being measured by the SJTs. The discussion and debate of exactly what SJTs tap into is garnering attention.

Christian, Edwards, and Bradley (2010) conducted a meta-analytic review of SJTs to possibly address the lacking literature on construct measures associated with SJTs in a quest to establish theory-driven frameworks. Their research resulted in support for SJTs assessing leadership skills, interpersonal skills, and where

applied, teamwork skills. Their findings included that matching predictors to criterion constructs improves criterion validity. They also suggest video-based SJTs possess stronger criterion validity than written scenarios.

A SJT designed to tap police integrity was conducted by de Meijer, Born, van Zielst, and van der Molen (2010) where they examined a multi-ethnic section of Dutch police. They found that when SJTs were used to measure one construct, in this case integrity, it had construct validity across several ethnic populations. Husbands, Rodgers, Dowell, and Patterson (2015) also found psychometric robustness for integrity SJTs among medical student applicants.

Lievens and Motowidlo (2016) have also engaged in an effort to generate interest in clarifying constructs of SJTs to better understand why SJTs actually work. They offer a model proposing SJTs as a measure of general domain knowledge that they define as, “knowledge about the utility or importance of traits such as these for effectiveness in a job that actually requires expressions of these traits for effective performance” (p. 4).

Validity specifically related to this study was addressed with the following considerations. First, all constructs were operationalized for clarity of constructs and definition. Second, the agency in the study deployed a selection battery from a third-party vendor specializing in the industry of law enforcement, thus ensuring that the selection instrument scores are uniform and consistent related to industry standards. Third, FTEP standards of daily performance (DORs) are established by

the Florida Department of Law Enforcement (FDLE) which mitigates, to a practical level, issues and assumptions of inter-rater reliability that were used for performance data.

Assumptions and Boundary Conditions

This section attempts to address recommendations previously discussed related to context and assumptions of turnover related research (Allen et al., 2014; Hausknecht & Trevor, 2011; Hom et al., 2017; Nyberg & Ployhart, 2013). The below assumptions and boundary conditions allow proper framing for possible generalization, potential future research suggestions, and applied application potential.

Assumptions. It is important to discuss key assumptions related to this research. Of particular importance are the underlying assumptions related to the overarching theories, SJTs, faking, field training, transfer of training, and turnover.

Overarching theories. Schneider's (1987) attraction-selection-attrition (ASA) model offers a linear theoretical foundation for this study. ASA related theories of person job (P-J) fit (Edwards, 1991), person organization (P-O) fit (Judge & Ferris, 1992), and shocks (Lee & Mitchell, 1994), such as unanticipated job offers, impact voluntary turnover (Arthur et al., 2006; Holland, 1993; Hom et al., 2017; Griffith, et al., 2000; Lee & Mitchell, 1994; Schneider, 1987; Tett & Meyer, 1993). Transfer of training (Blume et al., 2010; Baldwin, Ford, & Blume, 2009; Fecteau et al., 1995; Goldstein & Ford, 2002) has nexus to involuntary

turnover.

SJT and selection battery. First, this study did not explore specifics on the reliability or the validity of the SJT deployed by the organization. Instead, the assumption is that the SJT, and related selection battery components, are valid and reliable instruments for the police organization. Further, this assumption includes that the study organization deploys selection batteries consistent with employment law as well as selection and psychometric standards (EEOC, 1978; Gutman, 2011; SIOP, 2003).

Faking. The assumption for this research, related to the prevalence of faking, is that any faking falls within the concept of a *social desirability* type (Ones & Viswesvaran, 1998) of potential candidate manipulation. It is further assumed that this type of faking results in more positive scores, indicating that the candidate recognizes the correct “would do” response in the scenarios regardless of if their actual behavior and performance conforms (Ones & Viswesvaran, 1998; Oostrom et al., 2107). Additionally, in the absence of valid and reliable methods to detect faking (Birkeland et al., 2006; Hough et al, 1990; Griffith, et al., 2007; Ones et al., 1995) it leads to the assumption that some levels of “faking good” likely occur at unmeasurable or directly observable ways.

The video-based SJT, common to both police organizations, deploys item responses based on common situations officers may encounter on the job. These are considered behavioral “would do” responses (McDaniel & Nguyen, 2001). While it

is recognized that knowledge-based “should do” responses (McDaniel et al., 2007) can reduce the likelihood of faking, this would create additional problems. Organizations that deploy “should do” items create an advantage to those that have completed training and it furthers those with experience. Deploying a knowledge-based item response would significantly disadvantage applicants that have yet to attend the police academy let alone have any policing experience. Therefore, the assumption is that the deployed SJT, and related batteries, recognize that applicants tested have various degrees of police experience ranging from none to years of experience. While the “should do” response may reduce faking it would also disadvantage candidates who have yet to have the opportunity for basic academic training.

Field training. The overarching assumption regarding field training is that the field training and evaluation programs (FTEP) are based on common theoretical and technical frameworks (FDLE, 2011). This supports the assumption that the field training process is generalizable in the industry for the purposes of this research. The population agency’s field training program uses the popular San Jose model (Caro, 2011; Chappell, 2007; COPS Office, 2001; Fischer, 2015; Kaminski, 2002; Scott, 2000; Scott, 2010; Walker, 2005; Walker & Katz, 2002). The study agency complies with state requirements for the training of field training officers (FDLE, 2011). The field training officer course is a 40-hour course designed to facilitate training and evaluation of new officers consistent with the San Jose model

that includes aspects related to training theory, supervision, observation, evaluations using daily observations reports (DORs), remedial training, counseling, and field training structure (FDLE, 2011).

Transfer of training. The assumption is that new, non-experienced officers beginning field training from the basic police academy are not fully prepared for field duty. This is supported in the literature discussing directly that the basic police academy does not address the actual field performance (Buerger, 1998; Harr, 2001; Mastrofski & Ritti, 1995). The assumption includes that a minimum amount of KSAO development comes from the academy, as supported by Caro's (2011) study that only 10 percent of field training variance is explained by police academy performance. Caro's (2011) finding is consistent with other transfer of training literature suggesting that 10 percent of training is actually transferred (Baldwin & Ford, 1988; Clark, 2010; Fitzpatrick, 2001; Georgenson, 1982). This study further assumed that successful transfer of training rests, in part, on cognitive ability (Grossman & Salas, 2011; Gully et al., 2002; Maltarich et al., 2010; Saks et al. 2014; Van Iddekinge et al., 2011).

Turnover. The key assumption to this study is that turnover is considered, for the purposes of the study, as *overall* (Mobley, 1977), *voluntary* (Mobley et al., 1979), or *involuntary* (McElroy et al., 2001). The dependent variable was measured as such (overall, voluntary, or involuntary) depending on the specific hypothesis being tested. The theoretical assumption of this predictive research is that the

selection battery, including the SJT, has a predictive nexus to turnover through either P-J fit (Kristof-Brown et al., 2005), P-O fit (Maltarich et al., 2010), various shocks (Burton et al., 2010), or failures in the transfer of training (Baldwin et al., 2017).

Boundary Conditions. Both Allen et al. (2014) and Hom et al. (2017) contended that turnover related literature will improve with clear boundary conditions. The boundary conditions for this study include sector and job, selection battery and scores, field training, and turnover.

Sector and job domain. The context of the study is a public-sector industry with the domain of policing. The study population consisted of a Florida southeast mid-sized police department with less than 250 sworn officers. Generalizability may be inferred to U.S. police agencies who are using the San Jose model and who deploy a standardized selection process, including SJTs, but the boundary to this project is the Florida organization.

SJT and selection battery. The selection battery, including the video based SJT, originates from an industry vendor. The name of the vendor, along with the agency name and individual test takers, are intentionally unnamed to protect the identity of the agencies, subjects, and stake holders, in keeping with institutional review board (IRB) rules and regulations regarding confidentiality.

Field training. The study population is a Florida organization. While the context may have generalizability to police organizations across and outside of

Florida, the subjects and related organizational structure conform to Florida standards in basic training, field training, and advanced training. The context of the FTEP, including phases of training, DORs, and remedial training options are based on the San Jose model and meet the standards of Florida law enforcement (FDLE, 2014).

Turnover. Turnover will be explored as overall, voluntary, and involuntary bound with the theories outlined in the assumptions above. While the ASA (Schneider, 1987), P-J fit (Kristof-Brown et al., 2005), P-O fit (Maltarich et al., 2010), various shocks (Burton et al., 2010), and transfer of training (Baldwin et al., 2017) based models are consistent with literature, they are not the only potential models of police turnover.

Chapter 4

Findings

Overview

Chapter 4 is intended to present the results and findings of the exploratory study of situational judgement tests (SJTs) on training and turnover outcomes. This chapter begins with a summary of the research problem and research questions previously explored in chapter 1. Next, aspects of the data collection, the data itself, management of the data, and analysis implications will be explored. Research findings related to all hypotheses as well as supplemental analyses will be presented. The chapter continues with a synthesis and summary of the findings. The chapter concludes with a discussion related to the study validity, contributions of the study to applied practice, and the implications for the last chapter of this dissertation.

Summary of the Research Problem

This study is intended to explore the problem faced by police organizations to recruit, select, train, and retain qualified police officers (Cavanaugh, 2013; Haarr, 2001; Jordan et al., 2009; Koper et al., 2001; Maciag, 2018; National Research Council, 2004; Orrick, 2008; PERF, 2013; Selden et al., 2013; CSLGE, 2017; Smith, 2016; Wilson, 2012; Wilson, 2014; Wolfe & Nix, 2016; Woska, 2006; Woska, 2016). Police organizations are challenged in effectively addressing this problem due to a host of dynamic economic, social, and workforce issues.

An increasing challenge is a rapidly diminishing applicant pool where

applicants have more, and often more preferable, options in other industries and larger police organizations with more attractive benefits, pay, and career track options (Barrett et al., 2010; Cavanagh, 2003; Haarr, 2001; Koper, Maguire, & Moore, 2001; Orrick, 2008; Selden et al., 2013; Smith, 2016; Wilson, 2012; Woska, 2016). Small and mid-sized agencies make up the majority of police agencies across the country. Recruiting, selection, and training challenges can result in dysfunctional and costly turnover among new officers. This forces agencies to return to a diminishing candidate pool for quality applicants.

The use of SJTs in employee selection has substantial empirical support toward predicting job performance (Clevenger, Pereira, Wiechmann, Schmidt, & Harvey, 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Weekley & Jones, 1999). Chan and Schmitt (2002) and Weekley and Ployhart (2005). Research suggest that SJTs have higher selection validity over the combined validity of cognition, personality, and previous work experience (Chan & Schmitt, 2002; Weekley & Ployhart, 2005). Not only have SJTs demonstrated better predictive validity in job performance validity compared to other options, but they also produce less adverse impacts (Clevenger, Pereira, Wiechmann, Schmidt, & Harvey, 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Weekley & Jones, 1999). Recent research exploring “what” SJTs actually tapp suggests they may offer prediction of specific job relevant constructs as well as general domain knowledge (Lievens & Motowidlo, 2016).

In a review of a century of turnover research, Hom, Lee, Shaw, and Hausknecht (2017) found a severe lack of research on selection and turnover. Rubenstein, Eberly, and Lee (2015) conducted a meta-analysis of turnover and suggested that future research explore actual turnover, rather than intentions, and explore turnover as *voluntary* or *involuntary*.

This research gap in the selection process and turnover has the potential to inform selection partitions. Research specifically exploring the utility of SJTs, and other selection tools, in predicting field training and actual turnover outcomes for policing and other high consequence professions may provide valuable information to improve outcomes for employees and organizations. Identifying research and data to assist organizations strategically to improve training and turnover outcomes is important to practitioners. Resources for research and interventions, especially in smaller public sector organizations, can prove costly. Searching for additional “value added” utility from selection instruments to inform selection, recruiting, and retention practices and potential training interventions may allow agencies strategic opportunities that resources may otherwise not allow.

Purpose of the Study

The purpose of this study is to explore the existing gap in literature regarding SJT’s potential predictive nexus with turnover outcomes (Ployhart, 2006; Ryan & Ployhart, 2013; Price, 2011). This proposal seeks to better explain how SJTs and other selection instruments may predict actual voluntary, involuntary

(including training failure), and overall turnover. This exploratory research is searching for applied utility regarding predictive pathways to improve field training and other turnover outcomes.

Research Questions

This study explores the potential predictive nexus between SJTs, along with other selection instrument scores, to officer voluntary (choice), involuntary (performance, discipline), and overall (collective) turnover. The following overarching research questions will be explored with several hypotheses.

RQ1. Do Situational Judgement Tests (SJTs) offer predictive utility toward field training performance.

RQ2. Do SJTs offer predictive utility for turnover outcomes.

This exploratory study endeavors to better understand any predictive nexus between selection tests and retention. The applied goal is to seek potential added utility of existing testing and potentially inform police organizations to develop strategies to enhance the transfer of training outcomes along with other forms of voluntary and involuntary turnover. While this research is exploratory, specific hypotheses are necessary in the research. Based on literature review and specific knowledge of the job domain of law enforcement the following hypotheses will be explored, along with supplemental analyses based on exploratory findings.

Hypothesis 1 (H1): Scores on the video-based SJT (video score) and reading scores will offer prediction in field training performance.

Hypothesis 1a (H1a): Lower SJT sub scores in critical thinking will predict field training performance.

Hypothesis 1b (H1b): Lower scores in the reading assessment will predict field training performance.

Hypothesis 2 (H2): SJT sub-scores will add valid prediction to the three main scores (video, writing, and reading) toward overall turnover.

Hypothesis 2a (H2a): Lower SJT sub scores in critical thinking and reading will predict involuntary turnover.

Hypothesis 2b (H2b): Higher SJT sub scores in writing and critical thinking with lower sub scores in confrontation and organizational commitment will predict voluntary turnover.

Data Collection, Data Management, and Analysis Implications

Archival data from official records on officer selection, field training, turnover outcomes, and retention was used in this study from a southern mid-sized (150-200 officers) law enforcement agency. This data is limited to officers selected using the SJT selection battery deployed in October 2012. Officers hired prior to use of the October 2012 process are not part of this study and not reflected in turnover data. Selection tests include SJT video scores with sub-scores, writing scores, and reading scores. Field training and evaluation program (FTEP) performance data comes from daily observation reports (performance evaluations) and the retention outcome data denoted as retained, terminated (unintentional

turnover), or resignation (intentional turnover) comes from agency recruiting and separation records.

Predictor Measures

The researched agency utilizes a multiple hurdle selection process for hiring officers. These hurdles include a three-part video-based test implemented in October 2012. This is the source of the SJT video score, including sub-scores, writing score, and reading score. The exam is provided by a vendor so proprietary information or data on the exam was not accessible, but the vendor has validated the test used throughout the U.S.

SJT video score. This part of the exam utilizes a series of law enforcement video vignettes of one minute or less followed by time for candidates to select what they consider the best response for that scenario. Each vignette response is multiple-choice. This score is reported as a percentage with the passing cut score being 65 percent. Scores for the population (N=101) ranged from 67.93 to 92.39 percent.

The SJT video included specific sub-scores reported in terms of applicant suitability. These are ranked as *risk* (1), *some problems* (2), *normal* (3), or *good* (4). The sub-domains with distributions are:

Critical thinking. Accurately analyze immediate field situations in a timely manner. Distributions are high risk (n=0), some problems (n=3), normal (n=78), and good (n=20) reflecting poor range distribution.

Investigative. Effectively and analytically gaining information from pertinent parties. Distributions are high risk (n=0), some problems (n=0), normal (n=95), and good (n=6) reflecting very poor distribution.

Team orientation. Collaborates effectively on field situations. Distributions are high risk (n=5), some problems (n=10), normal (n=78), and good (n=8) reflecting adequate score distribution.

Confrontation. Initiative without delay or avoidance to confront situations. Distributions are high risk (n=0), some problems (n=1), normal (n=69), and good (n=31) reflecting poor distribution at the lower bounds.

Restraint. Uses only necessary and appropriate levels of force to resolve law enforcement situations. Distributions are high risk (n=1), some problems (n=7), normal (n=70), and good (n=23) reflecting some distribution.

Ethical orientation. Makes ethical decisions regardless of sacrifice. Distributions are high risk (n=1), some problems (n=13), normal (n=69), and good (n=18) reflecting adequate score distribution.

Empathy. Exhibits appropriate understanding when approaching emotional and challenging human situations. Distributions are high risk (n=0), some problems (n=2), normal (n=59), and good (n=50) reflecting poor distribution at the lower bounds.

Organization orientation. Contributes to organizational initiatives and supervisor efforts. Distributions are high risk (n=1), some problems (n=12), normal

(n=79), and good (n=9) reflecting adequate score distribution.

Community relations. Courteous and helpful to community policing goals. Distributions are high risk (n=1), some problems (n=3), normal (n=84), and good (n=13) reflecting score distribution.

SJT sub-score dichotomous recode. Range restriction limits utility on some of the aforementioned sub-scores. To fully explore potential predictive models, each sub-score was recoded into additional codes of “sub-score” risk, “sub-score” problem, “sub-score” either risk or problem, “sub-score” normal, and “sub-score” good. Each dichotomous additional sub-score was coded as “0” for no and “1” for yes as suggested by Field (2013), Pallant (2013), and Tabachnick and Fidell (2007).

Writing score. The writing score is derived from when candidates watch a video-based citizen-police scenario followed by writing a structured written report. The underlying constructs are not available from the vendor but the exercise offers face validity and a realistic representation of a police work product. This score is reported as a percentage with a passing cut score of 70 percent and is considered to be a continuous variable.

Writing dichotomous recode. While the writing scores are continuous (N=101), the score distribution was limited to seven scores. Over 40 percent of the candidates (n=41) scored a perfect 100. The second highest score recorded is a 93.3 accounting for nearly 40 percent (n=40). The remaining 9.8 percent of scores were

either 88.3 (n=3), 86.67 (n=11), 81.67 (n=1), 80 (n=3), or 73.33 (n=2). Given the restricted distribution, additional dichotomous variables were created with writing score “bands” of *writing 100*, *writing 96.67*, *writing 93.33*, *writing less than 90*, *writing less than 82* and *writing less than 80*. Each dichotomous writing score band was coded “0” for no or “1” for yes (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007).

Reading score. The reading test is the third part the of the selection exam. This test is a paper and pencil work-related situational judgment test where candidates are asked to read scenarios and select the best multiple-choice response. This score is reported as a percentage with a passing cut score of 70 percent. It is a continuous variable for the study.

Reading dichotomous recode. Interestingly, nearly half (n=48, 47.5 percent) of the 101 subjects scored 100 on the reading assessment. The next lower score, 96.67 percent, accounted for 33.7 percent of scores (n=34). A reading score of 93.33 accounted for 11.9 percent of the subject scores (n=12). The lowest scores were 90 (n=2), 86.67 (n=3), 83.33 (n=1), and 80 (n=1). Reading scores, like writing, had only seven scores although the score range is 80-100 percent. To maximize exploratory options additional variables were created with reading score “bands” of *100*, *96.67*, *less than 93.33*, and *less than 90*. Each dichotomous reading “band” is coded “0” for no and “1” for yes as suggested by Field (2013), Pallant (2013) and Tabachnick and Fidell (2007).

Additional predictor variables. Additional predictor variables were used in exploring various supplemental analyses models. These were added to models containing SJT video scores, writing scores, reading scores, and SJT sub-scores in the exploratory quest for predictive modeling in the scope of this research.

Candidate retest and retest (fail, pass) and retest (pass, pass). Variables used in reported analyses include *candidate retest* (n=14), indicating if candidates took the selection battery more than once. Retest was further explored with the variables of *Retest_FP*, indicating candidate retesting with an initial fail and subsequent pass (n=7). Candidates who passed and then retested (n=7) is represented in the *Retest_PP* variable.

Prior military experience. Officers with military experience (n=31) prior to selection testing was also explored. Officers with recorded military experience were coded “1” and those without recorded experience (n=66) were coded “0” as recommended (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007). Four officers had missing data regarding this variable therefore lowering the N for analyses utilizing this variable to 97 versus 101.

Prior law enforcement experience. Officers with prior law enforcement experience (N=15) was explored. Officers with previous law enforcement experience were coded as “1” and those without recorded experience (n=84) were coded “0” as recommended (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007). Two officers had missing data regarding any prior law enforcement

experience, therefore lowering the N for analyses utilizing this variable to 99 versus 101.

Dependent Variables

The dependent variables for this study are dichotomous measures related to turnover outcomes for the population of officers (n=101) hired since the adoption of the SJT selection battery in October 2012. Officers hired prior to deployment of the October 2012 selection battery are not part of this study. All outcome dependent variables are coded “0” for the outcome did not occur, and “1” indicating the outcome occurred (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007). The following dependent variables were utilized for predictive modeling in this study.

Field training and evaluation program (FTEP) passing. Passing FTEP is the most common type of retention variable in the study (n=85). This indicated that officers passed all four phases of FTEP. Some officers required remedial training and subsequently passed (n=7). While some officers ultimately left the agency, this outcome variable is important in better understanding attributes of those successful in passing field training.

Field training and evaluation program (FTEP) failure. Failure during FTEP is the most common type of involuntary turnover (n=16). This indicated that officers failed field training at any of the four phases of FTEP. These officers were given an opportunity for remedial training and either subsequently failed (n=13) or

quit when notified they were being remediated in training (n=3). Officers quitting upon receiving a remediation notification are considered *failing for performance* in this study given that they were rated as *not responding to training* and would have been terminated for not successfully completing the remediation.

Involuntary turnover. Involuntary turnover (n=21) accounts for both FTEP performance failure (n=16) and discipline (n=5). Discipline occurred during FTEP (n=3) and post FTEP (n=2). One subject was excluded from this category due to being injured (not duty related) and therefore this was not reflective of the construct of performance failure or discipline.

Voluntary turnover. Voluntary turnover (n=22) accounts for officers who made decisions to separate from the agency. These include changing their mind about law enforcement during FTEP (n=9), leaving for another law enforcement agency (n=4), changing professions (n=3), relocating due to spouse employment relocation (n=2), moving closer to home (n=2), and changing mind about law enforcement post FTEP (n=2). One officer, with substantial previous law enforcement experience, retired at normal retirement age and was excluded from this category. Given the observation period of October 2012 to present and the absence of all other collective turnover, this candidate, for the purposes of turnover in this population, is an outlier.

Overall turnover. Overall turnover (n=45) indicated that officers were not retained. This includes involuntary and voluntary turnover as well as the two

officers mentioned above who left due to medical and retirement reasons. These two officers were retained in the overall turnover.

Need for remedial training. Officers with performance deficiencies during FTEP and deemed *not responding to training* or needing additional time through *extension* in the phase of training are considered in *remedial training* (n=23). These officers are given additional training and time, as much as two weeks, to perform at standard or face training failure. Those who fail remedial training (n=13) are terminated and those successful in passing remedial (n=7). Three officers quit when informed they were being remediated for substandard performance issues.

Failing remedial training. Officers failing remedial training (n=13) are terminated. The most common performance domains, as reflected in DORs are officer safety, control of conflict, perception/judgment, problem solving/decision making, radio procedures, report writing, and understanding and performing in accordance with department policies and procedures. A full breakdown of performance deficiencies for all officers requiring remedial training is in the Appendix. While this study will not report models predicting any specific performance domain it is important for context related to remedial training failure models. Future research recommendations include exploring specific performance domains and will be discussed in chapter 5.

Methods and Techniques

This exploratory study deployed binary logistic regression analysis as it

allows for the prediction of group membership (Tabachnick & Fidell, 2007). Binary logistic regression analyses are appropriate when there is a dichotomous categorical dependent variable, no (0) and yes (1), with continuous, discrete, or dichotomous predictor variables (Pallant, 2013; Tabachnick & Fidell, 2007).

According to Tabachnick and Fidell (2007), the limitations of logistic regression are mostly related to the discrete requirement for the outcome dependent variable. Strengths include allowing exploration of interactions among predictors for possible multicollinearity where potential predictors are covariates in prediction related to the dependent variable (Tabachnick & Fidell, 2007).

Logistic regression was selected for this study because assumptions of the method include that the predictors have a linear relationship to the dependent variable (Tabachnick & Fidell, 2007). Therefore, the various selection scores and other pre-employment selection independent variables are assumed to predict the outcome variables related to turnover.

Tabachnick and Fidell (2007) stressed the importance of theory driven selection of predictors for this type of modeling. The models explored in this study are based on the independent variables related to selection testing, to include retesting, along with officer pre-employment attributes such as prior law enforcement experience, prior military experience, and college. Theoretically, these independent variables may offer a predictive nexus to outcome variables. FTEP training performance, FTEP remedial training necessity, and subsequent remedial

training success related to potential involuntary and overall turnover are theoretically grounded in transfer of training literature explored in chapter 2. Additional involuntary and voluntary outcomes are theoretically supported in research related to person-job (P-J) fit, person-organization (P-O) fit, and shocks.

SPSS analysis and relevant processes. IBM SPSS is relatively straight forward in running binary logistic regression. The default binary logistic regression procedure is *forced entry method* opposed to *stepwise* (Pallant, 2013). The *forced entry method* is preferred as the procedure as it allows theory-based model exploration accounting for simultaneous testing of predictors while controlling for the effects of other predictors in the model (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007). Tabachnick and Fidell (2007) stress the importance of theory driven selection of predictors over stepwise procedures. Pallant (2013) added that relying solely on statistical grounds in exploring predictive models lead to incomplete models due to the influence that random statistical variations may have on removing variables based solely on statistical grounds verses theoretically rational decisions.

One-tailed versus two-tailed hypothesis tests. Field (2013) discussed that a *one-tailed test* can be used in testing hypotheses when the direction of the relationship is assumed or known. When used, the one-tailed test disregards the possibility of the relationship occurring in the other direction (UCLA, n.d.). *Two-tailed tests* are more commonly used as direction of relationships can occur in both

directions. Decisions in using one-tailed versus two-tailed tests should be a decision made, when appropriate, in advance of running analyses (UCLA, n.d.).

The assumption in binary logistic regression is that the relationship between predictors and outcome variables is linear (Tabachnick & Fidell, 2007), which lends support for the possibility of using a one-tailed test of significance. This study will, however, use the more conservative two-tailed test of significance. The population of this study is adequate, but small. Future research recommendations, discussed fully in the next chapter, include adding additional agencies and subjects to increase the study population. Independent variables that would be considered *significant*, or *approaching significance*, using a one-tailed test will be noted in the results but not deployed for this study regarding significance testing for models.

Hosmer and Lemeshow goodness of fit test criticisms. Allison (2014) discussed drawbacks to the Hosmer and Lemeshow (HL) goodness of fit test used in logistic regression. He contends that, despite issues, the HL test is often the standard goodness of fit test of statistical software, including SPSS. In fact, Hosmer, Hosmer, Le Cessie, and Lemeshow (1997) point out the disadvantages of the HL test. Hosmer et al. (1997) point to a specific concern related to their logistic regression goodness of fit test regarding using fixed groups of estimated probabilities. They contend that the HL test grouping strategies based on these estimated probabilities can incorrectly assess goodness of fit when subjects when covariates have *widely different* values. Hosmer et al. (1997) ran simulated studies

and found that the HL test will suggest model fit for one group and reject the same model with a different group.

Allison (2014) added issues related to illogical inconsistencies in goodness of fit estimations based on adding, or removing, variables to the model. He contends that there is no logic to these changes. Given that binary logistic regression relies on theoretically driven models (Tabachnick & Fidell, 2007), the HL test was not used in the analysis.

Direct of the relationship (B values). The B value helps us understand the probability of an independent variable (continuous or dichotomous) predicting the outcome (dichotomous) variable (Pallant, 2013). The B value can be positive or negative, indicating the direction of the relationship. Negative B values indicate that an increase in the predictor score will result in a decreased probability of predicting the outcome (Field, 2013; Pallant, 2013). All outcome dependent variables in this study used codes of “0” for “no”, indicating subjects not belonging to the predictor group, and “1” for “yes”, indicating subjects belong to the predictor group. This coding is intended to facilitate proper interpretation of direction of model interactions.

Manual corrections in dichotomous independent variable odds ratios. Tabachnick and Fidell (2007) explained *odds ratios* as representing the change in odds that an independent variable predicts in an outcome variable by one unit. This holds true for continuous variables, such as the SJT score, writing score, and

reading score. For dichotomous independent variables the interpretations change from increase or decrease by “one unit”, and for continual variables, to increase or decrease by a percentage factor.

Pallant (2013) advised inverting the odds factors of less than 1 in predictor dichotomous variables, such as sub-scores, prior military, prior law enforcement, and retest. This inversion is accomplished by manually calculating *1 divided by the odds ratio value*. Corrected odds ratios can be run in SPSS by removing the designation for “categorical” but that also inverts the B from positive to negative and vice versa (Pallant, 2013). These manual inversion calculations will be reported in the results section. Dichotomous variables with odds ratios less than 1 will be identified as inverted in appendices containing SPSS outputs for each relevant hypothesis or supplemental analysis.

Research Findings

Hypothesis 1: SJT Video and Reading Scores Predicting FTEP Passing

Hypothesis 1 (H1) posited that scores on the video-based SJT (video score) and reading scores would offer prediction in officers passing the field training and evaluation program (FTEP). A direct logistical regression was performed to assess the impact of both factors on field training performance. This hypothesis was supported in this model.

The model contained two continuous independent variables consisting of the SJT video and reading scores. The full model containing both predictor

variables was statistically significant $\chi^2(2, N = 101) = 8.32, p < .05$, indicating that the model was able to distinguish between officers who took the selection testing and passed FTEP and officers who failed FTEP. The model as a whole explained between 7.9% (Cox and Snell R square) and 13.6% (Nagelkerke R square) of variance in officer success during FTEP and correctly classified 100% (85/85) of officers that passed training and 18.8% (3/16) of officers that failed training with an overall correct classification of 87.1.

Table 1 demonstrates that only one of the independent variables, reading score, contributes significantly to the model. Reading, being the only strong predictor, recorded an odds ratio of 1.19. This indicates that officers were 1.19 times more likely to pass training for every 1 point (unit) increase in reading score. Officers who failed during FTEP had lower reading scores controlling for the other factor, SJT video score, in the model.

Table 1

Logistic Regression: SJT Video and Reading Scores Predicting Training Pass (H1)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
SJT Video	-0.29	.048	.349	1	.554	.55	.88	1.068
Reading Score	.179	.065	7.672	1	.006*	1.19	1.05	1.35

*p<0.05

The results of this model support *H*. Readings scores are continuous, but the distribution was restricted to seven distinct scores across all 101 subjects.

Hypothesis 1a: Lower Critical Thinking Sub-scores Predicting FTEP Failure

Hypothesis 1a (H1a) asserted that lower SJT sub-scores in critical thinking would predict FTEP failure. The direct logistical regression to assess the impact of the critical thinking sub-scores on FTEP failure did not result in a supported model. H1a was not supported.

The full model containing the critical thinking sub-score as the predictor was not statistically significant $\chi^2(1, N = 101) = .183, p. 669$, indicating that the model was unable to significantly distinguish between officers who took the selection testing and failed FTEP and those who passed FTEP. This model displayed no prediction in FTEP failure. Table 2 reflects that the critical thinking sub-score is not a predictor based on these data and non-supported predictive model.

Table 2

Logistic Regression: Critical Thinking Predicting Training Performance (H1a)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
Critical Thinking Sub-score	.266	.630	.178	1	.673	1.305	.38	4.48

The independent variable, critical thinking sub-score, for H1a proved considerably restrictive. All nine SJT sub-scores are continuous reflective scores; 1 (risk), 2 (some problem), 3 (normal), and 4 (good). The critical thinking sub-score was severely restricted in range. There were no scores reflecting risk, three (3%)

with some risk, the majority (n=78, 77.2%) scored in the normal range, while 20 (19.8%) scored good.

Hypothesis 1b: Lower Reading Scores Predicting FTEP failure.

Hypothesis 1b (H1b) expected lower reading scores to predict FTEP failure. A direct logistical regression was performed to assess the impact of reading scores on FTEP failure, which resulted in a supported model. This model contained reading scores as the independent variable exploring prediction of FTEP failure.

The full model was statistically significant $\chi^2(1, N = 101) = 7.97, p < .05$, indicating that the model was able to distinguish between officers taking the selection testing who failed FTEP and those who passed FTEP. The model as a whole explained between 7.6% (Cox and Snell R square) and 13% (Nagelkerke R square) of variance in officers failing FTEP and correctly classified 18.8% (3/16) of officers who failed training and 97.6% (83/85) of officers who passed with an overall correct classification of 85.1 percent.

Table 3 demonstrates that reading score do contribute significantly to the model. In this model, reading recorded an odds ratio of .839. This indicates that officers were .839 times more likely to not fail FTEP per 1 point (unit) increase in reading score. Officers who passed during FTEP had higher reading scores in the model.

Table 3

Logistic Regression: Reading Score Predicting Training Performance (H1b)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
Reading Score	-.175	.064	7.498	1	.006*	.839	.740	.951

* $p < 0.05$

The results of this model support *H1b*, as lower reading scores in the model did significantly predict FTEP failure. Readings scores are continuous, but the distribution was restricted to seven distinct scores across all 101 subjects, as in H1.

Hypothesis 2: Selection Scores and Sub-scores Scores Predicting Overall Turnover

Hypothesis 2 (H2) anticipated that the three main selection scores (SJT video, writing, and reading) along with the nine SJT sub-scores would offer prediction toward overall turnover. A direct logistical regression was used to assess the impact of the twelve independent variables on overall turnover. While reading was significant and critical thinking approached significance ($p = .053$), the model was not statistically significant, resulting in an unsupported model. H2 was not supported.

The full model containing the three main scores (SJT video, writing, and reading) along with the nine SJT video sub-scores was not statistically significant $\chi^2(12, N = 101) = 13.146, = .358$, indicating that the model was unable to distinguish between officers who left the agency and those retained.

Table 4 reflects the independent variables explored in the model.

Table 4

Logistic Regression: Predictors Officer Turnover(H2)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
SJT Video Score	.063	.067	.872	1	.350	1.065	.933	1.215
Writing Score	.034	.037	.827	1	.363	1.034	.962	1.112
Reading Score	-.172	.066	6.744	1	.009*	.842	.739	.959
Critical Thinking	-1.203	.622	3.736	1	.053	.300	.089	1.017
Investigative	-.049	1.023	.002	1	.962	.953	.128	7.069
Team Orientation	-.251	.401	.391	1	.532	.778	.355	1.707
Confrontation	.083	.5	.027	1	.869	1.086	.407	2.896
Restraint	-.099	.478	.043	1	.837	.906	.355	2.314
Ethical Orientation	.132	.418	.1	1	.752	1.141	.503	2.591
Empathy	.163	.5	1.893	1	.745	1.176	.441	3.136
Org. Orientation	-.734	.533	1.893	1	.169	.480	.169	1.365
Community Relations	-.097	.611	.025	1	.874	.908	.274	3.004

*p<.05

Given the SJT sub-scores are derived from the SJT video score, another analysis was run without the SJT video score. This model was also not supported, $\chi^2(11, N = 101) = 12.262$, $p = .344$, with reading being significant ($p = .009$), and critical thinking approaching significance ($p = .084$).

This attempted exploratory model was challenged with some restricted variables. The nine continuous sub-scores were coded as 1 (risk), 2 (some problem), 3 (normal), and 4 (good). Some of these sub scores had range restrictions similar to the distribution restriction from H1a. Table 4a explores the nine SJT sub-score distribution.

Table 4a

Sub-score Distributions and Range Restrictions H2

	<u>Risk</u>	<u>Some Problems</u>	<u>Normal</u>	<u>Good</u>
Critical Thinking	0	3	78	20
Investigative	0	0	95	6
Team Orientation	5	10	78	8
Confrontation	0	1	69	31
Restraint	1	7	70	23
Ethical Orientation	1	13	69	18
Empathy	0	2	59	40
Org. Orientation	1	12	79	9
Community Relations	1	3	84	13

**Hypothesis 2a: Lower Reading and Critical Thinking Sub-scores Predicting
Involuntary Turnover**

Hypothesis 2a (H2a) theorized that lower reading scores, along with lower critical thinking sub-scores, would predict involuntary turnover. A direct logistical regression was performed to assess the impact of both factors on involuntary turnover. This predictive model did support the hypothesis.

The model contained two independent variables consisting of the reading scores and critical thinking sub-scores. The full model containing both predictor variables was statistically significant $\chi^2(2, N = 101) = 13.454, p < .01$, indicating that the model was able to distinguish between officers who took the selection testing and were involuntarily separated from those who voluntarily left or were retained by the agency. The model as a whole explained between 12.5%

(Cox and Snell R square) and 19.5% (Nagelkerke R square) of variance in officers involuntarily and voluntarily separated, and correctly classified 23.8% (5/21) of officers who involuntarily separated and 98.8% (79/80) of officers not involuntarily separated with an overall correct classification of 83.2 percent.

Table 5 demonstrates that only one of the independent variables, reading scores, contributes significantly to the model. Reading being the only strong predictor for officers leaving the agency recorded an odds ratio of .812. This indicates that officers were .812 times more likely to be separated involuntarily for every 1 point (unit) decrease in reading score. Officers involuntarily separated had lower reading scores controlling for the other factor, critical thinking sub-score, in the model.

Table 5

Logistic Regression: Reading and Critical Thinking Predicting Involuntary

Turnover (H2a)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
Critical Thinking	-.964	.683	1.992	1	.158	.382	.1	1.455
Reading Score	-.208	.068	9.371	1	.002*	.812	.711	.982

*p<.05

The results of this model support *H2a* as lower reading scores in the model did significantly predict involuntary turnover.

Hypothesis 2b: Writing and Select Sub-scores Predicting Voluntary Turnover

Hypothesis 2b (H2b) theorized that higher writing and critical thinking sub scores and lower sub scores in confrontation and organizational commitment will predict voluntary turnover. A direct logistical regression was used to assess the four independent variables on voluntary turnover. The model and H2b are not supported.

The full model consists of the writing score and the three SJT sub-scores of critical thinking, confrontation, and organizational orientation. This model was not statistically significant $\chi^2(4, N = 101) = 3.392, p = .494$, indicating that the model was unable to distinguish between officers who voluntary left the agency and those who did not. Table 6 reflects the independent variables explored in the model.

Table 6

Logistic Regression: Predictors of Voluntary Turnover (H2b)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
Writing Score	.078	.048	2.587	1	.108	1.081	.983	1.189
Critical Thinking	-.277	.571	.234	1	.628	.758	.247	2.324
Confrontation	-.015	.554	.001	1	.978	.985	.332	2.917
Org. Orientation	-.2	.513	.151	1	.697	.819	.3	2.239

The three sub-scores in this exploratory model suffered from range restrictions as reflected in Table 6a. Given these restrictions, the model presents

challenges to the assumptions of logistic regression (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007). Writing scores did approach statistical significance in a one-tailed perspective, $p = .108$, suggesting more robust data from a larger sample may offer future opportunities.

Table 6a

Sub-score Distributions and Range Restrictions H2b

	<u>Risk</u>	<u>Some Problems</u>	<u>Normal</u>	<u>Good</u>
Critical Thinking	0	3	78	20
Confrontation	0	1	69	31
Org. Orientation	1	12	79	9

Supplemental Analyses

Supplemental analyses intended to explore additional models related to both the RQs and to inform practitioners interested in exploring strategies toward improving turnover. The models reported in the supplemental analyses explore the following independent variables.

Reading score less than 90 (Read<90). The continuous reading scores do not have a good distribution and are limited to seven unique scores. Several reading dichotomous recodes (*100, 96.67, less than 93.33, and less than 90*) were created to further explore reading in a more meaningful exploration for practitioners. The lowest band of reading scores ($n=7$) is *Read<90* and coded “0” for no and “1” for yes as belonging to this group.

Candidate retest (Retest). Officers who completed the selection battery more than one time (n=14) make up the subjects in this variable. The variable of officers retesting is labelled *Retest* and coded “0” for no and “1” for yes regarding group membership.

Candidate retest fail then pass (Retest_FP). Officers retaking the selection battery who initially failed (n=7) before passing the battery are explored. The fail then pass retest group is labelled *Retest_FP* and coded “0” for no and “1” for yes indicating group membership.

Prior military experience (Prior Military). This group is comprised of officers with prior military experience (n=31). Officers with experience are labelled *Prior Military* with “0” coded for no and “1” coded for yes toward group membership. Analyses utilizing *Prior Military* has a reduced pool (N=97) as four candidates’ membership status was unknown.

Prior law enforcement experience (Prior LE). Officers taking the selection battery with prior law enforcement experience (n=15) are explored in the supplemental analyses. This group, *Prior LE*, is coded “0” for no and “1” for yes signifying group membership. Analyses utilizing *Prior LE* has a reduced group (n=99) because two officers’ previous law enforcement experience was unknown.

Writing score 100 percent (Writing 100). The continuous writing scores share only seven unique scores, as do reading scores. Several writing dichotomous recodes (*writing 100*, *writing 96.67*, *writing 93.33*, *writing less than 90*, *writing*

less than 82 and writing less than 80) were created to aid in exploratory modeling. The highest writing score (100 percent) is labelled *Writing 100* (n=41) with codes of “0” for no and “1” for yes in group membership.

Restraint sub-score (Restraint). Many of the sub-scores had limited score ranges as previously reflected in Table 4a. The restraint sub-score distributions have scores throughout the group (high risk, n=1; some problems, n=7; normal, n=70; good, n=23) thus offering the better potential to several restricted sub-scores to explore predictive models. This continuous variable is labelled *Restraint* and is intended to assess candidates’ ability to use only necessary and appropriate levels of force to resolve law enforcement situations.

Ethical orientation sub-score (Ethical Orientation). Ethical orientation is another continuous sub-score derived from the SJT video score (see Table 4a). As with *Restraint*, the ethical orientation sub-score offers reasonable distributions (high risk, n=1; some problems, n=13; normal, n=69; good, n=18). This variable is labelled *Ethical Orientation* and is intended to assess a candidate’s ability make ethical decisions.

Supplemental Analysis 1: Read<90 Predicting FTEP Performance Failure

Recall that in hypothesis H1b it was expected that lower reading scores would predict field training performance. The direct logistical regression supported the H1b model. Given the score distribution is limited, as discussed above, it was prudent to explore additional modeling utilizing score “bands” for potential utility

in predicting FTEP failure. Supplemental analysis 1 (S1) explores Read<90 and FTEP failure. Direct logistical regression was deployed to assess the predictive relationship, which resulted in a supported model.

The model contained Read<90 (n=7) as the independent variable exploring prediction of field training performance. The full model was statistically significant $\chi^2(1, N = 101) = 12.047, p < .01$, indicating that the model was able to distinguish between officers who took the selection battery and failed FTEP from those passing FTEP. The model as a whole explained between 11.2% (Cox and Snell R square) and 19.3% (Nagelkerke R square) of variance in officer failure during FTEP and correctly classified 31.3% (5/16) of officers who failed FTEP and 97.6% (83/85) of officers who passed with an overall correct classification of 87.1 percent.

Table 7 demonstrates that Read<90 contributes significantly to the model. In this model, not being in the Read<90 group (-B value) reported an odds ratio of 18.864 (inverted odds ratio per Pallant, 2013) indicating officers were nearly 19 times more likely to pass (not fail) if they had a reading score greater than 90. Given Read<90 is a dichotomous variable (no, yes) there are no units to increase as compared to the continuous reading score explored in H1b.

Table 7

Logistic Regression: READ<90 Predicting Training Performance (S1)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							<u>Lower</u>	<u>Upper</u>
READ<90	-2.937	.896	10.744	1	.001*	18.864	3.257	109.237

*p<.01

As supported in H1b, officers who failed during FTEP had lower reading scores in the model. The S1 model using Read<90 offers added predictability of FTEP failure than the H1b statistically, albeit small, (H1b, $p < .005$; S 1, $p < .001$) and in a practical sense. The S1 model predicted two additional officers those who failed FTEP (H1b 3/16; S1 5/16) then H1b. Both identified the same officers who passed (83/85). It is important to acknowledge that the population is relatively small and additional agencies and subjects would improve the power of the analyses. Interestingly, of the seven officers with Read<90, five failed FTEP, one was terminated for discipline during FTEP, and one was not retained.

Supplemental Analysis 2: READ<90 and Retest FP Predicting FTEP Failure

This supplemental analysis (S2) model adds Retest_FP to READ<90 to explore FTEP failure. Officers retaking the selection battery (n=14) fell into two outcome groups. The “pass, pass” group retested for various reasons including an attempt to improve one or more scores or the fact that time lapsed within a hiring window and the candidate needed to retest. The seven who failed then passed are the Retest-FP group. The direct logistical regression supported this model. Only

READ<90 is statistically significant, with Retest_FP approaching significance in a two-tailed model. In a one-tailed model both READ<90 and Retest_FP would be significant.

The S2 model explored Read<90 with Retest_FP and FTEP failure. The direct logistical regression used to assess the predictive relationship resulted in a supported model. The model contained Read<90 and Retest_FP as independent variables exploring prediction of FTEP failure. The full model was statistically significant $\chi^2(2, N = 101) = 14.715, p < .01$, indicating that the model was able to distinguish between officers who took the selection testing who failed FTEP and those who passed FTEP. The model as a whole explained between 13.6% (Cox and Snell R square) and 23.3% (Nagelkerke R square) of variance in officer success during FTEP and correctly classified 31.3% (5/16) of officers who failed FTEP and 97.6% (83/85) of officers who passed training with an overall correct classification of 87.1 percent.

Table 8 demonstrates that READ<90 contributes significantly to this model with Retest_FP approaching (two-tailed) significance. In this model, not belonging to the Read<90 group (-B value) reported an odds ratio of 18.867 indicating officers were almost 19 times more likely to pass (inverted odds ratios per Pallant, 2013) if they had a reading score higher than 90 in this model.

Table 8

Logistic Regression: READ<90 and RETEST_FP Predicting Training Failure (S2)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
Read<90	-2.945	.914	10.378	1	.001*	18.867	3.16	111.111
Retest	-1.557	.898	3.007	1	.083	4.739	.816	27.777

*p<0.01

The S2 model adds additional prediction above READ<90 toward FTEP outcomes as reflected in the variance explained. The S2 model explains 13.6%, opposed to 11.2% in S1, (Cox and Snell R square) and 23.3%, compared to 19.3% from S1, (Nagelkerke R square). Additional agencies and subjects to future analyses may produce more power and enhance the role of Retest_FP. Additional future recommendations and applied implications will be explored in chapter 5.

Supplemental Analysis 3: Predictive Model of Overall Turnover.

Supplemental analysis 3 (S3) revisits predictors of overall turnover initially explored in H2. H2 anticipated that the three main selection battery scores along with the SJT sub-scores would offer prediction of overall turnover. The H2 model was not supported. The S3 model contains the continuous writing and readings scores from the H2 model, with repeated selection battery, prior law enforcement experience, and prior military experience to explore potential prediction of overall turnover. S3 deploys a direct logistical regression and resulted in a supported model.

The model contained *Writing*, *Reading*, *Repeat*, *Prior LE*, and *Prior Military* as independent variables exploring prediction of overall turnover. The full model was statistically significant $\chi^2(5, N = 97) = 14.291, p < .05$, indicating that the model was able to distinguish between officers who took the selection testing who left and those retained. The model as a whole explained between 13.7% (Cox and Snell R square) and 18.5% (Nagelkerke R square) of variance in officer's leaving and correctly classified 42.5% (17/40) of officers who left and 82.5% (47/57) of retained officers with an overall correct classification of 66 percent.

Table 9 demonstrates that reading scores contribute significantly to this model. *Prior LE* approached (two-tailed) significance. In this model, reading recorded an odds ratio of .851. This indicates that officers were .851 times more likely to leave for every 1 point (unit) decrease in reading score. Officers who left the agency had overall lower reading scores in the model than those retained.

Table 9

Logistic Regression: Predictors Officer Turnover (S3)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
Writing Score	.044	.039	1.271	1	.260	1.045	.968	1.128
Reading Score	-.161	.067	5.741	1	.017*	.851	.746	.971
Repeat	-.925	.681	1.842	1	.175	.397	.104	1.508
Prior LE	1.391	.760	3.849	1	.067	4.019	.906	17.827
Prior Military	-.172	.473	.131	1	.717	.842	.333	2.129

* $p < 0.05$

While Prior LE only approached two-tailed significance ($p = .067$) it may

be of value to practitioners. Officers with prior experience (n=15) were retained 12/3. The model suggests that officers with prior experience are 4 times more likely to stay than officers without prior experience. Additional agencies and subjects would be necessary to add power with the potential to revisit the potential role Prior LE has on retention. Additional recommendations for future research and potential applied utility will be discussed in chapter 5.

Supplemental Analysis 4: READ<90 and Repeat Predicting Involuntary Turnover

Supplemental analysis 4 (S4) revisits prediction of involuntary turnover previously explored on H2a that explored lower reading and critical thinking scores. The H2a model was supported with reading being statistically significant in the model. The H2a model explained between 12.5% (Cox and Snell R square) and 19.5% (Nagelkerke R square) of involuntary turnover variance. Given the restricted range in critical thinking sub-scores (see Table 4a) exploring other potential predictors, in addition to low reading, may produce applied utility. S4 explores READ<90 (n=7), as opposed to the previous reading continuous score, and repeated selection battery (Retest, n=14) in this model. A direct logistical regression was performed to assess the impact of both factors on involuntary turnover resulting in a supported model where both dichotomous factors significantly contribute.

The S3 model with READ<90 and Repeat exploring potential prediction for

involuntary turnover. The full model containing both predictor variables was statistically significant $\chi^2(2, N = 101) = 18.977, p < .001$, indicating that the model was able to distinguish between officers who took the selection testing and experienced involuntary turnover and those who did not (left voluntarily or retained). The model as a whole explained between 17.1% (Cox and Snell R square) and 26.8% (Nagelkerke R square) of variance in officers leaving the agency involuntarily, and correctly classified 28.6% (6/21) of officers who involuntarily separated as well as 98.8% (79/80) of officers who left voluntarily or retained with an overall correct classification of 84.2 percent.

Table 10 demonstrates that both independent variables contribute significantly to the model. READ<90 (-B value) for officers not leaving the agency involuntarily recorded an odds ratio of 25.641 indicating officers were more than 25 times more likely to not be involuntarily separated (inverted odds ratios per Pallant, 2013) if their reading score was more than 90. RETEST (-B value) resulted in an odds ratio of 4.081 indicating that officers who did not retest were four times more likely (inverted odds ratios per Pallant, 2013) to not be involuntarily separated. Officers not involuntarily separated, in this model, had higher reading and were less likely to have retaken the selection battery test than those who were involuntarily separated.

Table 10

Logistic Regression: READ<90 and Retest Predicting Involuntary Turnover (S4)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							<u>Lower</u>	<u>Upper</u>
READ<90	-3.251	1.141	8.116	1	.004	25.641	2.754	250
RETEST	-1.407	.680	4.289	1	.038	4.081	1.078	15.384

* $p < 0.05$

The S4 model provided an improved explanation of the variance; 17.1% over 12.5% (Cox and Snell R square) and 26.8% over 19.5% (Nagelkerke R square) to that of H2a, which explored reading scores and critical thinking. Correct classification was improved with successful prediction of one additional officer (6 vs 5) involuntarily separated. Potential applied utility will be discussed in chapter 5.

Supplemental Analysis 5: Predictive Model of Officers Requiring Remedial Training

This supplemental analysis (S5) is exploring an aspect important to practitioners charged with facility transfer of training to newly hired officers. Officers not completing the FTEP process account for 28 percent of overall turnover (28/101) The majority of those not completing FTEP are due to failure (n=16). Interestingly, many without performance issues simply change their mind about law enforcement during FTEP (n=9). Three officers were disciplinary terminations during FTEP. These issues related to officers not successfully clearing FTEP are worthy to explore. S5 explores officers needing remedial training, the

most common issue with turnover consequence during the FTEP process.

This S5 explores officers requiring remedial training. There were 23 officers requiring remedial training during FTEP. Thirteen officers ultimately failed with seven officers successfully completing FTEP. This analysis explores the independent variables of READ<90, writing scores of 100 (Writing 100), a restraint sub-score of good (Restraint_Good), team orientation sub-score, ethical orientation sub-score, organizational orientation sub score, community relations sub-score, and prior military experience to explore any predictive nexus to officers requiring remedial training. The direct logistical regression of these factors on the need for remedial training resulted in a supported model. The model was supported with READ<90 as statistically significant in the model. Writing 100 and Restraint_Good approached statistical significance (two-tailed) in the model.

The S5 model with the dichotomous variables of READ<90 (n=7), Writing 100 (n=41), Restraint_Good (n=23), Prior Military (n=31), and the continuous sub-scores of team orientation, ethical orientation, organizational orientation and community relations along with the need for remedial training was explored. The full model containing all predictor variables was statistically significant $\chi^2(8, N = 97) = 17.986, p < .05$, indicating that the model was able to distinguish between officers who took the selection testing and needed remedial training and those who passed FTEP remedial training. The model as a whole explained between 18% (Cox and Snell R square) and 29.3% (Nagelkerke R square) of variance in officers

requiring remedial training and correctly classified 33.3% (6/18) of officers who required remediation and 96.2% (73/76) of officers who passed FTEP without remedial training with an overall correct classification of 84.5 percent.

Table 11 demonstrates that READ<90 contributes significantly to the model, and Writing 100 and Restraint_Good each approached significance (each at $p = .059$). READ<90 (-B value) for officers not requiring remedial training recorded an odds ratio of 20.408 indicating officers were more than 20 times more likely not to need remedial training (inverted odds ratios per Pallant, 2013) if their reading score is more than 90. This suggests that reading scores above 90 are a protective factor regarding the need for remedial training. The variable of Writing 100 ($p < .059$) resulted in an odds ratio of 3.556 indicating that officers with a writing score of 100 were roughly 3.5 times more likely to require remedial training. Restraint ($p < .059$) had a negative B value with an odds ratio of 2.86 indicating that officers were 2.86 times more likely to not need training (inverted odds ratios per Pallant, 2013), suggesting that it is also a protective factor. This model suggests that there is a greater likelihood that officers can avoid remedial training with the protective factors of reading above 90 percent and the upper bound restraint score (good) and will more likely need remedial training if they scored perfectly in writing.

Table 11

Logistic Regression: Predictors of Remedial Training (S5)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
READ<90	-3.008	.987	9.281	1	.002*	20.408	2.923	142.857
Writing 100	1.287	.682	3.556	1	.059	3.621	.951	13.793
Restraint_Good	-1.403	.743	3.563	1	.059	2.863	.947	17.543
Prior Military	-.875	.697	1.577	1	.209	.417	.106	1.634
Team Orientation	.109	.574	.036	1	.849	1.115	.362	3.435
Ethical Orientation	-.362	.5	.524	1	.469	.696	.261	1.856
Org. Orientation	-.983	.622	2.494	1	.114	.374	.111	1.267
Community Rel.	.969	.787	1.517	1	.218	2.635	.564	12.314

*p<.01

This analysis only had 97 (of 101) subjects, as four officer's data regarding prior military service was unknown. Discussion on the Writing 100 group and higher restraint scores will be discussed in the context of potential applied utility in chapter 5.

Supplemental Analysis 6: Predicting Officers Failing Remedial Training

It may be helpful to the field to distinguish those requiring remedial training that pass and fail. This supplemental analysis explores the officers that fail remedial training. This supplemental analysis (S6) searches for a predictive model of remedial training failure. The model analysis explores the independent continuous variables of SJT video and reading scores with dichotomous variables of Writing 100, Repeat_FP, Prior Military, and Prior LE to explore for a predictive nexus to remedial training failure. The model was supported with reading and repeating the

selection battery with initial fail being statistically significant in the model. Writing 100 approached statistical, two-tailed, significance at $p = .081$.

The S6 model SJT video, reading, Writing 100, Repeat_FP, Prior Mil, and Prior LE along with remedial training failure was explored. The full model containing all predictor variables was statistically significant $\chi^2(6, N = 97) = 17.454$ $p < .01$, indicating that the model was able to distinguish between officers who took the selection testing and needed remedial training and those who did not require remedial training during FTEP. The model as a whole explained between 16.5% (Cox and Snell R square) and 32.5% (Nagelkerke R square) of variance in officers leaving the agency involuntarily and correctly classified 27.3 (3/11) of officers who failed remediation, and 96.5 (83/86) of officers who did not need remedial training with an overall correct classification of 88.7 percent.

Table 12 demonstrates that two independent variables, reading and Retest_FP, contribute significantly to the model. In this model, reading recorded an odds ratio of 1.261. This indicates that officers were .793 times more likely to pass remedial for every 1 point (unit) increase in reading score. Retest_FP (-B value) recorded an odds ratio of 11.111 indicating that officers were more than 11 times less likely to fail remedial training if they did not fail the selection battery initially. This model suggests that officers with lower reading scores are more likely to fail remedial training as are those who failed on their first selection test attempt.

Table 12

Logistic Regression: Predictors of Remedial Training Failure (S6)

	<u>B</u>	<u>S.E.</u>	<u>Wald</u>	<u>df</u>	<u>p</u>	<u>Odds Ratio</u>	<u>95% C.I. for EXP(B)</u>	
							Lower	Upper
SJT Video	.064	.071	.809	1	.369	1.066	.927	1.226
Reading Score	-.232	.081	8.252	1	.004*	.793	.667	.929
Writing 100	1.674	.958	3.051	1	.081	5.334	.815	34.903
Prior Military	-1.234	.798	2.393	1	.122	.291	.061	1.39
Prior LE	-.2	1.098	.033	1	.856	.819	.095	7.050
Repeat FP	-2.403	1.184	4.119	1	.042*	11.111	1.085	111.111

*p<.05

This analysis only had 97 (of 101) subjects as four officers' data regarding prior military service was unknown. Writing 100, which approached two-tailed significance, will be further discussed in chapter 5 along with other applied utility opportunities.

Synthesis and Summary of Data

The goal of this exploratory study was to search for potential predictive nexuses between selection battery tests and turnover outcomes. Specifically, do law enforcement SJT selection batteries offer predictive utility toward field training performance (RQ1) and turnover outcomes (RQ2). This research is of interest to the study agency in their constant search for strategies to improve field training outcomes and retention. These findings may offer empirical insights to impact interventions toward improving success rates of training and other retention related efforts.

Specific findings. The study population was limited to officers hired at the mid-sized Florida agency who were selected, in part, using scores from the SJT selection battery deployed in October 2012. This limits the scope of study for the agency population. The sample does not include any transfer of training or retention data for officers selected prior to the battery first deployed in October 2012. Officers hired as early as 1989 have yet to retire.

Context relevant limitations related to findings interpretations. Functional turnover, such as retirement, was not part of the study. Only one officer from the study population retired at retirement age. Turnover related to shocks, where external events *push* or *pull* employees to leave (Burton, Holtom, Sablinski, Mitchell, & Lee, 2010), were also limited by the sample. The scope of the study is a limitation as only recent agency members (since October 2012) are included.

While the population sample only explores officers hired during the past seven and a half years, it does offer a good snapshot of contemporary recruits for exploring field training failure. While the sample size is relatively small for exploratory analysis, which is not uncommon in applied studies, it is sufficient and offers a conservative approach as power is lower than a more robust population, and due to the use of a two-tailed test of significance.

Field training performance. Hypothesis 1 explored SJT video and reading scores in a model to test prediction of passing field training. Reading was statistically significant ($p = .006$) in the model shared with the SJT video score.

The H1 model (Table 1) suggests that officers are 1.19 times more likely to pass for every point increase in reading score. For example, an officer scoring 96.67 is 7.93 times more likely to pass than an officer scoring 90. The same officer is nearly 20 (19.83) times more likely to pass than an officer at the minimum cut score of 80.

Hypothesis 1b explored reading as a sole predictor related to field training failure. That model was supported, and reading was statistically significant ($p = .006$) in predicting training failure. The H1b model (Table 3) suggests that officers are .839 times more likely to fail for every point decrease in reading score. For example, an officer scoring 90 is 5.59 times more likely to fail than an officer scoring 96.67. An officer scoring 80, the minimum cut score, is nearly 14 (13.98) times more likely to fail than an officer scoring 96.67.

It is clear reading scores predict field training performance. Additional models were explored in supplemental analyses (S1& S2) regarding reading scores' prediction for training failure. These analyses may prove more beneficial for practitioners. Recall the recode of the seven distinct scores for reading (100, 96.67, 93.33, 90, 86.67, 83.33, and 80) into reading score "bands" of *100*, *96.67*, *less than 93.33*, and *less than 90*.

The S1 model (Table 7) explored the relationship between having a reading score of less than 90 percent (Read<90) and field training failure. The supported model showed reading as statistically significant ($p = .001$). This showed that a reading score above 90 is a protective factor against failure where

officers scoring higher than 90 were 18.86 times more likely not to fail than those scoring less than 90. The model also had improved statistical significance compared to H1b, as well as positively predicting two additional officers in the failed group.

The S2 model (Table 8) explored whether having a reading score of less than 90 percent (Read<90) and having a selection battery retest, where the officer initially failed, (Retest_FP) predicted field training failure. This model was supported with Read<90 being statistically significant ($p = .001$) and Retest_FP approaching statistical significance ($p = .083$). This model interacted similarly to the S1 model where it showed that reading above 90 is a protective factor against failure, where officers scoring higher than 90 were 18.86 times more likely not to fail than those scoring less than 90. Retest_FP, while only approaching two-tailed significance, suggests the potential of an additional protective factor for passing. Retest_FP suggest that officers who did not retest with an initial failure were 4.73 less likely to fail than those who did fail their initial attempt.

Field training outcomes: Bottom line. The findings related to predicting field trading performance is that reading scores predict. Both the continuous reading scores and banded reading less than 90 predict failure and success in field training. There is some support that officers retesting after an initial selection battery failure may also have a predictive nexus.

Turnover. Hypothesis 2 posited that all three SJT selection battery main

scores, SJT video, writing, and reading, along with all SJT sub-scores would offer prediction of overall turnover. The model was not supported although reading remained statistically significant. The H2 model (Table 4) had a challenge of range restriction of some of the sub-scores (see Table 4a).

Hypothesis 2a explored lower reading scores and the critical thinking sub-score to predict involuntary turnover. Despite range restriction in critical thinking, the model was supported. The Reading score was statistically significant ($p = .002$) and the model was supported. The H2a model (Table 5) suggests that officers are .812 times more likely to be involuntarily separated for every point decrease in reading score. According to this model, an officer with a score of 90 is 5.41 times more likely to be involuntary separated than an officer scoring 96.67. An officer with the minimum cut score of 80 is over 13 (13.53) times more likely to be terminated.

The H2b model (Table 6) considered the writing score and the sub-scores of critical thinking, confrontation, and organizational commitment as a predictor model of voluntary turnover. Several models predicting voluntary turnover were explored with no supported models identified. Overall turnover was revisited in supplemental analyses.

Supplemental analysis 3 revisited overall turnover with potential predictors of writing score, reading score, repeated selection battery, prior law enforcement experience, and prior military experience resulting in a supported model. Reading

again proved statistically significant ($p = .017$) with Prior_LE approaching two-tailed significance ($p = .067$).

The S3 model (Table 9) suggests that officers are .851 times more likely to leave for every point decrease in reading score. Continuing with the previous examples, an officer scoring 90 is 5.67 times more likely to leave than an officer scoring 96.67. An officer passing with the minimum score (80) is 14.18 times more likely to leave compared to an officer scoring 96.67. Prior law enforcement experience approached two-tailed significance ($p = .067$) and may possibly be a protective factor. The study suggests Prior LE may make officers with prior experience 4.019 times less likely to turnover for any reason.

Involuntary turnover was revisited in the S4 model. S4 explored reading scores less than 90 percent (Read<90) and officers who repeated the selection battery involuntarily, regardless if the initial attempt was failed or passed (Repeat), and turnover. This model was supported with both variables, Read<90 ($p = .004$) and Repeat ($p = .038$), being statistically significant.

The S4 model (Table 10) suggests that reading scores above 90 are a protective factor against involuntary turnover. This suggests that officers scoring higher than 90 were 25.64 times more likely not to be terminated than those scoring less than 90. The model also suggests that Retest is a protective factor, while also suggesting that officers who did not retest are 4.08 times less likely to involuntarily separate than those who did not retest.

Turnover outcomes: Bottom line. Reading scores, as with training outcomes, have a predictive nexus with overall turnover and involuntary turnover. Continuous reading scores and banded reading scores less than 90 predict turnover outcomes, as well as failure and success in field training. Officers retaking the selection battery impacts involuntary turnover. Potential support exists that officers with previous law enforcement experience are shielded from overall turnover.

Remedial training. The two final supplemental analyses explore officers requiring remedial training (S5) and those who failed remedial training (S6). Police agencies, field training officers, and their supervisors invest substantial resources and time in preparing officers for solo duty. Officer performance deficiencies result in mandated remedial training (n=23) with most subsequently failing (n=13).

The need for remedial training. The S5 model exploring officers requiring remedial training included nine independent variables and was supported. READ<90 was statistically significant ($p = .002$), while writing score 100 and restraint sub-scores in the “good” category each approached (two-tailed) significance at ($p = .059$).

The S5 model (Table 11) suggests that reading scores above 90 are protective against needing remediation. This suggests that officers scoring higher than 90 were 20.408 times more likely not to need remediation than those scoring less than 90. This model also suggests that perfect writing scores (Writing 100)

may lead to the need for remedial training at 3.621 times more than officers scoring less than 100 percent on the writing test. The SJT restraint sub-score of “good” may also be a protective factor in that officers scoring “good” in restraint are 2.863 times more likely to avoid remedial training in the model.

Remedial training failure. The final supplemental analysis in the study (S6) explored officers failing remedial training and produced a supported model. Reading scores ($p = .004$) and repeated selection battery with initial failure ($p = .042$) were statistically significant in the model. Writing 100 approached significance ($p = .081$), as a potential predictor consistent with the need for remedial training from S5.

The S6 model (Table 12) suggests that officers are 1.261 times more likely to fail remedial training for every point decrease in their reading score. Using the continuing example, an officer scoring 90 is 8.41 times more likely to fail remedial training than an officer scoring 96.67. An officer with a reading score of 80, the minimum passing score, is 21.02 times more likely to fail remedial than an officer scoring 96.67. Officers not retaking the selection battery following an initial failing score have a protective factor and are 11.111 times less likely to fail remedial training than those who did retest following a failure.

Officers with writing scores of 100 ($n=41$) approached two-tailed significance ($p = .081$). As in S5, Writing 100 may indicate an increased likelihood of failing remedial training at 5.334 times more than officers scoring

less than 100 percent on the writing test.

Remedial training outcomes: Bottom line. The need for remedial training and those officers who were failing remedial training were explored in the models. As with training and turnover outcomes, reading proved to offer the best and most consistent predictor related to remedial training performance. Those officers who initially failed the selection battery were unlikely to pass remedial training.

Perfect writing scores of 100 percent and “good” restraint sub-scores approached significance and warrant additional inquiry. “Good” ratings in restraint as a protective factor against the need for remedial appears logical if selection battery factors tapped into positive performance outcomes. Writing scores of 100 (41/101) requires additional study. It is possible that the writing battery taps attention-to-detail issues that, in excess, can hamper pragmatic process. It is also possible the relationship is spurious based on the group size and psychometric properties of the test.

Several models exploring prediction for remedial training success were not supported. Given the group that passed ($n=7$), the study may be likely lacking predictive power. Discussions related to future research in the following chapter will include recommendations for studying successful officers in remedial training.

Findings Informing the Problem Statement.

The research questions for this study inquired into whether SJT selection batteries offer prediction to field training performance and turnover outcomes. Although not all of the hypotheses models were supported, those that were, along with the supplemental analyses, do affirmatively respond. This research offers new insights regarding selection instruments and prediction of actual turnover (voluntary and involuntary), including field performance failure. Immediate applied utility may be possible. For example, utilizing officer's reading scores can assist in offering them additional training prior to starting the FTEP, if they have scores below 90.

Contribution to applied practice. This research is of interest to the agency observed as well as law enforcement administrators in general. In an environment where law enforcement continues to struggle with qualified applicants, the quest for proactive strategic initiatives is ongoing. Interventions and strategies to facilitate improved training outcomes and retention of new officers saved agencies valuable resources, in both time and finances. Selection process decisions, training officer assignments, and proactive pre-FTEP training programs are potential data-driven initiatives.

Applied practice contributions. The potential impact of this exploratory study includes training opportunities and selection process policy review. Given that selection battery information exists and is available, data can offer added

utility in exploring strategies. It is important, however, to recognize and discuss the reality that any intervention or strategy must consider potential unintended consequences. These will be discussed with each possible contribution.

It is also critical, in an applied context, that this study be used as an initial exploratory study to establish a foundation for additional study. With respect to this critical consideration, immediate applied practice contributions are purposefully limited. The potential for future applied contributions will be further explored in chapter 5.

Pre-FTEP training opportunities. Reading scores predict outcomes in this study. Officers with reading scores less than 90 ($n=7$) are readily identifiable prior to starting field training. All seven officers in this study with reading scores below 90 are no longer with the agency. Six of the seven did not make it out of FTEP. Five officers failed, all recording critical report writing deficiencies. One officer was fired for discipline. The one that completed FTEP left after one year.

The reading cut score is set by the vendor. It would not be advised to consider raising the cut score without additional research, including adverse impact concerns. It is, however, prudent to explore creating a structured pre-FTEP training program for officers scoring 90 or below. Initial steps would include a focus group with field training officers to explore critical performance deficiencies and explore pedagogically sound training. Report writing, for example, is an identified performance deficient area that training may explore.

Selection battery retest. Both retaking the test and taking the test with an initial fail were either statistically significant, or approached significance, in several models. Fourteen officers retook the exam with only five retained. Additional exploration of the three test scores (SJT video, writing, and reading) among the retest officers is warranted. Data suggests that those with low reading scores that retook the exam did not succeed in training. A search for more specific interactions may produce better policy frameworks on determining retest opportunities.

Changing mind about law enforcement. Nine officers did not complete field training because they changed their mind about law enforcement while in FTEP. None had recorded performance deficiencies in their DORs. In comparison, two officers changed their minds about law enforcement and left post FTEP with three changing professions. While exploratory modeling failed to identify prediction to voluntary turnover in this study, the fact that nine left stating they changed their minds, compared to 16 who failed training, warrants review.

An initial recommendation is for the agency to review any literature or media related to *realistic job previews (RJPs)*. When deployed correctly, RJPs can reduce turnover (Griffith et al., 2002; Dalessio, 1994; Jones et al., 2002; Scott & Jones, 2003). Additionally, exploring ride-along initiatives for candidates may prove worthwhile. The ultimate goal is to help candidates assess if law enforcement is a good career (P-J) or organization (P-O) fit prior to costly

training.

Chapter Summary.

This study was successful in establishing that SJT selection battery scores do have a predictive nexus to field training and turnover outcomes. While some hypotheses did not result in supported exploratory models in binary logistic regression analyses, the supported hypotheses and supplemental analyses lay the groundwork to explore SJT selection battery scores, particularly reading, thus meeting the goal to inform the gap in literature. Chapter four reported and discussed the analyses and findings, offered a synthesis of these findings, and discussed applied application considerations.

Chapter 5

Discussion, Implications, and Recommendations

Overview

Chapter 5 is intended to further discussion of the study results, study implications, and empirical and applied recommendations for future study. This final chapter will commence with an overview of the applied problem, review major literature that serves as the study foundation, as well as provide a summary of findings and implications.

The chapter continues with exploring the potential theoretical contributions, as discussed in chapter one, and will be revisited with the study's implications on those expected contributions. Contributions to literature and practice will be explored, with a focus on findings that support and differ from existing theory. The chapter continues by exploring conclusions and implications related to informing literature, potentially influencing application, and discussing limitations and boundary conditions. Recommendations for applied practice as well as future research will precede concluding the discussion.

Applied research problem. The research problem explored in this study focuses on the public sector turnover issue in police officers. Issues related to police turnover include recruiting, selection, training, and retention (Cavanaugh, 2013; Haarr, 2001; Jordan et al., 2009; Koper et al., 2001; Maciag, 2018; National Research Council, 2004; Orrick, 2008; PERF, 2013; Selden et al., 2013; CSLGE,

2017; Smith, 2016; Wilson, 2012; Wilson, 2014; Wolfe & Nix, 2016; Woska, 2006; Woska, 2016). Various dynamic challenges, such as economic, social, and workforce issues make understanding and responding to this threat to the capabilities of most (small and mid-sized agencies) police agencies increasingly challenging (Jordan et al., 2009; Orrick, 2008; Smith, 2016; Wilson, 2014; Woska, 2016).

Law enforcement is a high consequence industry (Barrett et al., 2010) where the cost of turnover extends beyond the immediate resource loss (Barrick & Zimmerman, 2009; Hom et al., 2017; Maltarich, Nyberg, & Reilly, 2010; Van Iddekinge et al., 2011). Ployhart and Weekley (2010) discussed the critical learning curve required in complex roles, such as policing, that is needed for an officer to contribute fully to the agency.

Organizations must consider the cost of onboarding and training new employees. The average employee, regardless of industry or sector, has a learning curve before the new employee fully contributes to the organization. This learning curve takes even longer in complex roles, such as policing (Ployhart & Weekley, 2010). Diminishing applicant pools compound the turnover issue for agencies (Barrett et al., 2010; Cavanagh, 2003; Haarr, 2001; Koper, Maguire, & Moore, 2001; Orrick, 2008; Selden et al., 2013; Smith, 2016; Wilson, 2012; Woska, 2016).

There are significant workforce challenges facing policing, such as labor market conditions, generational differences, norms related to shorter lengths of

employment, and job mobility (Chen & Chang, 2010; Dess & Shaw, 2001; Hom et al., 2017; Kacmar et al., 2006; Mabon, 1994; Ployhart, 2006; Ployhart & Weekley, 2010). Recruiting, selection, and training challenges related to turnover result in costly, and dysfunctional, turnover to police ranks. This dysfunctional turnover forces agencies back to searching for qualified candidates in a diminishing candidate pool. Returning to this pool in a quest for quality applicants is especially problematic for small and mid-sized law enforcement agencies that comprise the majority of U.S. agencies.

SJT selection batteries. SJTs have been researched related to employee selection resulting in substantial empirical support in predicting job performance (Clevenger, Pereira, Wiechmann, Schmidt, & Harvey, 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Weekley & Jones, 1999). Chan and Schmitt (2002) and Weekley and Ployhart (2005) suggested that SJTs have higher selection validity over other common factors, such as cognition, personality, and work experience. In addition to improved predictive validity in job performance, SJTs have demonstrated less adverse impacts than previously mentioned factors (Clevenger, Pereira, Wiechmann, Schmidt, & Harvey, 2001; Motowidlo & Tippins, 1993; Pulakos & Schmitt, 1996; Weekley & Jones, 1999).

Study goals toward addressing the problem. This study was intended to specifically explore SJT selection battery scores, and other candidate attributes, toward predicting field training performance and other turnover outcomes

specifically in a law enforcement organization. The quest for using existing candidate data strategically to better understand field training and retention outcomes has potential value-added opportunities for agencies to proactively retain officers. These types of data-driven exploratory initiatives are challenging to most law enforcement agencies where budgets are constrained. Opportunities to expand the utility of existing data may inform strategies toward improved selection, recruiting, and retention practices.

Major literature themes. The overarching theoretical framework guiding this study is the attraction-selection-attrition (ASA) model (Schneider, 1987) with additional related support from Edward's (1991) person-job (P-J) fit and Judge and Ferris's (1992) person-organization (P-O) fit. The other major literature themes comprising the theoretic model for the study are situational judgment tests (Annell et al., 2015; Lievens & Motowidlo, 2016; Whetzel & McDaniel, 2009), transfer of training (Baldwin & Ford, 1998), turnover (Bills, 1925), voluntary versus involuntary turnover (Price, 1977), functional (Dalton et al., 1982), dysfunctional (Dalton et al., 1982), and optimal turnover (Abelson & Baysinger, 1984), and the unfolding model of turnover including shocks (Lee & Mitchell, 1994).

Attraction-selection-attrition model. Schneider's (1987) ASA model is an extension of Lewin's (1936) behavior theory equation, which theorized that behavior is a function of the person and environment, $B = f(P, E)$. Schneider (1987) extends Lewin's behavior theory by exploring the organizational dynamics

of selection. Schneider (1987) included the *actors* who behave within Lewis' notion of environment. According to Schneider (1987), the ASA model views these actors as critical to selection of new members. Rather than new members being randomly attracted or selected, ASA recognizes selection based on criteria established by existing members (Schneider, 1987).

The ASA cycle of attraction, selection, and attrition provides a framework to explore the organizational culture and dynamics associated with turnover. Schaubroeck et al. (1998) contended that as people are attracted, selected, and remain in an organization they become actors who shape the factors that help determine the environment. In the ASA model poor perceived matches, such as bad fits, lead to attrition (De Cooman et al., 2009; O'Reilly, Chatman, & Caldwell, 1991; Schneider et al., 1998).

Schaubroeck et al. (1998) further discussed the reality of nature for people to have an attraction to others with similar interests, attitudes, and values. This attraction of the various factors considered "similar" can lead to homophily of members. McPherson et al. (2001) cautioned that organizations must be aware of *homophily bias* where group, culture, and behaviors among organizational members become unhealthy to organizational capabilities.

Job fit. Ployhart (2006) discussed the importance of *fit* between individuals and jobs or organizations, and turnover outcomes. Some turnover is simply a result of poor work performance or failures in the transfer of training for new members.

Other employees leave for reasons unrelated to performance. These are considered, in the ASA model concept, as either poor person-job (P-J) or person organization (P-O) fit.

Person-job fit. Edwards (1991) operationalized the fit between individual members' KSAOs and the job or industry as P-J fit. P-J fit factors are important in proper job task analysis processes and subsequent selection, training, promotion, and retention policies and practices (Breugh & Starke, 2000; Kristof, 1996; Kristof-Brown, 2002; Werbel & Gulliland, 1999). Lough and Von Treuer (2013) point to P-J fit as critical for optimal selection in high stakes professions. Poor P-J fit can result in reduced job performance, lower organizational commitment, poor job satisfaction (Kristof-Brown, Zimmerman, & Johnson, 2005) and job burnout (Maslach & Leiter, 1997).

Person-organization fit. Judge and Ferris's (1992) theory of P-O fit discusses how new employees are accepted by other members in the organizational. Piasentin and Chapman (2006) observed the strong theoretical support of P-O fit from Schneider's (1987) ASA model. For example, Kristof (1996) and Verquer et al. (2003) explored P-O fit and reported findings on the importance of shared characteristics and retention.

P-O fit literature (Piasentin & Chapman, 2006) identified four common operationalizations of P-O fit. *Supplementary* P-O fit indicates when a person has similar characteristics to organizations. *Complementary* P-O fit is operationalized

as instances where a person fills an organization's vacant need. *Needs-supplies* P-O fit is defined as instances when the organization fulfills the needs of an individual. Lastly, *demands-abilities* P-O fit describes fit as when an individual's abilities meet demands of an organization (Piasentin & Chapman, 2006).

Situational judgement tests. Whetzel and McDaniel's (2009) meta-analytic study reviewed the popularity of SJTs in the U.S. and Europe. They pointed to empirical support for SJTs including improved predictive validity related to job performance when compared to selection options, such as personality tests and cognitive ability (Christian, Edwards, & Bradley, 2010; Whetzel & McDaniel, 2009).

SJTs offer organizations another critical selection advantage as they are shown to produce less adverse impact than other options. This allows agencies to deploy SJT selection batteries that are more likely to produce results aligned with employment case law and EEOC guidelines (Barrett et al., 2010; Callanin & Robertson, 2000; Chan & Schmidt, 1997; Motowidlo & Tippins, 1993; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Weekley & Jones, 1997; Whetzel, McDaniel, & Nguyen, 2008).

Previous police SJT and related studies. Anell et al. (2015) studied Swedish police recruits (n=750) to explore the relationship between selection tests, cognitive ability, personality inventories, and fitness, and if they prove predictive for job performance, job satisfaction, retention, and health. Their longitudinal study

measured outcomes after the police academy, after field training (6 months), and after one year on the job (Annell et al., 2015).

Their findings suggest that verbally loaded cognitive ability tests predicted academic and work performance (Annell et al., 2015). Additionally, general cognition scores negatively predicted job satisfaction over all three observation periods on the study (Annell et al., 2015). Annell et al. (2015) findings included prediction between personality tests job satisfaction, retention, and health but not predictive of job performance, the most common criticism for using personality in selection batteries.

Transfer of training. Baldwin and Ford (1998) operationalized transfer of training as the generalization, retention, or enhancement of job relevant KSAOs acquired as a function of training. Goldstein and Ford (2002) argued that exploring the relationship between the transfer of training and actualized performance should be the *paramount goal* of training functions in any organization.

A transfer of training metaanalysis (Bloom et al., 2010) explored various predictive factors of transfer. Their metaanalysis suggested that an empirical identification of a “magic bullet” resulting from research to applied utility is elusive. However, there is substantial research on transfer of training producing significant predictors such as cognitive ability, conscientiousness, motivation, and supervisor support, but few prove consistently strong across job domains and organizations (Grossman & Salas, 2011; Saks, Salas, & Lewis, 2014).

Turnover. Turnover simply equates to employees leaving. Marion A. Bills (1925) is credited with the first turnover literature looking to predict why clerical staff were leaving. Turnover concepts and models evolved including the concept of movement desirability (March & Simon, 1958). More contemporary turnover research is framed from Mobley et al.'s (1979) content model that has informed broad conceptual notions about why people quit.

Mobley's (1977) sequential path model suggested that turnover starts with job dissatisfaction that leads to thinking about leaving. The employee, according to the model, then evaluates one's subjective expected utility (SEU) resulting from quitting. Mobley (1977) stated that SEU is followed by intentions to look for an alternative job, evaluating alternative jobs, and comparing the potential job to the present job. These steps can then lead to intentions to quit, with some employees actually leaving (Mobley et al., 1979; Mom et al., 2017).

Hom et al.'s (2017) exploration of the century of turnover research revealed that applied research with the potential to inform industry is lacking. This lack of applied research is due to the fact that it is challenging to observe turnover and control for the collective selection processes (Hom et al., 2017). Hom et al. (2017) also pointed out that most turnover research focuses on turnover prevention rather than actual turnover.

Voluntary versus involuntary turnover. Turnover can be categorized as voluntary or involuntary (Price, 1977). Voluntary turnover is employee self-

initiated separation (Mobley et al., 1979). Involuntary turnover is argued by McElroy et al. (2001) as avoidable turnover resulting from substandard performance or disciplinary issues resulting in dismissal. Dalton et al. (1982) assessed that attempts at reducing turnover focus mainly on voluntary turnover, which is considered preventable.

Functional, dysfunctional, and optimal turnover. Dysfunctional turnover occurs when an employee wants to leave, but the organization prefers they stay (Dalton et al., 1982). Dalton et al. (1982) contended that some turnover is not dysfunctional. They contended that functional turnover occurs when an employee, such as marginal performers or disgruntled workers, desires to leave and the organization does not object (Dalton et al., 1982). Functional turnover can be organizationally advantageous by reducing member homogeneity (Schaubroeck et al., 1998), thereby improving organizational performance in adaptation to change, groupthink, and innovation (Astakhova et al., 2015; Schneider et al., 1998).

Optimal turnover. Abelson and Baysinger (1984) contended that “dysfunctional” turnover can imply organizational malfunction and argued that may not be the case. Abelson and Baysinger (1984) operationalized the notion of *optimal turnover* as turnover resulting in organizational economic balances between the costs of turnover against the cost of reducing turnover.

This return on investment approach offers optimum outcome potential according to Abelson and Baysinger (1984). Employees voluntarily separating

when the organization prefers they stay is still considered dysfunctional in the Abelson and Baysinger (1984) optimum turnover model, but the model recognizes the organizational reality in inevitable turnover, such as retirement, and the resulting organizational value. As Ployhart and Weekely (2010) reminded us; everyone leaves the organization at some point.

Unfolding model of turnover (shocks). Turnover research during the 1990's witnessed a theoretical refinement leading to new and innovative models such the unfolding model of turnover by Lee and Mitchell (1994). Unlike earlier frameworks developed from March and Simon's (1958) model, the unfolding model of turnover argued that turnover rests primarily on job dissatisfaction (Lee & Mitchell, 1994). This dissatisfaction, operationalized a bit differently than job satisfaction in that it takes out those employees that are "neutral" in either direction, is the cause of employees looking for other opportunities perceived as better. Lee and Mitchell's (1994) unifying theory is considered the dominant current turnover model informing both research and practice (Holstrom, Mitchell, Lee, & Eberly, 2008; Hom, 2011; Hom et al., 2017).

The concept of *shocks* was later introduced as part of Lee and Mitchell's (1994) unfolding model. Shocks are events external to employee perceptions of job satisfaction (satisfied, neutral, or dissatisfied) that *push* or *pull* employees to contemplate leaving. These *jarring* events, pushing or pulling, help conceptualize alternative paths in turnover models (Burton, Holtom, Sablinski, Mitchell, & Lee,

2010). Shocks, according to Lee and Mitchell (1994), offer explanations to four distinct pathways toward turnover.

Shocks activating preexisting plan. The first path is a push or pull activated by an individual's preexisting desire or plan to leave (Mitchell & Lee, 1994). In a law enforcement context, examples could include a decision to stay home and raise children, leave for another career, or return to college for an advanced degree.

Negative job shocks violating organizational commitment. Lee and Mitchell's (1994) second path is negative job shocks that violate employees' values or goals, thereby leading to reduced organizational commitment. Examples for this shock, in the study context, include being falsely accused of violating department policies or procedures, being passed over for promotion, salary compression where much junior officers earn the same pay, or violations to perceptions of fairness or organizational justice.

Shocks from unsolicited job offers. The third shock takes the form of unsolicited and unexpected job offers leaving employees to weigh the potential advantages and disadvantages of a new opportunity regardless of satisfaction or commitment in the current organization (Lee & Mitchell, 1994). Law enforcement examples would include job and career offers resulting from an officers' existing network of friends, family, classmates, and colleagues, former organizational members in new police administrative roles elsewhere "poaching" an employee to

their new agency, or lucrative opportunities in the private sector related to the improved economy and reduced labor pool.

Conventional dissatisfaction. Lee and Mitchell's (1994) final pathway, number four, addresses the conventional turnover paradigm where dissatisfied employees leave to take a perceived better job. Examples relative to the study population include leaving for another agency due to compared stagnant wages and promotions, perceived poor organizational commitment to officers, or bad leader-member relationships with supervisors.

Unfolding model with job embeddedness. Mitchell and Lee (2001) added job embeddedness theory to their 1994 model. The addition to the unfolding model rests with the idea that embedded protective factors pull employees to remain part of the organization, therefore mitigating the pushes and pulls of shocks (Burton et al., 2010). The addition of Mitchell et al.'s (2001) job embeddedness adds scope to the unfolding model as it addresses the reasons people may stay versus why they leave. Literature is robust related to staying that include constructs including kinship (Price & Mueller, 1986), job sacrifices (Meyers & Allen, 1997), nonwork influences (Mobley et al., 1979), occupational embeddedness (Feldman & Ng, 2007), family embeddedness (Feldman et al., 2012; Mitchell et al., 2001), community embeddedness (Feldman, Ng, & Vogel, 2012), and training opportunities (Hom et al., 2017).

Law Enforcement Turnover Model. Law enforcement offers unique

research advantages to explore the potential nexus between SJT selection batteries and turnover related outcomes. This context specific research project capitalizes on the unique nature of daily field training performance evaluations (DORs), which adds unique richness of data for new hire on-the-job training that is not easily available in other industries.

Given the applied nature of this study, the law enforcement turnover model (Figure 3) visually depicts the law enforcement recruiting and selection process from field training to eventual turnover or retention. The model starts with the applicant pool entering the selection process, depicted by the “funnel”, including the SJT selection battery. Selections processes, by design, assess candidate’s job specific knowledge, skills, abilities, and other attributes leading to hiring new officers.

Newly hired officers spend time in orientation events, which last normally two to four weeks, before starting a fourteen-week field training and evaluation program (FTEP) (Fischer, 2015; Walker & Katz, 2002). Upon completing orientation, new officers begin the four phase FTEP program (Fischer, 2015; Walker, 2005). Phase one of FTEP required new officers to complete 10% of tasks associated with the job. Phase two expectations rise, where new officers handle 50 percent of the duties. Phase three should see the new officer able to handle 90 percent of the duties as they prepare for the final two-week phase where they are expected to handle 100 percent of the duties. Officers are assigned a different field

training officer in each phase, except in the final phase where the initial phase one officer completes the “check off” that the officer is ready for solo work (FDLE, 2011; Haberfield, 2002; MPCTC, 2017).

All four phases are depicted in the model (Figure 3) as arrows. The top series of arrows reflect failing field training (fail), which also reflects as a failure in the transfer of training (Baldwin & Ford, 1998). Officers failing field training most often relate to issues with P-J fit (Edwards, 1991) but some can be P-O fit (Judge & Ferris, 1992) related, as shown by some candidates who later became successful law enforcement officers in a different agency. Officers can change their mind about law enforcement (CMLE), and these are reflected in all green arrows indicating that they had no performance deficiencies. This group is considered a P-J fit issue as they all left, with none of these candidates currently working in policing.

The middle arrows reflects an officer who needs remedial training, in the model it depicts during phase two, which resulted in correction of the deficient performance and therefore the officer is retained. This is retention, but also may result in functional turnover if the field training performance issues reemerge later. The bottom series of arrows reflect a new officer who completes all four phases without the need for remedial training, thereby leading to retention.

The last long red arrow signifies shocks (Lee & Mitchell, 1994) and is unrelated to the transfer of training. In this study, shocks equate to candidates

leaving for another agency (n=4), changing professions (n=3), relocating due to spouse employment relocation (n=2), and moving closer to home (n=2).

Dysfunctional turnover for the study group is conservatively figured at 11 (of 101) officers.

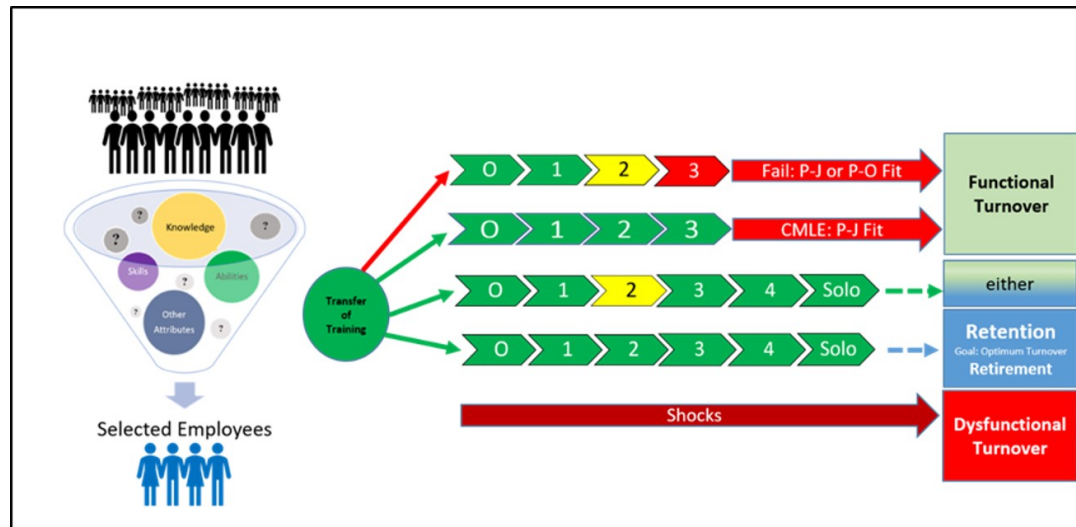


Figure 3. The Law Enforcement Turnover Model. This figure illustrates, based on study findings, how the Law Enforcement Turnover Model accounts for the impact of transfer of training on turnover or retention outcomes in police agencies.

Methods and findings summary. Binary multiple regression was deployed to explore the prediction of group membership in this study. The study hypotheses, as well as supplemental models, were explored for the predictive models given that the categorical dependent variables were all dichotomous (no=0; yes=1) with continuous, discrete, or dichotomous predictor variables (Field, 2013; Pallant,

2013; Tabachnick & Fidell, 2007).

Major key finding. The key finding of this study is that the reading test component of the SJT selection battery offers consistent prediction in overall training outcomes, training failure, overall turnover, involuntary turnover, the need for remedial training, and failure during remedial training. The reading portion of the test is a paper and pencil law enforcement based SJT. Candidates read scenarios and then select the multiple-choice response they believe best addresses the scenario.

Field training outcome findings. The prediction of overall field training performance and training failure consistently rests on reading scores as being statistically significant. Continuous reading scores, and the banded dichotomous factor identifying members with reading scores less than 90 percent, were cornerstones of all supported models. Officers who attempted the selection battery when they initially failed may have a predictive nexus with training performance, but it only approached (two-tailed) significance.

Turnover outcome findings. The predictive nexus between both continuous and banded reading scores exists with overall and involuntary turnover outcomes. Officers attempting the selection battery more than once, regardless of initial pass or fail, was statistically significant in successful models predicting involuntary turnover. Officers with previous experience approached two-tailed significance as a protective factor toward overall turnover and this finding warrants additional

inquiry.

Remedial training outcome findings. Understanding dynamics of why officers may require remedial training is important to practitioners charged with transfer of training related duties. Supplemental analyses explored various potentially predictive models. Supported models, in both needing remedial training and those subsequently failing training, shared the predictive constant of reading score. Additionally, the variable of officers who initially failed the selection battery was a statistically significant factor in supported models that predicted failing of remedial training.

There were factors that approached statistical significance in supported models related to remediation that warrant discussion and additional exploration. The dichotomous factor, which consisted of group members all scoring a perfect 100 percent in the writing score, indicated potential prediction on both the need for remedial training and subsequent failure of remediation. Writing scores of 100 (n=41) initially may seem counter intuitive but does warrant additional review. It is possible the relationship is spurious due to 41 percent of candidates being members of the group, or it may be tapping attention-to-detail, or other personal attributes, that may be adding contextual stress in this high-consequence profession.

Another potential protective factor is for members who received a score of “Good” in the sub-score of restraint. Members of the “good” restraint group were less likely to fail remedial training. Additional inquiry with other predictive

methods, such as path analysis, will be discussed below.

Contributions of the Study

Hom et al. (2017) reviewed turnover research from the past century and suggested that there is a severe lack of research related to selection and turnover. Rubenstein et al.'s (2015) meta-analysis exploring turnover resulted in their identification for a need in future research to explore actual turnover rather than turnover intentions, as well as to dichotomize turnover as voluntary or involuntary. These are among the contributions from this study.

Expected contributions. This exploratory study started with discussing the significance of the study and its explored domains, within existing literature, and where the study may add to that literature in chapter one. Each of these domains are revisited to discuss the study findings that potentially add to the literature.

Prediction versus prevention of turnover. Hom et al. (2017) called for more applied research on turnover based on their review findings that the majority of existing research focuses on prevention rather than prediction of turnover. This void of research between SJT and turnover along with the potential to inform practitioners is echoed by Barrick and Zimmerman (2009) and Whetzel and McDaniel (2009).

Study contributions. This study responded to Rubenstein et al.'s (2018) suggestion that turnover research explore actual turnover as opposed to intentions, by exploring actual turnover as a dependent variable. Additionally, the study dichotomizes

voluntary from involuntary turnover, also called for by Rubenstein et al. (2018), with predictive findings toward actual turnover.

This study also adds to the literature by answering a need expressed by Van Iddekinge et al. (2011) calling for research to identify prediction criteria in selection research. This study identified predictive factors of new officer field training outcome as overall turnover.

Contributes to limited turnover prediction literature. The second potential contribution discussed was the potential to capitalize on predictor constructs, such as personality, cognitive ability, and other KSAOs as called for by Campbell et al. (1993) and Schmidt et al. (1986). Additionally, this study contributes by showing potential predictors offered from SJTs as called for by Schmidt and Hunter (1998).

The results of this study do offer opportunities to capitalize on predictor constructs, from selection SJTs, with the potential to inform applied practice. In doing so, the study also succeeded by adding literature to limited research on turnover prediction as identified by Barrick et al. (2009) and Van Iddekinge et al. (2011).

Temporal order and KSAOs with turnover. Rubenstein et al. (2018) called for future research attempts to explore the temporal order associated with turnover. They discussed the majority of literature that looks at turnover at a single point in time. The use of DORs in this study allowed for temporal measurement by phase of training, and even day of training if warranted. Turnover and failure was controlled for temporally as during FTEP and post FTEP. Data was also collected related to exact

training phase and day or training deficiency but was not used in the hypothesis testing or supplemental analyses. The data exists and can be further explored in future research, which will be discussed later in this chapter.

What the study actualized very well, as recommended by Rubenstein et al. (2015), was measuring specific events in the predictive analysis of turnover. The study's exploration of failing training, the need for remedial training, retesting aspects, and identification of those changing their minds about law enforcement are examples that add to this lacking temporal order literature.

Additionally, the National Research Council (2004) called for exploration of knowledge, skills, abilities, and other characteristics (KSAOs) in measuring actual police field performance. This study suggests various predictors such as reading ability, prior experience, and restraint constructs predict performance.

Public sector research. Selden et al. (2013) examined new employee turnover in state governments using high-performance work systems (HPWS) framework. Selden et al.'s (2013) findings point to three HR practices (recruiting and selection, compensation, and training and development) that predict new hire turnover. Ployhart (2006) questioned the generalizability of public sector selection research to private organizations. Ployhart (2006) recognized that results are generalized without proper exploration.

This study explores public sector turnover, thus adding to the literature on new hire turnover such as Selden et al. (2013). Additionally, the use of SJTs with

scores for SJT video, writing, and reading offer promising future studies to explore generalizability to private sector jobs, as suggested by Ployhart (2006).

Police training effectiveness. The National Research Council (2004) contended that police training effectiveness lacks the rigorous research needed to help shape training development. There are practitioner concerns that without research designs, beyond simply correlational designs, progress is unlikely (Geller & Toch, 1995; National Research Council, 2004). This study deployed a predictive modeling methodology; specifically examining field training performance with immediate applied potential for new training interventions based on reading score indicators. Additionally, Caro (2011) pointed to a host of applied advantages in exploring police field training performance. This study actualizes what Caro (2011) calls for in future research.

SJTs predicting performance and turnover. The sixth anticipated contribution relates to the evolving nature of SJT research. Ryan and Ployhart (2014) suggested that future SJT research should include prediction of performance in unexplored domains such as turnover. Ployhart and Weekley (2006) contended that such research has empirical applied utilities. Ployhart and Weekly (2006) further urged that research extend beyond correlational designs and called for more robust experimental analysis.

This study specifically adds to these calls related to future research and improved methodologies. First, the study directly explored prediction of SJT

selection scores to various turnover outcomes. Second, the binary logistical regression used offers more robust and predictive analyses between SJTs and the dependent variables, thus addressing the literature gaps identified by Ployhart and Weekley (2006), Ryan and Ployhart (2014) and The National Research Council (2004).

Selection battery use toward improved outcomes. Ployhart (2006) discussed the irony that nearly every organization deploys some type of selection process with no guarantee regarding appropriateness, effectiveness, or optimal utility. Ployhart (2006) called for future research to explore this gap between selection and outcomes to maximize selection processes. Hausknecht and Holwerda (2013) suggested that research methodologies using member tracking strategies may produce robust turnover results and therefore be more promising for applied utility.

A major applied goal of this study was to explore whether or not existing SJT selection battery results inform training and turnover outcomes. Actionable results would leverage existing data and offer potentially cost neutral exploration toward training development, selection strategies, and retention programs. This optimization addresses the call from Ployhart (2006). This study also responds to the type of exploratory and predictive analysis suggested by Hausknecht and Holwerda (2013) regarding potential applied utility.

Police transfer of training and retention strategies. The eighth identified

contribution relates to transfer of training. Baldwin et al. (2009) discussed the applied importance of transfer of training research. Baldwin et al. (2009) contended that evidence-based interventions and strategies can be critical to improve transfer of training in the field. Hom et al. (2017), echoing Baldwin et al.'s (2009) called for future research to explore factors that may inform improved training designs, training readiness, and supervisor support. Hom et al. (2017) also urged for improved empirical observations of various management practices and the impact on turnover outcomes.

This study contributes to transfer of training literature. Study findings, such as reading scores predicting training and turnover, may lead to training interventions for lower scoring officers. This aligns precisely with the call from Baldwin et al. (2009). Improving training readiness through a training intervention may improve transfer of training and ultimately improve training success and turnover outcomes, as suggested by Hom et al. (2017). This study also explored management practices, such as retaking the selection battery, with a potential predictive nexus to turnover. This improved empirical observation may be modest based on the population size, but it meets the research mark set by Hom et al. (2017).

Specific context and operational definitions. Allen et al.'s (2014) review of empirical turnover studies (n=447) led them to conclude that any growth in turnover literature depends on having clear and precise articulation of operational

definitions and study context. Hausknecht and Trevor (2011) and Hom et al. (2017) add to Allen et al.'s (2014) identification that job context is critical in exploring the scope and frame of various turnover antecedents (Allen et al., 2014; Hausknecht & Trevor, 2011; Hom et al., 2017; Nyberg & Ployhart, 2013).

Allen et al. (2014) made an interesting observation about the current state of turnover research. They contended that this lack of context has created a dominant analytical mindset (DAM) wherein assumptions in the body of turnover literature hamper evolution. Allen et al. (2014) contended that the response to the DAM is to broaden turnover research to add specificity in study populations, explore actual turnover beyond the tendency to use intentions to quit, and to dichotomize turnover to at least voluntary and involuntary turnover.

The study methodology includes actual turnover, including dichotomized voluntary and involuntary turnover, and sector and industry specificities. The archival data adds the specificities called for by researchers such as Allen et al. (2014), Hausknecht and Trevor (2011), Hom et al. (2017), and Nyberg and Ployhart (2013).

Findings related to the research questions. This exploratory study examined various predictive models that were exploring for predictive nexuses between law enforcement SJT selection battery scores and turnover. The dependent variables included field training performance (RQ1) and turnover outcomes (RQ2). The study findings include statistically significant and supported predictive models

related to both training performance and turnover. Predictive factors include reading scores and whether a subject retook the selection battery, both overall and with initial fail. Two factors approached (two-tailed) significance as potential predictors that warrant discussion. Officers with perfect writing scores (100 percent), “good” ratings in the sub-score of restraint, and prior law enforcement experience show the potential of prediction. These will be discussed further in recommendations regarding future research.

Findings and theoretical support for the study framework. The Law Enforcement Turnover Model (see Figure 3) conceptualizes law enforcement turnover. This model offers unique pathways, specific to the industry, which capitalize on the daily observation and documentation of standardized performance expectations offered by police FTEPs. Study findings offer support for this model. These will be explored below along with study findings supported by theory. Empirical support for the study findings, the model, and the nexus linking both are discussed below.

Selection SJT predicting performance and turnover. Whetzel and McDaniel’s (2009) meta-analytic review of SJTs in selection included improved predictive validity for job performance (Christian, Edwards, & Bradley, 2010; Whetzel & McDaniel, 2009). Additionally, SJTs produce less adverse impact than other options (Barrett et al., 2010; Callanin & Robertson, 2000; Chan & Schmidt, 1997; Motowidlo & Tippins, 1993; Oswald, Schmitt, Kim, Ramsay, & Gillespie,

2004; Weekley & Jones, 1997; Whetzel, McDaniel, & Nguyen, 2008).

Police SJT study. Annell et al. (2015) conducted a longitudinal study of police recruits (n=750) in Sweden, exploring possible prediction of selection batteries, consisting of cognitive ability, personality, and fitness as predictors to performance, satisfaction, retention, and health. Annell et al. (2015) found that verbally loaded cognitive ability tests predicted academic and job performance. They also found that general cognitive scores negatively predicted only job satisfaction.

SJT format: video-based vs. written scenarios. Early SJTs were delivered in a written scenario using paper and pencil format before the technology and costs associated with multimedia video formats allowed for a more robust version (Brown, Jones, Serfass, & Sherman, 2016; Dalessio, 1994; Funke & Schuler, 1998; Lievens & Sackett, 2006; Weekley & Jones, 1997). Lievens and Sackett (2006) explored the practical impact of the evolution from paper and pencil SJT tests to video-based SJTs with offer more robust scenarios. They studied this transition at a Belgian medical school (n=1,100+) where prospective students took SJT admission tests, initially paper and pencil then video-based (Lievens & Sackett, 2006). Results suggested that video-based SJTs offer better prediction related to intrapersonal related criteria. However, they did find that the written scenario-based paper and pencil better predicted cognitive ability.

Study findings supported by theory: SJTs. Study findings suggested that

reading scores predict training and turnover outcomes. The reading test is a verbally loaded test that is arguably an indirect measure that taps into cognitive ability. This finding is consistent with Annell et al.'s (2015) Swedish police study, suggesting that verbally loaded cognitive ability tests predicted academic and job performance.

SJT improved validity. The study also finds support in Whetzel and McDaniel's (2009) meta-analytic review, which identified SJTs as improving predictive validity for job performance (Christian, Edwards, & Bradley, 2010; Whetzel & McDaniel, 2009). The SJT battery did offer predictive utility in reading, writing, and restraint scores as supported by Whetzel and McDaniel's (2009) study.

Adverse impact avoidance. Perhaps most important to law enforcement management is the reduced adverse impact of SJTs related to other selection options (Barrett et al., 2010; Callanin & Robertson, 2000; Chan & Schmidt, 1997; Motowidlo & Tippins, 1993; Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Weekley & Jones, 1997; Whetzel, McDaniel, & Nguyen, 2008). The study agency deploys this SJT selection battery specifically to avoid adverse impact. While this study did explore adverse impact directly, correlations were reviewed related to race and gender and the various independent and dependent variables. The only significant results related to ethnicity and gender offers some support for the avoidance of adverse impact supported by the studies listed above.

This study found that African-American ethnicity was statistically

significant with reading scores ($p = .000$). Given the nexus between reading scores and training outcomes, African-American ethnicity was also statistically significant ($p = .049$) in correlations to failing training, ($p = .036$) not completing FTEP, and ($p = .058$) needing remedial training. There were no statistically significant correlations with African-American ethnicity and any other dependent outcome variable.

The other factor approaching significance was gender in relation to SJT video scores ($p = .051$). Correlations with gender and video score indicate that females score lower than males. It is also noteworthy that gender's interplay with leaving the agency, which resulted in shocks, approached significance ($p = .088$). This indicates that female officers were more likely to leave as a result of shocks than male officers. Other than with the SJT video score, gender showed no additional significance with any dependent variable related to training outcomes, the need for remedial training, or any turnover outcomes.

While the gender differences are unlikely to result in adverse impact in a verbally loaded cognitive ability test, ethnicity related to reading scores most likely would. This likelihood adds theoretical support of literature suggesting that SJTs avoid adverse impact.

Paper-pencil versus multimedia SJTs. This study gets support from literature which suggested that paper and pencil SJT tests are better predictors of cognitive ability. The reading portion of the SJT selection battery is a paper and

pencil SJT and is arguably tapping into cognitive ability consistent with Lievens and Sackett's (2006) findings.

Transfer of training. Transfer of training equates to the generalization, retention, or enhancement of job relevant KSAOs acquired through the function of training (Baldwin & Ford, 1998). Saks and Burke-Smalley (2014) found that on-the-job training is preferable to classroom or computer, with a better relationship to performance outcomes.

Quinones (1995) explored trainee perceptions, finding that labels of *remedial* or *advanced* training may be important. Quinones (1995) added that the training frame can impact participant perceptions related to fairness and willingness to learn.

Gully et al. (2002) explored training effectiveness resulting from mistakes or errors in decision-making scenarios. Gully et al.'s (2002) findings suggest that those with higher cognitive ability and openness to experience best learned from mistakes. Trainees higher in conscientiousness were found to struggle as encouragements by training supervisors to look at mistakes as training opportunities were found to impact self-efficacy (Gully et al., 2002). Several additional studies link cognitive ability to transfer of training outcomes (e.g. Grossman & Salas, 2011; Maltarich et al., 2010; Saks et al. 2014; Van Iddekinge et al., 2011).

Study findings supported by theory: Transfer of training. This study

explored transfer of training outcomes specifically by identifying members failing field training. Failing field training (n=16) was examined as well as officers requiring remedial training (n=23) resulting in both passing (n=7) and failing (n=16). Predictors related to transfer of training failure were reading scores and retest of selection battery.

Baldwin and Ford's (1998) transfer of training supports the study findings. Specifically, those failing training were a direct result of a transfer of training failure. This on-the-job training, as discussed by Saks and Burke-Smalley (2014), displayed direct nexus to performance outcomes concurrent with their study findings.

There is additional support regarding trainee perceptions of training frames as posited by Quinones (1995). Three officers quit upon learning they were required to take remedial training. While it cannot be determined if those decisions were due to the frame of "remedial training", inferences can be made. At least one of those quitting had prior law enforcement experience leaving the possibility that the frame played a part in the decision.

Additional theoretical support comes from literature aligning cognition and transfer of training. Reading scores, in the study context, appear to be a proxy for cognitive ability, therefore garnering support from the studies associating cognition with transfer of training outcomes (e.g. Grossman & Salas, 2011; Gully et al., 2002; Maltarich et al., 2010; Saks et al. 2014; Van Iddekinge et al., 2011).

P-J fit. P-J fit is the fit between employee KSAOs and the job (Edwards, 1991). P-J fit factors influence job task analyses that, in turn, impact organizational processes related to selection, training, promotion, and retention (Breaugh & Starke, 2000; Kristof, 1996; Kristof-Brown, 2002; Werbel & Gulliland, 1999). P-J fit is considered critical in selecting members for high stakes professions (Lough & Von Treuer, 2013).

Study findings supported by theory: P-J fit. Overall turnover (n=46) in the study (n=46) includes officers who left for various reasons including involuntary turnover (e.g. training failure, disciplinary termination) and voluntary turnover. While FTEP failure (n=16) accounted for most officers who did not complete FTEP, there were nine officers who changed their mind about law enforcement and resigned prior to completing training. None of these nine officers recorded performance deficiencies during training. Additionally, two officers changed their mind about law enforcement post FTEP, and another three more left due to changing professions. Findings related to those who changed their mind about law enforcement find theoretical support in P-J fit (Edwards, 1991) literature.

P-O fit and ASA model. P-O fit explores how new employees are accepted by the organization comprised of its incumbent members (Judge & Ferris, 1992). P-O fit has strong theoretical support from Schneider's (1987) ASA model (Piasentin & Chapman, 2006). Kristof (1996) and Verquer et al. (2003) reported study findings suggesting P-O fit rests on new and incumbent members sharing

characteristics for retention.

Study findings supported by theory: P-O fit and ASA model. Overall turnover in the study (n=46) includes officers who left for various reasons including involuntary turnover (e.g. training failure, disciplinary termination) and voluntary turnover. Of those officers who left, eleven are reportedly still in law enforcement and two in corrections. These findings are supported as P-O fit (Ferris, 1992) wherein officers may have not been a good fit for the particular agency as well as the ASA model (Schneider, 1987).

Functional turnover. Dalton et al.'s (1982) notion of functional turnover includes those members not performing at standards and discipline. This type of turnover is considered functional since retaining the officers would be of detriment to the agency.

Study findings supported by theory: Functional turnover. The most frequent group in the study population were those who failed field training (n=16). Additional turnover classified as functional included those with P-J fit issues, who specifically changed their mind about law enforcement (n=11). Discipline terminations (n=4) are also considered functional. Subjects from this study (n=101) included 31 members who were conservatively classified in the functional turnover group. These study findings are supported in functional turnover theory.

Dysfunctional turnover and shocks. Dysfunctional turnover occurs when members leave but the agency would prefer that they stay (Dalton et al., 1982).

Shocks are the external pushes and pulls that can lead to turnover (Lee & Mitchell, 1994). These two factors share nuances and therefore are discussed together given that the reason for most dysfunctional turnover is related to shocks.

Study findings supported by theory: Dysfunctional turnover and shock. The dysfunctional and shock group share membership, which is comprised of members who left the agency while in good standing. This dysfunctional, and shock group, include officers leaving for another agency (n=4), changing professions (n=3), relocating due to spouse employment relocation (n=2), and moving closer to home (n=2). Dysfunctional turnover for the study group is conservatively figured at 11 (of 101) officers. These findings are theoretically supported by theories related to both dysfunctional turnover (Dalton et al., 1982) and shocks (Lee & Mitchell, 1994).

Functional turnover or retention “either” area. The Law Enforcement Turnover Model (Figure 3) identifies pathways to functional turnover, retention, and dysfunctional turnover. There is an area reflected in the model as “either”, indicating that officers with remedial training may eventually turnover or may be retained. For example, marginally performing officers who pass field training may not equally contribute to the organization mission. Additionally, members who left for another profession (n=3) may or may not have been functional members within the agency. It is assumed that those leaving for another agency were dysfunctional, but it is important to point out that officers (n=3) not successfully completing the

FTEP program due to issues including performance failure and discipline are still in law enforcement.

Study findings supported by theory: Functional turnover or retention “either” area. There is no theoretical support for an “either” functional or retention grey area. The logical basis for including this area visually in the model is recognizing that some turnover or retention outcomes may, or may not be, good for the organization or the officer. This notion is important in the model to reflect the caution practitioners should use in assessing functional turnover and retention concepts.

Optimal turnover. Abelson and Baysinger (1984) discussed that optimal turnover occurs in instances such as normal retirement, when marginally performing officers leave, or other situations where the agency realized a cost-benefit in the turnover. The model includes financial advantages but also hedges against homophily and increases new thought and potential innovation.

Study findings supported by theory: Optimal turnover. This population was limited to the time frame of the SJT selection battery since October of 2012. However, the cohort did contain one officer, with previous law enforcement experience, who retired. This is an instance of an optimal turnover classification. As time passes, officers from this cohort who retire will be considered as optimal turnover also.

Findings that differ from previously reviewed studies. More recent research focusing on “what” SJTs actually predict is drawing increased attention and debate. Lievens and Motowidlo (2016) suggested that SJTs actually tap into specific job relevant content as well as *general domain knowledge*. Their proposed SJT model stated that SJTs measure general domain knowledge that is not contingent on context (Lievens & Motowidlo, 2016). Lievens and Motowidlo (2016) defined general domain knowledge as, “knowledge about the utility or importance of traits such as these for effectiveness in a job that actually requires expressions of these traits for effective performance” (p. 4).

Lievens and Motowidlo’s (2016) general domain knowledge argument garnered some support in meta-analyses. Clevenger et al.’s (2001) meta-analysis explored federal investigative officers, customer service personnel, and manufacturing engineers. Their results were mixed related to SJTs tapping into cognitive ability as some had predictive validity based on industry or job domain. Clevenger et al.’s (2001) study resulted in low correlations and a lack of statistical significance between SJT scores and previous job experience.

Difference with study findings. Unlike Clevenger et al.’s (2001) meta-analysis that aligned with Lievens and Motowidlo’s (2016) general domain knowledge, the current study did find support for previous job experience and performance. Prior law enforcement experience in the current study was explored with results suggesting that it is a protective factor against turnover.

Discussions and Implications

Hom et al. (2017) explored decades of turnover research and concluded that research with the potential to inform practice is lacking, with most turnover research focused on prevention versus prediction of turnover. Implications of the study are discussed related to contributing to the literature. Aspects critical to the study context and scope will be discussed as limitations and boundary conditions.

Key implications informing literature. The implication of this study in informing literature rests in answering the question regarding SJT selection tests and their potential predictive nexus to training outcomes and turnover. This study's findings suggest that yes, there is additional predictive utility possible from SJT selection battery test results. The law enforcement context may be important to potential generalizability along with psychometric properties associated with this particular SJT selection battery. Despite potential context nuances that may limit replication, the groundwork is laid to possibilities previously unexplored in the literature.

Limitations. This study has several limitations that may impact replication or generalizability to other sectors, industries, or organizations. While each may, or may not, limit future research efforts in similar predictive studies, they must be identified and discussed. Limitations include methodology nuances, sample constraints, the SJT selection battery source, the unfolding model of turnover issues, and SJT construct questions. Each limitation will be discussed in the

following paragraphs.

Methodology. Binary logistic regression was used in this study as it offers predictive testing of theoretical models, opposed to statistically driven models, thus allowing for better potential applied utility (Field, 2013; Pallant, 2013; Tabachnick & Fidell, 2007). Additionally, this study assumed that the dependent variables had a linear relationship with the dependent variables (Tabachnick & Fidell, 2007). The strengths of logistic regression, according to Tabachnick and Fidell (2007), include exploration of multicollinearity between dependent variables (predictors) where potential predictors are covariates in prediction related to the dependent variable.

It is important to point out that logistic regression is not without limitations. According to Tabachnick and Fidell (2007) limitations of logistic regression, while limited, rest in the requirement for dichotomous outcome variables. This was a strength for this study as the outcome variables coded as “0” for no and “1” for yes aligned well with overall turnover, involuntary turnover, voluntary turnover, and training performance measures.

Sample cohort and size. The study population consisted of officers hired using a SJT selection battery commencing in October of 2012. This study, therefore, did not explore officers hired using earlier selection processes. This limits observation and study for officers hired prior to October 2012 that also experience shocks, involuntary, voluntary, or optimal turnover.

Limited functionally optimal turnover. This seven-and-a-half-year snapshot of the agency's turnover drastically limits observation of optimal turnover, mostly due to officers successfully completing a career and retiring. Only one officer retired that was selected using this selection battery. While this example affirms, albeit in a singular example, the optimal turnover equated with functional turnover, it did not have enough group membership to explore predictive models toward successful career completion.

Sample size. Based on both the sample size related to the number of predictor variables in the models reported, (Hsieh, 1989) and related to the rule of ten events per predictor (Vittinghoff & McCulloch, 2006) the sample size (N=101) for the logistic regression is appropriate. It is also an applied study so easily adding subjects, as done in controlled experiments, is not realistic aside from the passing of time potentially revealing additional subjects.

Sample size: Range restriction. The limitation of this study relates to sample size in the range restriction that occurred in some of the nine SJT sub-scores (e.g. critical thinking, investigative, confrontation, and empathy). While these range restrictions statistically excluded them from any statistically supported models, it did limit the potential of these sub-scores may offer. All sub-scores were recoded into dichotomous (no, yes) variables for each of level of the sub-scores (risk, some problems, normal, and good), therefore allowing them to be included in attempted models.

Sample size: Missing data. The study data did have a few missing data points. The data missing is related to officers who had been terminated, and in which cases the files had been transferred to long-term storage. The missing data impacted four subjects' data related to them having any prior military experience, prior law enforcement experience, or having earned any college degrees.

Source of SJT selection battery. Data for this study was derived from the study agency's SJT selection battery scores and sub-scores that were provided by an outside vendor. Psychometric data was not available aside from basic scoring discussion and interpretation. This limits direct replication unless additional agencies are added using the same SJT selection battery. Future meta-analyses would likely require SJT tests with more available psychometric information in order to allow proper cross-coding related to various tests from various studies.

Unfolding model. Hom et al. (2017) considered the unfolding model of turnover a major advancement in turnover research. They also pointed out limitations of the theory. Most notably, a limitation of their research is that the unfolding model of turnover rests thus far on qualitative research (Hom et al., 2017). Hom et al. (2017) contended that quantitative predictive research is necessary for proper theory development.

Limitation specific to this study. This study did discuss instances that were considered shocks. Unfortunately, the hypotheses and supplementary analyses did not directly explore the predictive nexus to shocks. This potential will be further

addressed in future study recommendations.

What subconstructs do SJTs tap? SJTs enjoy considerable evidence related to criterion-related validity given that they relate to job performance (Campion, Ployhart, & MacKensie Jr., 2014; Christian, Edwards, & Bradley, 2010; Lievens & Motowidlo, 2016; McDaniel et al., 2001). The exact constructs of “what” SJTs measure remain elusive (Lievens & Motowidlo, 2016; Motowidlo, Dunnette, & Carter, 1990).

Meta-analytic support exists for a nexus between SJTs and leadership, interpersonal skills, and team skills (Christian et al., 2010). A nexus to SJTs and integrity was identified with a police population (Meijer et al., 2010) as well as among medical school applicants (Husbands et al., 2015). Lievens and Motowidlo (2016) argued that SJTs tap a variety of relevant job domains, especially interpersonal behaviors, and that they are better at predicting academic success than cognition.

Limitation specific to this study. The results of this study are consistent with the body of research supporting the criterion-related validity of SJTs (Campion, Ployhart, & MacKensie Jr., 2014; Christian, Edwards, & Bradley, 2010; Lievens & Motowidlo, 2016; McDaniel et al., 2001). The limitations are also consistent with the current academic discourse into “what” constructs SJTs measure (Lievens & Motowidlo, 2016; Motowidlo, Dunnette, & Carter, 1990).

An argument can be made that the most predictive aspect of the SJT

selection battery explored in this study is the reading score. A plausible argument also would contend that reading scores, being verbally loaded SJTs, are a proxy for cognitive ability. The goal of this study was not to resolve this widely acknowledged quest and controversy of the literature. It is, however, critical to discuss as a potential limitation of the study.

Boundary Conditions. Allen et al.'s (2014) review of empirical turnover studies (n=447) led them to conclude that any growth in turnover literature depends on having a clear and precise articulation of operational definitions and the study context. Hausknecht and Trevor (2011) and Hom et al. (2017) added to Allen et al.'s (2014) identification that job context is critical in exploring the scope and frame of various turnover antecedents (Allen et al., 2014; Hausknecht & Trevor, 2011; Hom et al., 2017; Nyberg & Ployhart, 2013). The boundary conditions in this study include sector, industry, job, selection battery deployed, selection battery scores, field training, and turnover.

Sector, industry, and job domain. This context of this study is a public-sector, law enforcement industry, and the job of a police officer. The study agency is a Florida police department with less than 250 officers, which is considered mid-sized. Generalizability should be limited to U.S. police agencies that utilize SJT selection batteries, preferably from the same vendor, who also deploy the San Jose FTEP model and use certified field training officers.

SJT and selection battery. The SJT selection battery results come from an

industry vendor. The agency and vendor are intentionally unnamed to protect the identity and confidentiality of all stakeholders as per normal and customary IRB expectations related to protection of human subjects.

Field training and evaluation program. The study agency deploys the San Jose FTEP and ensures that all field training officers attend the training per FDLE requirements. Generalizability, outside of agencies using the San Jose model and concepts consistent with the FDLE field training officer model, may be limited.

Turnover Constructs. Turnover is explored as overall turnover, involuntary turnover, and voluntary turnover as suggested by Hom et al.'s (2017) meta-analysis recommendations.

Law Enforcement Turnover Model. The law enforcement turnover model developed is based on assumptions and theory associated with the ASA model of turnover (Schneider, 1987), P-J fit (e.g. Kristof-Brown et al., 2005), P-O fit (e.g. Maltarich et al., 2010), transfer of training (e.g. Baldwin et al., 2017), and various shocks postulated in the unfolding model of turnover (e.g. Burton et al., 2010).

Recommendations

The results of this exploratory study produced actionable findings. These findings may prove insightful and actionable to the study agency, as all elements of the study are derived from that specific context. Potential applied recommendations will be offered, along with cautions related to potential unintended consequences which are specifically noted. Strategies aimed to improve performance and

turnover outcomes based on post-selection data offer advantages to new officers and the agency. Altering the selection process, based on this study, warrants discussion. However, any action needs to involve analysis related to employment issues such as adverse impact.

Recommendations for future research are less cautious. The study succeeded in establishing that the SJT selection battery scores do offer prediction for field training and turnover outcomes in law enforcement settings. The additional utility of these tests, normally shelved by the HR unit post selection, offer researchers rich opportunities. This data can help inform identified research gaps in employment selection outcomes, SJT criterion and construct validity, transfer of training outcomes, quantitative exploration of the pathways related to the unfolding model of turnover, and more precise turnover dichotomies.

Recommendations for applied utility. The study results offer actionable findings for the study agency, as well as generalizable treatment opportunities for other law enforcement agencies. The two most promising actionable findings will be explored along with one aspect related to selection test policy. The first recommendation addresses the potential officer deficiencies that are related to reading scores, which predict performance, such as report writing domains in field training.

The other immediate actionable strategy recommendation from these findings relates to the number of officers in field training that changed their minds

about policing and left. This points to a need to revisit all recruiting and selection aspects related to realistic job previews. The third recommendation involves reviewing selection battery retest policies. Any potential intervention or strategy that leads to improved training outcomes and retention saves law enforcement valuable time and financial resources.

Recommendation for pre-FTEP reading related training. Reading scores were significant predictors throughout this study. Reading scores even remained statistically significant in the few models that were not statistically supported overall. Reading scores alone are shown to predict training success or failure, overall turnover, and involuntary turnover. No officer who scored less than 90 percent (n=7) is still with the agency.

Applied recommendation. The agency should explore pre-FTEP training for those reporting reading scores that are less than 90 percent. Issues with a clear nexus to reading that were identified among officers who required remedial training included the officers' ability to understand and act upon department policies and procedures, as well various report writing issues.

Training identification and development. A search for existing trainings related to improved reading and retention, along with report writing would be warranted. In the absence of existing pedagogically sound training, the agency may want to work with a local university to more precisely assess the training need, perhaps via focus groups with field training officers, in order to develop the

training need, goals, and learning objectives.

Framing and potential broader utility. The ultimate objective of this pre-FTEP training would be to help new officers start FTEP with an improved strategy for reading and report writing, thereby leading to training success. As Quinones (1995) pointed out, the frame of the training matters. Rather than labelling the training “remedial” or “developmental”, perhaps a proficiency approach would pay dividends. This scenario would have modules in the orientation phase prior to the start of FTEP. Proficient officers would quickly progress and be ready to start FTEP. Officers requiring more time to complete these modules would be afforded more time, within reason. This may also help identify other officers who may benefit from this type of training, therefore leading to expected improved success in field training performance.

Recommendation for improved realistic job preview strategies. The study revealed that nine officers changed their minds about law enforcement during the past 7.5 years. None of the nine officers had performance deficiencies reported in their DORs. Officers changing their mind about law enforcement resulted in not making it out of FTEP, which accounted for 56 percent of losses compared to those who failed (n=16).

The comparisons strongly suggest that there is a disconnect between new officer expectations about policing and actual on-the-job realities. These disconnects are most commonly alleviated with *realistic job previews* (RJPs) that

have been shown to reduce turnover (Griffith et al., 2002; Dalessio, 1994; Jones et al., 2002; Scott & Jones, 2003).

Potential disconnect between recruiting and RJP. The recruiting function's goal is to attract hopefully qualified candidates to job openings. In policing, these recruiting efforts take place through posters, websites, social media, and videos commonly showing the "cool" aspects and various functions of policing. These positive mechanisms are designed to attract interest, but can also set certain expectations.

RJPs, on the other hand, are designed to give potential employees a realistic and unfiltered snapshot of what the job entails (Griffith et al., 2002; Dalessio, 1994; Jones et al., 2002). These RJP snapshots are intended to adjust expectations to realistic levels for new officers, thus reducing inconsistencies between actual and received outcomes, which is consistent with Vroom's (1964) tenants of expectancy theory.

Actionable strategy. Agencies should allocate resources to updating RJPs when recruiting efforts increase. RJPs should have parallel types of materials, especially if video-based, to maintain a consistent message. Recruiting videos can be developed with RJP in mind, but recruiting efforts designed to gain attention without intentional RJPs do not offer effective expectation balance. The bottom line is that agencies need both recruiting materials designed to attract candidates and RJP materials designed to create clear and realistic expectations of the actual

job of an officer.

Additional RJP strategy: Mandated ride-alongs. Most law enforcement agencies offer ride-along programs. Some are liberal and let any citizen without a criminal record ride with an officer for a period of time. Some agencies are more restrictive and limit ride-alongs to college students, applicants, or candidates. A potential enhancement to improve expectations would be mandatory ride-alongs. A recommendation of at least three ride-alongs, with different officers, during parts of different shifts would allow a reasonable opportunity to provide a variety of realistic observations. This variability of both time and officer allows for advantageous snapshots of the actual job that can have positive outcomes related to P-J and P-O fit.

Recommendation to review selection battery policies on retest. Study findings included models where officers retaking the selection battery, regardless of initial pass or fail outcomes and retaking with an initial score of fail were predictive. While this data warrants additional review with the agency, management review of the policy to allow retesting is warranted. Caution should be observed before making any change, such as the potential adverse impact implications, but it may reduce involuntary turnover based on training performance, especially if additional research can add more clarity to the retest outcomes.

Recommendations for Future Research.

Several meta-analyses acknowledge the gap in literature between SJTs and turnover (e.g. Barrick & Zimmerman, 2009; Whetzel & McDaniel, 2009) as well as actual turnover rather than simple intentions (Rubenstein et al., 2018). The only prior study exploring SJT's potential prediction to turnover was done by Dalessio (1994), where he studied post selection SJT training scores and insurance agent turnover. Dalessio's (1994) findings were consistent finding regarding realistic job previews and job enrichment on turnover. Dalessio (1994) recommended the use of SJTs in selection decisions versus the post-selection training used in his study.

This study actualized Dalessio's (1994) recommendation to study SJTs used in selection batteries and turnover, thus filling an identified gap in the literature (e.g. Anell et al., 2015; Dalassio, 1994; Lievens & Motowidlo, 2016; Whetzel & McDaniel, 2009). Results offer a starting point for future research into answering the unexplored question of whether SJT selection batteries offer additional utility, beyond selection, to turnover outcomes.

Replication studies. Given the existing gap, replication studies would add to literature by exploring generalizability as well as establishing literature for future meta-analyses. Replication, both in the context of law enforcement, other high-stakes professions, and other industries should pay attention to the context, scope, and frame (Allen et al., 2014; Hausknecht & Trevor, 2011; Hom et al., 2017; Nyberg & Ployhart, 2013).

Law enforcement replications. Future replications in the context of law enforcement are important toward establishing generalizable findings. Any opportunity to add additional agencies utilizing the same SJT selection battery should be seized. Additionally, increased attention on identified FTEP specific performance domains would add to the literature. A limitation of the current study is the longevity of the SJT selection battery use. Additional agencies may offer a longer temporal duration and capture data that allows for a more detailed and robust analysis of optimal turnover, such as retirement and shocks. The findings related to perfect writing scores perhaps serving as a predictor for remedial training would benefit from replication observations.

Non-law enforcement populations. Other sectors and industries would find applied utility in the findings related to SJT selection scores and turnover outcomes. It may be inferred that high-consequence professions may eventually be more generalizable to the law enforcement study, but this needs to be tested. Any opportunity allowing a researcher to SJT selection scores and turnover outcomes should be explored. A limitation for non-police agencies regards the common lack of initial training observations and ratings. This is an area where high-stakes professions may offer a better opportunity to capture early performance data.

Transfer of training researchers (e.g. Bloom et al., 2010; Grossman & Salas, 2011; Saks, Salas, & Lewis, 2014) discussed the elusive “magic bullet” predictor for training success in existing literature. While predictors exist, such as

cognitive ability, conscientiousness, motivation, and supervisor support, none prove to be consistent across domains (Grossman & Salas, 2011; Saks, Salas, & Lewis, 2014). Replication studies with mindful designs that are conscientious of context and operationalization may not find a universal “magic bullet”, but they may offer common predictors for specific industries or other emerging classification of jobs, such as those with high-stakes consequences.

Expanded predictive methodologies. Logistic regression is an appropriate method for this study. Eventually, structure equation modeling (Tabachnick & Fidell, 2007), such as path and factor analysis, may offer more precision in identifying more generalize clusters leading to meta-analysis. The larger the sample size and more robust the variables are among group membership, the greater the opportunity for expanded multivariate analysis (Tabachnick & Fidell, 2007).

SJT selection batteries predicting and superior performance. The current study explored performances that lead to either training failure or success. Those who did not fail are assumed to have performed at a minimally acceptable level. Organizations utilizing SJT selection batteries and valid performance management systems may be able to capitalize in exploring prediction of good performance.

Conclusion

Surprisingly, selection battery prediction of turnover in literature is scarce aside from three studies identified by Hom et al.’s (2017) meta-analysis. Barrick and Zimmerman (2009) studied selection batteries and found a predictive nexus

between personality traits and turnover. Maltarich et al. (2010) found that cognitive ability, mediated through P-O fit and cognitively demanding jobs, predicts turnover. The third study is a meta-analysis conducted by Van Iddekinge et al. (2011) who reported findings suggesting alignment of both employee interest and organizational characteristics (construct-focused interest scales) resulted in longer tenure. The current study definitely adds reading scores, which are likely a proxy for cognitive ability, and various potential other predictors to the literature. Given the applied utility and lack of research, it forces one to wonder why this actionable research is so void?

Practitioner and academic partnerships. Exploratory research in the public sector can be challenging. Most law enforcement agencies do not have resources, in either expertise or funding, to explore research regardless of the potential benefit. The closed culture common to law enforcement also makes academic partnership challenging. The current study is an example of the possible win-win opportunities that exist when police managers and academic researchers partner to explore a phenomenon that is of interest to both practice and literature.

Results of this study may offer no utility to anything beyond the study agency. Results of this study led to precise recommendations related to reading score prediction of transfer of training and turnover outcomes, with actionable potential interventions. Those officers leaving during FTEP who simply changed their minds are addressed with grounded theory and practice associated with

realistic job previews. Potential selection policy reviews may lead to the realization that allowing candidates to retake the selection battery costs the agency money with no concern for adverse impact. If any one of these three findings and recommendations improves officer success and retention, then the endeavor was successful.

Practitioners charged with recruitment, selection, field training, and evaluation are always on the lookout for potential means to improve outcomes. The best advice may be to conduct a thorough academic and practitioner literature search, consult with colleagues in other departments, and find an academic partner interested in conducting applied research. These partners may be faculty or graduate students working under the direct observation of faculty who are desperately trying to find a meaningful thesis or dissertation study.

References

- Abelson, M. A. (1987). Examination of avoidable and unavoidable turnover. *Journal of Applied Psychology, 72*, 382–386.
- Abelson, M. A., & Baysinger, B. D. (1984). Optimal and dysfunctional turnover: Toward an organizational level model. *The Academy of Management Review, 9*, 331–341.
- Academy of Management. (nd). Our mission and values. Retrieved from: <http://aom.org/about/>
- Ackerman, P. L., & Humphreys, L. G. (1990). Individual differences theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.) *Handbook of industrial and organizational psychology*, (Vol. 1, pp. 223–282). Palo Alto, CA: Consulting Psychologists Press.
- Agarwal, R., Ganco, M., & Ziedonis, R. (2009). Reputations for toughness in patent enforcement: Implications for knowledge spillovers via inventor mobility. *Strategic Management Journal, 30*, 1349–1374.
- Aguilar-Moya, R., Melero-Fuentes, D., Aleixandre-Benavent, R., & Valderrama-Zurián, J.C. (2013). Production and scientific collaboration in police training (1987-2011). *Policing, 36*(4), 768-786.
- Aguinis, H., & Kraiger, K. (2009). Benefits of training and development for individuals and teams, organizations, and society. *Annual Review of Psychology, 60*, 451–74.

- Albemarle Paper Company v. Moody*, 422 U.S. 405 (1975).
- Allen, D. G., Hancock, J. I., Vardaman, J. M., & McKee, D. N. (2014). Analytical mindsets in turnover research. *Journal of Organizational Behavior*, 35, S61–S86.
- Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, 60(1), 59-72.
- Allison, P.D. (2014). Measures of fit for logistic regression (paper number 1485-2014). *SAS Institute Inc.* Retrieved May 21, 2019 from https://pdfs.semanticscholar.org/cea1/65dfb9094361e5cc37ecc108a85961c5fc52.pdf?_ga=2.118053065.1511151885.1559436714-245417319.1559436714
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C.: American Educational Research Association.
- Americans With Disabilities Act of 1990, Pub. L. No. 101-336, § 1, 104 Stat. 328 (1990).
- Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational Researcher*, 25(4), 5–11.

- Anderson, N. (2011). Perceived job discrimination: Toward a model of applicant propensity to case initiation in selection. *International Journal of Selection and Assessment*, 19(3), 229-244.
- Anderson, N., Salgado, J. F., & Hülsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment*, 18(3), 291-304.
- Annell, S., Lindfors, P., & Sverke, M. (2015). Police selection – implications during training and early career. *Policing*, 38, 221-238.
- Arnold, J. (2004). The congruence problem in John Holland's theory of vocational decisions. *Journal of Occupational and Organizational Psychology*, 77(1), 95-113.
- Arthur, W., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: An assessment of its criterion-related validity. *Journal of Applied Psychology*, 91(4), 786-801.
- Arthur, W., Jr., Bennett, W., Jr., Edens, P. S., & Bell, S. T. (2003). Effectiveness of training in organizations: A meta-analysis of design and evaluation features. *Journal of Applied Psychology*, 88, 234-245.

- Association to Advance Collegiate Schools of Business (2010, April). Applying the business school conceptual framework: An approach for individual business schools. Retrieved from <https://www.aacsb.edu/-/media/aacsb/docs/business-education-intelligence/tools/applying-the-framework.ashx?la=en>
- Astakhova, M.N. (2016). Explaining the effects of perceived person-supervisor fit and person-organization fit on organizational commitment in the U.S. and Japan. *Journal of Business Research*, 69, 956-963.
- Astakhova, M. N., Beal, B. D., Schriesheim, C. A., & Camp, K. M. (2015). The dark side of job fit: A three-country study of the fit-risk taking propensity relationship. *Academy of Management Proceedings*, 12296-12296.
- Arthur Jr, W., Bell, S. T., Villado, A. J., & Doverspike, D. (2006). The use of person-organization fit in employment decision making: an assessment of its criterion-related validity. *Journal of Applied Psychology*, 91(4), 786.
- Baker, T. A., & Gebhardt, D. L. (2012). The assessment of physical capabilities in the workplace. In N. Schmitt (Eds.), *Handbook of workplace assessment: Selecting and developing organizational talent* (pp. 274-296), New York: Oxford University Press.
- Baldwin, T. T. (1992). Effects of alternative modeling strategies on outcomes of interpersonal-skills training. *Journal of Applied Psychology*, 77(2), 147-154.

- Baldwin, T.T., & Ford, J. K. (1988). Transfer of training: A review and directions for future research. *Personnel Psychology*, 41, 63–105.
- Baldwin, T. T., Ford, J. K., & Blume, B. D. (2009). Transfer of training 1988–2008: An updated review and agenda for future research. *International Review of Industrial and Organizational Psychology*, 24, 41-70
- Baldwin, T. T., Kevin Ford, J. and Blume, B. D. (2017). The state of transfer of training research: Moving toward more consumer-centric inquiry. *Human Resource Development Quarterly*, 28. 17-28.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*, Prentice-Hall, Englewood Cliffs, NJ.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman and Co.
- Barrett, G. V., Doverspike, D., & Young, C. (2010). *The special case of public sector police and fire selection*. In J. C. Scott & D, Reynolds. *Handbook of workplace assessment: Evidence-based practices for selecting and developing organizational talent*. (pp. 437-462) San Francisco, CA: Jossey-Bass.

- Barret, G., Polonsky, M., & McDaniel, M. (1999). Selection tests for firefighters: A comprehensive review and meta-analysis. *Journal of Business and Psychology, 13*, 507-514.
- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel psychology, 44*(1), 1-26.
- Barrick, M. R., & Zimmerman, R. D. (2005). Reducing voluntary, avoidable turnover through selection. *Journal of Applied Psychology, 90*(1), 159.
- Barrick, M. R., & Zimmerman, R. D. (2009). Hiring for retention and performance. *Human Resource Management, 48*(2), 183-206.
- Bassok, M., & Holyoak, K. J. (1989). Interdomain transfer between isomorphic topics in algebra and physics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 153–166.
- Bartol, C.R., Bergen, G.T., Volckens, J.S. and Knoras, K.M. (1992). Women in small-town policing: Job performance and stress. *Criminal Justice and Behavior, 19*, 240-259.
- Bauer, T. N., Maertz, C., Dolen, M., & Campion, M.A. (1998). A longitudinal assessment of applicant reactions to an employment test. *Journal of Applied Psychology, 83*, 892–903.
- Bayley, D.H., & Bittner, E. (1984) Learning the skills of policing. *Law and Contemporary Problems, 47*, 35-59.

- Becker, T. E. (1998). Integrity in organizations: Beyond honesty and conscientiousness. *Academy of Management: Academy of Management Review*, 23(1), 154-161.
- Becker, B., & Huselid, M. (2006). Strategic human resource management: Where do we go from here? *Journal of Management*, 32(6), 898–925.
- Bergman, M. E., Drasgow, F., Donovan, M. A., Henning, J. B., & Juraska, S. E. (2006). Scoring situational judgment tests: Once you get the data, your troubles begin. *International Journal of Selection and Assessment*, 14(3), 223-235.
- Bernardin, H.J., & Russell, J.E.A. (2013). *Human resource management (6th ed.)*. New York, NY: McGraw-Hill.
- 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.
- Bhatti, M.A., & Kaur, S. (2010). The role of individual and training design factors on training transfer. *Journal of European Industrial Training*, 34(7), 656-672.
- Biddle, D.A., & Morris, S.B. (2011). Using Lancaster's mid-P correction to the Fisher's exact test for adverse impact analyses. *Journal of Applied Psychology*, 96, 956–965.

- Biddle, R.E. & Biddle, D.A. (2013). What public-sector employers need to know about promotional practices, procedures, and tests in public safety promotional processes: After *Ricci v. DeStefano*. *Public Personnel Management*, 42(2), 151-190.
- Bing, M. N., Kluemper, D., Davison, H.K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, 116(1), 148-162.
- Birdi, K., Clegg, C., Patterson, M., Robinson, A., Stride, C. B., Wall, T. D., & Wood, S.J. (2008). The impact of human resource and operational management practices on company productivity: A longitudinal study. *Personnel Psychology*, 61, 467-501.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T. and Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment* 14(4), 317–335.
- Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its relationship to performance. *Personnel Psychology*, 62, 229–258.
- Bluedorn, A. C. (1978). A taxonomy of turnover. *Academy of Management Review*, 3, 647-651.
- Blume, B. D., Ford, J. K., Baldwin, T. T. & Huang, J. L. (2010). Transfer of training: A meta-analytic review. *Journal of Management*, 36, 1065–1105.

- Bobko, P., & Roth, P.L. (2013). Reviewing, categorizing, and analyzing the literature on black–white mean differences for predictors of job performance: verifying some perceptions and updating/correcting others. *Personnel Psychology*, 66, 91–126.
- Bonsignore v The City of N.Y.*, 521 F. Supp. 394 (S.D.N.Y.1981), *aff'd* 683 F.2d 635 (2d Cir. 1982).
- Borneman, M.J. (2016). Future considerations in SJT development. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 55-59.
- Bourke, B. (2014). Positionality: Reflecting on the research process. *The Qualitative Report*, 19(33), 1-9.
- Bowen, D.E., Ledford, G.E., & Nathan, B.R. (1991). Hiring for the organization, not the job. *Academy of Management Executive*, 5(4), 35-51.
- Bradford, D., & Pynes, J. (1999). Police academy training: Why hasn't it kept up with practice? *Police Quarterly*, 2, 283–301.
- Bradley, D., & Nixon, C. (2009). Ending the ‘dialogue of the deaf’: Evidence and policing policies and practices, an Australian case study. *Police Practice and Research: An International Journal*, 10, 423–435.
- Breaugh, J. A. (2003). Effect size estimation: Factors to consider and mistakes to avoid. *Journal of Management*, 29(1), 79-97.

- Bretz Jr, R. D., & Judge, T. A. (1994). The role of human resource systems in job applicant decision processes. *Journal of Management*, 20(3), 531-551.
- Brough, P. (2005). A comparative investigation of the predictors of work-related psychological well-being within police, fire and ambulance workers. *New Zealand Journal of Psychology*, 34(2), 127-134.
- Brown, N.A., Jones, A.B., Serfass, D.G., & Sherman, R.A. (2016). Reinvigorating the concept of situational judgment tests. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 38-42.
- Brunetto, Y., Teo, S., Shacklock, K., Farr-Wharton, R., & Shriberg, A. (2017). The impact of supervisor-subordinate relationships and a trainee characteristic upon police officer work outcomes. *Journal of Management and Organization*, 23(3), 423-436.
- Buerger, M. (1998). Police training as a pentecost: Using tools singularly ill-suited to the purpose of reform. *Police Quarterly*, 1, 27-63.
- Buerger, M. E. (2010). Policing and research: Two cultures separated by an almost-common language. *Police Practice & Research*, 11(2), 135-143.
- Burbeck, E., & Furnham, A. (1985). Police officer selection: A critical review of the literature. *Journal of Police Science and Administration*, 13(1), 58-69.
- Burke, L. A., & Baldwin, T. T. (1999). Workforce training transfer: A study of the effect of relapse prevention training and transfer climate. *Human Resource Management*, 38(3), 227-241.

- Burkhart, B. R. (1980). Conceptual issues in the development of police selection procedures. *Professional Psychology, 11*(1), 121-129.
- Burns, G. N., Shoda, E. A., & Roebke, M. A. (2017). Putting applicant faking effects on personality tests into context. *Journal of Managerial Psychology, 32*(6), 460-468.
- Burton, J., Holtom, B., Sablinski, C., Mitchell, T., & Lee, T. (2010). The buffering effects of job embeddedness on negative shocks. *Journal of Vocational Behavior, 76*, 42–51.
- Call, M. L., Nyberg, A. J., Ployhart, R. E., & Weekley, J. (2015). The dynamic nature of collective turnover and unit performance: The impact of time, quality, and replacement. *Academy of Management Journal, 58*, 1208–1232.
- Callinan, M., & Robertson, I. T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8*(4), 248-260.
- Calvasina, G.E., & Calvasina, R.V. (2016). Using personality testing as part of the employee selection process: Legal and policy issues for employers. *Journal of Legal, Ethical and Regulatory Issues, 19*, 112-120.
- Campbell, D. J., Christman, B. D., & Feigelson, M. E. (2000). Improving the recruitment of women in policing: An investigation of women's attitudes and job preferences. *Police Chief, 67*, 18–28.

- Campbell, J. (1971). Personnel training and development. *Annual Review of Psychology*, 22, 565–602.
- Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35–70). San Francisco, CA: Jossey-Bass.
- Campbell, K. (2010). Will “equal” again mean equal? Understanding *Ricci v. DeStefano*. *Texas Review of Law and Politics*, 14(2), 387-422.
- Campion, M. C., Ployhart, R. E., & MacKenzie Jr., W. I. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance*, 27(4), 283-310.
- Caro, C. A. (2011). Predicting state police officer performance in the field training officer program: What can we learn from the Cadet’s performance in the training academy? *American Journal of Criminal Justice*, 36, 357-370.
- Cavanagh, M. E. (2003). *Policing within a professional framework*. Upper Saddle River, NJ: Pearson Prentice Hall.
- Center for State and Local Government Excellence. (2018). *State and local government workforce: 2018 data and 10 year trends*. Washington, DC: International Public Management Association for Human Resources.

- Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology*, 82(1), 143.
- Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance*, 15, 233-254.
- Chan, D., Schmitt, N., DeShon, R. P., Clause, C. S., & Delbridge, K. (1997). Reactions to cognitive ability tests: The relationships between race, test performance, face validity perceptions, and test-taking motivation. *Journal of Applied Psychology*, 82, 300-310.
- Chappell, A.T. (2007). Community policing: Is field training the missing link? *Policing: An International Journal of Police Strategies & Management*, 30, 498-517.
- Chen, L., Fan, J., Zheng, L., & Hack, E. (2016). Clearly defining constructs and specific situations are the currency of SJTs. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 34-38.
- Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and meta-analytic of their criterion-related validities. *Personnel Psychology*, 63, 83-117.
- Chatman, J.A. (1989). Improving interactional organizational research: A model of person-organization fit. *Academy of Management Review*, 14, 333-349.

- Chuang, C. & Liao, H. (2010). Strategic human resource management in service context: Taking care of business by taking care of employees and customers. *Personnel Psychology*, 63, 153–96.
- City of Canton, Ohio v. Harris*. 109 S. Ct. 1197, 489 U. S. 378 (U. S. Ohio, 1989).
- Civil Rights Act of 1991 § 109, 42 U.S.C. § 2000e et seq (1991).
- Clark, H. C. (2010). *Field training officer program: Beyond San Jose*. Retrieved on February 12, 2004 from:
http://www.fdle.state.fl.us/fcjei/slppapers/clark_hc.pdf
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Veronica, S. H. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology*, 86(3), 410-417.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cohen, D.B., Aamodt, M.G., & Dunleavy, E.M. (2010). *Technical advisory committee report on best practices in adverse impact analyses*. Washington, DC: Center for Corporate Excellence.
- Colquitt, J., LePine, J. & Noe, R. (2000). Towards an integration theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678-707.

- Converse, P., Peterson, M., & Griffith, R. (2009). Faking on personality measures: Implications for selection involving multiple predictors. *International Journal of Selection and Assessment*, 17, 47–60.
- COPS Office. (2001). *PTO: An overview and introduction. A problem-based learning manual for training and evaluating police trainees*. Washington, DC: US Department of Justice.
- Coutts, L., & Schneider, F. (2004). Police officer performance appraisal systems: How good are they? *Policing: An International Journal of Police Strategies and Management*, 27(1), 67–81.
- Cox, S.M. (1996). *Police: Practices, perspectives, problems*. Needham Heights, MA: Allyn & Bacon.
- Crook, A.E. (2016). Unintended consequences: Narrowing SJT usage and losing credibility with applicants. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 59-63.
- Cucina, J. M., Su, C., Busciglio, H. H., Thomas, P.H., & Peyton, S.T. (2015). Video-based testing: A high-fidelity job simulation that demonstrates reliability, validity, and utility. *International Journal of Selection and Assessment*, 23, 197–209.
- Dalessio, A.T. (1994). Predicting insurance agent turnover using video-based situation judgement test. *Journal of Business and Psychology*, 9(1), 23-32.

- Dalton, D. R., Tudor, W.D., & Krackhardt, D. M. (1982). Turnover overstated: The functional taxonomy. *Academy of Management Review*, 7, 117-123.
- Dalton, D. R., Krackhardt, D. M., & Porter, L. W. (1981). Functional turnover: An empirical assessment. *Journal of Applied Psychology*, 66, 716-721.
- Daniels, S., & King, E. (2002), The predictive validity of MMPI-2 content scales for small-town police officer performance. *Journal of Police and Criminal Psychology*, 17(2), 54-60.
- 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.
- Dalen, L. H., Stanton, N. A., & Roberts, A. D. (2001). Faking personality questionnaires in personnel selection. *Journal of Management Development*, 20(8), 729-742.
- Dalessio, A. T. (1994). Predicting insurance agent turnover using video-based judgment test. *Journal of Business and Psychology*, 9(1), 23 – 32.
- D'Aquila, B.J. (2010). A management employment lawyer's perspective on *Ricco v. DeStefano*. *ABA Journal of Labor and Employment Law*, 25(2), 213-225.
- Dawis, R. V., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual differences model and its application*. Minneapolis, MN: University of Minnesota.

- De Cooman, R., Gieter, S. D., Pepermans, R., Hermans, S., Bois, C. D., Caers, R., & Jegers, M. (2009). Person–organization fit: Testing socialization and attraction–selection–attrition hypotheses. *Journal of Vocational Behavior*, 74(1), 102-107.
- de Meijer, L.A. L., Born, M. P., van Zielst, J., & van der Molen, H.T. (2010). Construct-driven development of a video-based situational judgment test for integrity: A study in a multi-ethnic police setting. *European Psychologist*, 15(3), 229-236.
- Dess, G. G., & Shaw, J. D. (2001). Voluntary turnover, social capital, and organizational performance. *Academy of management review*, 26(3), 446-456.
- Dejong, C. S., Mastrofski, & Parks, R.B. (2001) Patrol officers and problem solving: An application of expectancy theory. *Justice Quarterly*, 18, 31-61.
- Dickerson, J.D. (2009). *Ricci v. DeStefano*: A failed resolution. *Berkeley Journal of Employment & Labor Law*, 30(2), 557-570.
- Doener, W. G., & Hunter, R. D. (2006). Post-FTO performance evaluations of rookie police officers. *Journal of Ethnicity in Criminal Justice*, 4(1/2), 113-128.
- Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance*, 16(1), 81-106.

Dulebohn, J.H., Bommer, W.H., Liden, R.C., Brouer, R.L., & Ferris, G.R. (2012).

A meta-analysis of antecedents and consequences of leader-member exchange: Integrating the past with an eye toward the future. *Journal of Management*, 38, 1715-1759.

Dwight, S. A. (1999). *An assessment of the different effects of warning applicants not to fake*. (Unpublished doctoral dissertation), State University of New York at Albany, Albany, NY.

Dwyer, J. M. (2015). *Florida law enforcement academy performance and state certification examination scores* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3718786).

Edwards, J. R. (2008). Four person–environment fit in organizations: An assessment of theoretical progress. *Academy of Management Annals*, 2(1), 167-230.

Edwards, J. R. (1991). Person–job fit: A conceptual integration, literature review, and methodological critique. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 283–357). New York: Wiley.

Equal Employment Opportunity Commission. (1978). Uniform guidelines on employee selection procedures. *Federal register*, 43(166), 38295-38309

- Equal Employment Opportunity Commission. (1995). ADA enforcement guidance: Preemployment disability-related questions and medical examinations. *Appendix to enforcement guidance.*
- Equal Employment Opportunity Commission (2010). Employment Tests and Selection Procedures: *Fact Sheet on employment tests and selection procedures.* Washington D.C.: The U.S. Equal Employment Opportunity Commission.
- England, K. V. L (1994). Getting personal: Reflexivity, positionality, and feminist research. *The Professional Geographer*, 46(1), 80–89.
- Espinoza, C., Ukleja, M., & Rusch, C. (2010). *Managing the millennials: Discover the core competencies for managing today's workforce.* Hoboken, New Jersey: John Wiley and Sons.
- Facteau, J. D., Dobbins, G. H., Russell, J. E. A., Ladd, R. T., & Kudisch, J. D. (1995). The influence of general perceptions of the training environment on pretraining motivation and perceived training transfer. *Journal of Management*, 21, 1-25.
- Fagan, S. L. (2017). *Trainees' perceptions on supervisor factors that influence transfer of training* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 10688591)

- Fan, J., Stuhlman, M., Chen, L., & Weng, Q. (2016). Both general domain knowledge and situation assessment are needed to better understand how SJTs work. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 43-47.
- Farrington, J. (2011). Training transfer: Not the 10% solution. *Performance Improvement Quarterly, 24*, 117-121.
- Feldman, D. C., Ng, T. W., & Vogel, R. M. (2012). Off-the-job embeddedness: A reconceptualization and agenda for future research. *Research in Personnel and Human Resources Management, 31*, 209–251.
- Feldman, D. C., & Ng, T. W. (2007). Careers: Mobility, embeddedness, and success. *Journal of Management, 33*, 350–377.
- Fell, C.B., & König, C.J. (2016), Cross-cultural differences in applicant faking on personality tests: A 43-nation study, *Applied Psychology, 65*(4), 671-717.
- Fetzer, M., Tuzinski, K., & Freeman, M. (2010). 3D animation, motion capture, and SJTs: IO is finally catching up with IT. Proceedings from the *25th Annual Conference of Industrial and Organizational Psychology, Atlanta, GA*.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (5th ed.). Thousand Oaks, CA: Sage.

- Fischer, M.A. (2015). *IACP best practices guide: Field training for today's recruits*. Available from <http://www.theiacp.org/portals/0/documents/pdfs/BP-FieldTrainingforTodaysRecruits.pdf>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological bulletin*, 51 (4), 327.
- Florida Department of Law Enforcement. (2011). Criminal justice standards and training commission: Field training officer course for law enforcement officers (Course 890). Tallahassee, FL: US.
- Funke, U., & Schuler, H. (1998). Validity of stimulus and response components in a video test of social competence. *International Journal of Selection and Assessment*, 6(2), 115-123.
- Fry, W. (2017, May 18). Thousands apply to be SDPD officers each year, few make the cut [Video file]. Retrieved from <https://www.nbcsandiego.com/news/local/Thousands-Apply-to-be-SDPD-Officers-Each-Year-423089874.html>
- Gaines, L. K., & Falkenberg, S. (1998). An evaluation of the written selection test: Effectiveness and alternatives. *Journal of Criminal Justice*, 26(3), 175–183.
- Gardner, T. (2005). Interfirm competition for human resources: Evidence from the software industry. *Academy of Management Journal*, 48, 237–256.

- Gaudine, A. P., & Saks, A. M. (2004). A longitudinal quasi-experiment on the effects of posttraining transfer interventions. *Human Resource Development Quarterly*, 15(1), 57–76.
- Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, C. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology*, 72, 493–511.
- Geller, W.A., & Toch, H. (1995). *And justice for all: Understanding and controlling police abuse of force*. Washington, DC: Police Executive Forum.
- Georgenson, D. L. (1982). The problem of transfer calls for partnership. *Training and Development Journal*, 36(10), 75–78.
- Gist, M. E., Stevens, C. K., & Bavetta, A. G. (1991). Effects of self-efficacy and post-training intervention on the acquisition and maintenance of complex interpersonal skills. *Personnel Psychology*, 44, 837–861.
- Glenn, R.W., Raymond, B., Barnes-Proby, D., Williams, E., Christian, J., Lewis, M.W., Gerwehr, S., & Brannan, D. (2003). *Training the 21st century police officer: redefining police professionalism for the Los Angeles Police Department*. Santa Monica, CA: RAND Corporation.
- Gold, A. (2015, June 5). Why has the murder rate in some US cities suddenly spiked? *BBC News*. Retrieved from <http://www.bbc.com/news/world-us-canada-32995911>

- Goldstein, H. (1979). Improving policing: A problem-oriented approach. *Crime and Delinquency* 25(2), 236-258.
- Goldstein, I., & Ford, J. K. (2002). *Training in organizations* (4th ed.). Belmont, CA: Wadsworth.
- Goldstein, I. L., Zedeck, S., & Schneider, B. (1993). *An exploration of the job analysis- content validity process: Personnel selection in organizations*. San Francisco, CA: Jossey- Bass Publishers
- Gonder, M., & Gilmore, D.C. (2004), Validity study: Personality profiles of police officers who successfully complete academy training, *Applied HRM Research*, 9(2), 59-62
- Gottfredsonm, L., (1996, October). Racially Gerrymandered Police Tests. *Wall Street Journal*.
- Gottfredson, L. (1996). Racially gerrymandering the content of police tests to satisfy the U.S. Justice Department: A case study. *Psychology, Public Policy, and Law*, 2(3), 418-446.
- Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? *Journal of Business and Psychology*, 16(4), 573-592.
- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium. *Journal of Management*, 26, 463–488.

- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review*, 36, 341–357.
- Griffith, R. L., & Converse, P. D. (2011). The rules of evidence and the prevalence of applicant faking. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 34–52). New York, NY: Oxford University Press.
- Griffith, R.L., Jones, M.A., McChrystal, E., & Kung, M. (2002). Florida law enforcement recruiting and retention series part seven: Managing members of generation X. *The Florida Police Chief*. 28(10), 19-22.
- Griffith, R.L., & Peterson, M.H. (2011). One piece at a time: The puzzle of applicant faking and a call for theory. *Human Performance*, 24(4), 291-301.
- Griggs v. Duke Power Co.*, 401 U.S. 424 (1971).
- Grohmann, A., Beller, J. & Kauffeld, S. (2014), Critical role of motivation to transfer. *International Journal of Training and Development*, 18, 84-103.
- Grossman, R., & Burke-Smalley, L.A. (2018). Context-dependent accountability strategies to improve the transfer of training: A proposed theoretical model and research propositions. *Human Resource Management Review*, 28(2), 234-247.

- Guion, R. M. (1991). Personnel assessment, selection and placement. In M.D. Dunnette & L.M. Hough (Eds.). *Handbook of industrial and organizational psychology* (pp. 327-397). Palo Alto: Consulting Psychologists Press.
- Guion, R. M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah: Lawrence Erlbaum.
- Guion, R. M., & Highhouse, S. (2004). *Essentials of personnel assessment and selection*. New York: Psychology Press.
- Guller, I. (1994). *Evolution of a bio-data instrument for the screening of police and other public safety personnel*. Oakland, NJ: Institute for Forensic Psychology.
- Guller, M. (2004). Predicting performance of law enforcement personnel using the candidate and officer personnel survey and other psychological measures (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3130121).
- Gully, S. M., Payne, S. C., Koles, K. L., & Whiteman, J. K. (2002). The impact of error training and individual differences on training outcomes: An attribute-treatment interaction perspective. *Journal of Applied Psychology*, 87, 143-155.
- Gutman, A. (2009). Major EEO issues related to personnel selection decisions. *Human Resource Management Review*, 19(3), 232-250.

- Gutman, A., Koppes, L.L., & Vodanovich, S.J. (2011). *EEO law and personnel practices* (3rd ed.). New York: Taylor and Francis.
- Haarr, R.N. (2001). The making of a community policing officer: The impact of basic training and occupational socialization on police recruits. *Police Quarterly*, 4, 402-433.
- Haberfield, M.R. (2002). *Critical issues in police training*. Upper Saddle River, NJ: Prentice-Hall.
- Hale, D., Ployhart, R., & Shepherd, W. (2015). A two-phase longitudinal model of a turnover event: Disruption, recovery rates, and moderators of collective performance. *Academy of Management Journal*, 59, 906–929.
- Hancock, J. I., Allen, D. G., Bosco, F. A., McDaniel, K. R., & Pierce, C. A. (2013). Meta-analytic review of employee turnover as a predictor of firm performance. *Journal of Management*, 39, 573–603.
- Hatch, N., & Dyer, J. (2004). Human capital and learning as a source of sustainable competitive advantage. *Strategic Management Journal*, 25, 1155–1178.
- Hanson, M. A., & Ramos, R. A. (1996). Situational judgment tests. *Fair employment strategies in human resource management*, 119-124.
- Hargrave, G.E., Hiatt, D. and Gaffney, T.W. (1986). A comparison of MMPI and CPI test profiles for traffic officers and deputy sheriffs. *Journal of Police Science & Administration*, 14, 250-258.

- Harris, A.M., Siedor, L.E., Fan, Y., Listyg, B., & Carter, N.T. (2016). In defense of the situation: An interactionist explanation for performance on situational judgment tests. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 23-28.
- Harris, W. E. (2014). *Examining the relationship between police officer education and job performance* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global. (Order No. 3625852).
- Harter, J.K., Schmidt, F.L., & Hayes, T.L. (2002). Business-unit-level relationship between employee satisfaction, employee engagement, and business outcomes: A meta-analysis. *Journal of Applied Psychology, 87*, 268–279
- Harvey, R.J. (2016). Scoring SJTs for traits and situational effectiveness. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 63-71.
- Hausknecht, J.P., Day, D.V., & Thomas, S.C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 53*, 639-683.
- Hausknecht, J. P., & Holwerda, J. A. (2013). When does employee turnover matter? Dynamic member configurations, productive capacity, and collective performance. *Organization Science, 24*(1), 210-225.

- Hausknecht, J. P., & Langevin, A. M. (2010). Selection for service and sales jobs. In J. L. Farr & N. Tippins (Eds.), *Handbook of employee selection* (pp. 765–779). New York: Routledge, Taylor, & Francis.
- Hausknecht, J. P., Trevor, C. O., & Howard, M. J. (2009). Unit-level voluntary turnover rates and customer service quality: Implications of group cohesiveness, newcomer concentration, and size. *Journal of Applied Psychology, 94*, 1068–1075.
- Heavey, A. L., Holwerda, J. A., & Hausknecht, J. P. (2013). Causes and consequences of collective turnover: A meta-analytic review. *Journal of Applied Psychology, 98*(3), 412.
- Heimbeck, D., Frese, M., Sonnentag, S., & Keith, N. (2003). Integrating errors into the training process: The function of error management instructions and the role of goal orientation. *Personnel Psychology, 56*, 333-361.
- Henson, B., Reynolds, B., Klahm, C. and Frank, J. (2010). Do good recruits make good cops? Problems predicting and measuring academy and street-level success. *Police Quarterly, 13*(1), 5-26.
- Hirsh, H., Schmidt, F., & Hunter, J. (1986) Estimation of employment validities by less experienced judges. *Personnel Psychology, 39*, 337-344.

- Hitt, M., Bierman, L., Shimizu, K., & Kochhar, R. (2001). Directing and moderating effects of human capital on strategy and performance in professional firms: a resource based perspective. *Academy of Management Journal*, 44(1), 13–28.
- Ho, T. (2001). The interrelationships of psychological testing, psychologists' recommendations, and police departments' recruitment decisions. *Police Quarterly*, 4, 318-342.
- Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology*, 37(2), 379-86.
- Holladay, C. L., & Quiñones, M. A. (2003). Practice variability and transfer of training: The role of self-efficacy generality. *Journal of Applied Psychology*, 88, 1094-1103.
- Holtom, B. C., Mitchell, T. R., Lee, T. W., & Eberly, M. B. (2008). Turnover and retention research. *The Academy of Management Annals*, 2, 231–274.
- Holton, E., III, Bates, R., & Ruona, W. (2000). Development of a generalized learning transfer system inventory. *Human Resource Development Quarterly*, 11, 333–360.
- Hom, P. (2011). Organizational exit. In S. Zedeck (Ed.), *Handbook of industrial/organizational psychology* (Vol. 2, pp. 67–117). Washington, DC: American Psychological Association.

- Hom, P. W., & Griffeth, R. W. (1995). *Employee turnover*. Cincinnati, OH: South-Western College Publishing.
- Hom, P. W., Lee, T. W., Shaw, J. D., & Hausknecht, J. P. (2017). One hundred years of employee turnover theory and research. *Journal of Applied Psychology, 102*(3), 530.
- Hom, P. W., Mitchell, T. R., Lee, T. W., & Griffith, R. W. (2012). Reviewing employee turnover: Focusing on proximal withdrawal states and an expanded criterion. *Psychological Bulletin, 138*(5), 831-858.
- Hogg, A., & Wilson, C. (1995) *Is the psychological screening of police applicants a realistic goal? The successes and failures of psychological screening*. (Report Series 124.) Payneham, South Australia: National Police Research Unit.
- Hooper, A. C., Cullen, M. J., & Sackett, P. R. (2006). Operational threats to the use of SJTS: Faking, coaching, and retesting issues. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hooper, A. C., Jackson, H. L., & Motowidlo, S. J. (2004). Proceedings from the Society for Industrial and Organizational Psychology: *Faking on personality-based measures: SJTs compared to a traditional format*. Chicago, IL.

- Hoover, J. H. (2002). *The Reno model of post-academy training* (Executive Summary). Reno, NV Police Department.
- Hoover, J., Pitts, S., & Ponte, D. (2003) *Reno Police Department PTO Manual*. Reno, NV: Reno Police Department.
- Hosmer, D.W., Hosmer, T., Cessie, S.L., & Lemeshow, S. (1997). A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in Medicine*, 16, 965–980.
- Hough, L. M. (1998). The millennium for personality psychology: New horizons or good old daze. *Applied Psychology*, 47, 233-261.
- Hough, L.M., Eaton, N.K., Dunnette, M.D., Kamp, J.D., & McCloy, R.A. (1990) Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, 75, 581–595.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment*, 9(1-2), 152-194.

- Hough, L., & Paullin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *Biodata handbook: Theory, research, and use of biographical information in selection and performance prediction* (pp. 109-145). Palo Alto, CA, US: CPP Books.
- Hsieh, F.Y. (1989). Sample size tables for logistic regression. *Statistics in Medicine*, 8, 795-802.
- Huang, J.L., Blume, B.D., & Ford, J.K. (2015). A tale of two transfers: Disentangling maximum and typical transfer and their respective predictors. *Journal of Business and Psychology*, 30, 709-732.
- Hülsheger, U. R., & Anderson, N. (2009). Applicant perspectives in selection: Going beyond preference reactions. *International Journal of Selection and Assessment*, 17(4), 335-345.
- Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment*, 19(1), 62-81.
- Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology*, 13(4), 373-386.
- Hunter, J.E., and Schmidt, F.L. (1990). Dichotomization of continuous variables: The implications for meta-analysis. *Journal of Applied Psychology*, 75, 334-349.

- Husbands, A., Rodgerson, M.J., Dowell, J., & Patterson, F. (2015). Evaluating the validity of an integrity-based situational judgment test for medical school admissions. *BMC Medical Education*, 15, 144-152.
- Hutchinson, S. T., Villalobos, J. R., & Beruvides, M. G. (1997). Effects of high turnover in a serial assembly environment. *International Journal of Production Research*, 35, 3201-3223.
- Huy, L. & Marcus, J. (2012). The overall odds ratio as an intuitive effect size index for multiple logistic regression. *Educational and Psychological Measurement*, 72, 6.
- Inwald R., Hurwitz, H., & Kaufman, J. (1989). Uncertainty reduction in retail and public safety/private security screening. *Forensic Reports* 4(2), 171–212.
- Inwald, R., Kaufman, J., & Roberts, M. (1991). Alcohol use, drug use, and past psychiatric history. *Criminal Justice Digest* 10(7), 1–8.
- Inwald, R.E., & Shusman, E.J. (1984). The IPI and MMPI as predictors of academy performance for police recruits. *Journal of Police Science & Administration*, 12(1), 1-11.
- Janik, J. (1994). Why psychological screening of police candidates is necessary: the history and rationale. *Journal of Police and Criminal Psychology*, 10(2), 18–23.
- Jepsen, D. A., & Dilley, J. S. (1974). Vocational decision-making models: A review and comparative analysis. *Review of educational research*, 331-349.

- Johnson v. Cincinnati, 39 F. Supp 2d 1013 (S. D. Ohio, 1999).
- Johnsen, B. H., Espevik, R., Saus, E., Sanden, S., Olsen, O. K., & Hystad, S. W. (2017). Hardiness as a moderator and motivation for operational duties as mediator: The relation between operational self-efficacy, performance satisfaction, and perceived strain in a simulated police training scenario. *Journal of Police and Criminal Psychology, 32*, 331-339.
- Jones, M.A., McChrystal, E., & Kung, M. (2002). Florida law enforcement recruiting and retention series part one: An introduction and overview. *The Florida Police Chief, 28*(4), 21.
- Jones, M.A., McChrystal, E., Kung, M. & Griffith, R.L. (2002). Florida law enforcement recruiting and retention series: An introduction and overview. *The Florida Police Chief, 28*(4).
- Jones, M.A., McChrystal, E., Kung, M. & Griffith, R.L. (2002). Florida law enforcement recruiting and retention series part two: Recruiting and retention: A singular problem. *The Florida Police Chief, 28*(5), 21-22.
- Jordan, M. H., Lindsay, D. R., & Schraeder, M. (2012). An examination of salient, non-monetary, factors influencing performance in public sector organizations: A conceptual model. *Public Personnel Management, 41*, 661-684.

- Jordan, W. T., Fridell, L., Faggiani, D., & Kubu, B. (2009). Attracting females and racial/ethnic minorities to law enforcement. *Journal of Criminal Justice*, 37(4), 333-341.
- Juraska, S. E., & Drasgow, F. (2001, April). Faking situational judgment: A test of the Conflict Resolution Skills Assessment. Paper presented at the 16th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Kacmar, K. M., Andrews, M. C., Van Rooy, D. L., Steilberg, R. C., & Cerrone, S. (2006). Sure everyone can be replaced... but at what cost? Turnover as a predictor of unit-level performance. *Academy of Management journal*, 49(1), 133-144.
- Kaminsky, G. F. (2002). *The field training concept in criminal justice agencies*. Upper Saddle River, NJ: Prentice Hall.
- Kaminski, R.J. (1993). Police minority recruitment: Predicting who will say yes to an offer for a job as a cop. *Journal of Criminal Justice*, 21(4), 395-409.
- Kammeyer-Mueller, J. D., Wanberg, C. R., Glomb, T. M., & Ahlburg, D. (2005). The role of temporal shifts in turnover processes: It's about time. *Journal of Applied Psychology*, 90, 644-658.
- Kessler, E. H. (2013). *Encyclopedia of management theory*. Thousand Oaks, California: SAGE Publications, Inc.

- Kiefer, C., & Benit, N. (2016). What is applicant faking behavior? A review on the current state of theory and modeling techniques. *Journal of European Psychology Students, 7*(1), 9-19.
- Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology, 76*, 698-707.
- Koper, C. S., Maguire, E. R., & Moore, G. E. (2001). *Hiring and retention issues in police agencies: Readings on the determinants of police strength, hiring and retention of officers, and the federal COPS program*. Washington, DC: Urban Institute.
- Krackhardt, D., & Porter, L. (1985). When friends leave: A structural analysis of the relationship between turnover and stayers' attitudes. *Administrative Science Quarterly, 30*, 242-261.
- Krackhardt, D., & Porter, L. (1986). The snowball effect: Turnover embedded in communication networks. *Journal of Applied Psychology, 71*, 50-55.
- Kristof, A. L. (1996). Person-organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49*(1), 1-49.
- Kristof-Brown, A.L. (2000). Perceived applicant fit: Distinguishing between recruiters' perceptions of person-job and person-organization fit. *Personnel Psychology, 53*, 643-671.

- Kristof-Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individuals' fit at work: A meta-analysis of person-job, person-organization, person-group, and person-supervisor fit. *Personnel Psychology, 58*, 281-342
- Kuncel, N. R., & Hezlett, S. A. (2010). Fact and fiction in cognitive ability testing for admissions and hiring decisions. *Current Directions in Psychological Science, 19*(6), 339-345.
- Landy, F.J, Farr, J.L., Saal, F.D., Freytag, W.R. (1976). Behaviorally anchored scales for rating the performance of police officers. *Journal of Applied Psychology, 61*, 750–758
- Legree, P. & Psotka, J. (2006). Refining situational judgment test methods. Proceedings of ASC06: *The 25th Army Science Conference*. Department of the Army.
- Lee, T. W., & Mitchell, T. R. (1994). An alternative approach: The unfolding model of voluntary employee turnover. *The Academy of Management Review, 19*, 51–89.
- Lee, T. W., Mitchell, T. R., Holtom, B. C., McDaniel, L., & Hill, J. W. (1999). The unfolding model of voluntary turnover: A replication and extension. *Academy of Management Journal, 42*, 450–462.
- Lee, T. W., Mitchell, T. R., Wise, L., & Fireman, S. (1996). An unfolding model of voluntary employee turnover. *Academy of Management Journal, 39*, 5–36.

- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107, 34-47.
- Lewin, K. (1936). *Principles of topological psychology*. New York: McGraw-Hill
- Lievens, F. (2017). Construct-driven SJTs: Toward an agenda for future research. *International Journal of Testing*, 17(3), 269-276.
- Lievens, F., Buyse, T., & Sackett, P. R. (2005). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology*, 90(3), 442-452.
- Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management*, 41, 1604-1627.
- Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Eds.), *Handbook of assessment and selection* (pp. 383–410). New York: Oxford University Press.
- Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 3-22.

- Lievens, F., & Peeters, H. (2008). Impact of elaboration on responding to situational judgment test items. *International Journal of Selection and Assessment*, 16(4), 345-355.
- Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: A review of recent research. *Personnel Review*, 37(4), 426-441.
- Lievens, F., & Sackett, P. R. (2006). Video-based versus written situational judgment tests: A comparison in terms of predictive validity. *Journal of Applied Psychology*, 91(5), 1181.
- Lievens, F., & Sackett, P. R. (2007). Situational judgment tests in high-stakes settings: Issues and strategies with generating alternate forms. *Journal of Applied Psychology*, 92(4), 1043.
- Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology*, 97(2), 460-468.
- Lievens, F., Sackett, P.R., & Buyse, T. (2009). The effects of response instructions on situational judgment test performance and validity in a high-stakes context. *Journal of Applied Psychology*, 94, 1095–1101
- Liu, Y., Combs, J. G., Ketchen Jr, D. J., & Ireland, R. D. (2007). The value of human resource management for organizational performance. *Business Horizons*, 50(6), 503-511.

- Locke, E., & Latham, G. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ. Prentice-Hall.
- Lough, J., & Von Treuer, K. (2013). A critical review of psychological instruments used in police officer selection. *Policing*, 36(4), 737-751.
- Lough, J., Wald, E., Byrne, K., & Walker, G. (2007). The impact of psychological profiling of Australian correctional officers. *Corrections Compendium*, 32(4), 1-6.
- Lunenburg, F. C. (2012). Organizational structure: Mintzberg's framework. *International Journal of Scholarly, Academic, Intellectual Diversity*, 14(1), 1-8.
- Mabon, H. (1996). The cost of downsizing in an enterprise with job security. *Journal of Human Resource Costing & Accounting*, 1(1), 35-62.
- Maciag, M. (2018). With fewer police applicants, departments engage in bidding wars. Making the grade in today's schools. *Governing*, 31(7), 62-63.
- MacKenzie Jr, W. I., Ployhart, R. E., Weekley, J. A., & Ehlers, C. (2009). Contextual effects on SJT responses: An examination of construct validity and mean differences across applicant and incumbent contexts. *Human Performance*, 23(1), 1-21.
- Makinen, G. (2002). *Intelligence issues for Congress* (CRS Report No. RL31617). Retrieved from Congressional Research Service website:
<https://fas.org/irp/crs/RL31617.pdf>

- Maltarich, M. A., Nyberg, A. J., & Reilly, G. (2010). A conceptual and empirical analysis of the cognitive ability–voluntary turnover relationship. *Journal of Applied Psychology*, 95(6), 1058.
- Maryland Police and Correctional Training Commission. (2017). Field training and evaluation guide: A sample policy and procedure for Police Training Commission, Chapter 01 – General Regulations, Regulation 17. Retrieved from http://mdle.net/pdf/FTO_Policy-5-27-16.pdf
- Maslach, C. & Leiter, M.P. (1997). *The truth about burnout*. San Francisco: Jossey Bass.
- Mastrofski, S. D., & Ritti, R.R. (1996). Police training and the effects of organization on drunk driving enforcement. *Justice Quarterly*, 13, 291–320.
- Mathieu, J.E., Tannenbaum, S.I. and Salas, E. (1992). Influences of individual and situational characteristics on measures of training effectiveness. *Academy of Management Journal*, 35, 828-47.
- May, G.L. & Kahnweiler, W.M. (2000). The effect of a mastery practice design on learning and transfer in behavior modeling training. *Personnel Psychology*, 53, 353-73.
- McCampbell, M. (1986). *Field training for police officers: State of the art*. Washington DC: National Institute of Justice.

- McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*(1), 63-91.
- McDaniel, M. A., Kepes, S., & Banks, G. C. (2011). The Uniform Guidelines are a detriment to the field of personnel selection. *Industrial and Organizational Psychology, 4*(4), 494-514.
- McDaniel, M.A., List, S.K., & Kepes, S. (2016). The “hot mess” of situational judgment tests construct validity and other issues. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 47-51.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*(4), 730.
- McDaniel M.A., & Nguyen N.T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103 – 113.
- McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A. P., & Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal of Applied Psychology, 96*(2), 327.

- McEvoy, G.M., & Cascio, W.F. (1985). Strategies for reducing employee turnover: A meta-analysis. *Journal of Applied Psychology, 70*, 342-353.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across non-cognitive measures. *Journal of Applied Psychology, 85*, 812-821.
- McFarland, L. A., Ryan, A. M., & Aleksander, E. (2002). Item placement on a personality measure: Effects on faking behavior and test measurement properties. *Journal of Personality Assessment, 78*, 348-369.
- McPherson, M., Smith-Lovin, L., & Cook, J.M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*, 415–444.
- Meier, R.D., Farmer, R.E., & Maxwell, D. (1988). Psychological screening of police candidates: Current perspectives. *Journal of Police Science and Administration, 15*, 210-215.
- Melchers, K.G., & Kleinman, M. (2016). Why situational judgement is a missing component of the theory of SJT. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 29-34.
- Metcalf, H. C. (1912). *Industrial and social justice: Trial outline and bibliography*. Chicago, IL: Tufts College Press.
- Miller, B.H. (2015). Situational judgement tests: An exploration of reliability estimates and scoring models. (Unpublished doctoral dissertation). California School of Professional Psychology, Los Angeles, CA.

- Mitchell, C.E. (2013). An analysis of the U.S. Supreme Court's decision in *Ricci v. DeStefano*: The New Haven firefighter's case. *Public Personal Management, 41*(1), 41-54.
- Mitchell, T. R., Holtom, B. C., Lee, T. W., Sablinski, C. J., & Erez, M. (2001). Why people stay: Using job embeddedness to predict voluntary turnover. *Academy of Management Journal, 44*(6), 1102-1121.
- Mitchell, T. R., & Lee, T. W. (2001). The unfolding model of voluntary turnover and job embeddedness: Foundations for a comprehensive theory of attachment. *Research in Organizational Behavior, 23*, 189–246.
- Mobley, W. H. (1977). Intermediate linkages in the relationship between job satisfaction and employee turnover. *Journal of Applied Psychology, 62*, 237–240.
- Mobley, W. H., Griffeth, R. W., Hand, H. H., & Meglino, B. M. (1979). Review and conceptual analysis of the employee turnover process. *Psychological Bulletin, 86*, 493–522.
- Monell v. New York City Department of Social Services*, 436, U. S. 658, 98 S. Ct. 2018 (1978).
- Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60*(3), 683-729.

- Motowidlo, S. J., & Beier, M.E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, 95, 321–33.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, 75(6), 640.
- Motowidlo, S. J., & Tippins, N. (1993). Further studies of the low-fidelity simulation in the form of a situational inventory. *Journal of Occupational and Organizational Psychology*, 66(4), 337-344.
- Mumford, T.V., Van Iddekinge, C.H., Morgeson, F.P., & Campion, M.A. (2008). The team role test: Development and validation of a team role knowledge situational judgment test. *Journal of Applied Psychology*, 93, 250–67.
- Murphy, K. J. (2012). Executive Compensation: Where we are, and how we got there. In Constantinides, G., Harris, M., & Stulz, R., (Eds.), *Handbook of the Economics of Finance*, (pp. 192–196). North Holland: Elsevier Science.
- Myers, R. (2004), What future(s) do we want for community policing? In Fridell, L. and Wycoff, M.A. (Eds), *Community policing: The past, present, and future* (pp. 169-82), Washington, DC: Police Executive Research Forum.

- Naemi, B., Martin-Raugh, M., & Kell, H. (2016). SJTs as measures of general domain knowledge for multimedia formats: Do actions speak louder than words? *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 77-83.
- National Criminal Justice Reference Service (2006). *Bad cops: A study of career-ending misconduct among New York City police officers*. Washington, D.C.: American University.
- National Research Council (2004). *Fairness and effectiveness in policing: The evidence*. Washington, D.C.: National Academies Press.
- Ng, K. H. (2015). Supervisory practices and training transfer: Lessons from Malaysia. *Asia Pacific Journal of Human Resources, 53*, 221-240.
- Nguyen, N. T., Biderman, M. D., & McDaniel, M. A. (2005). Effects of response instructions on faking a situational judgment test. *International Journal of Selection and Assessment, 13*(4), 250-260.
- Noe, R. A., Sears, J., & Fullenkamp, A. M. (1990). Relapse training: Does it influence trainees' post training behavior and cognitive strategies? *Journal of Business and Psychology, 4*, 317-328.
- Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human performance, 11*(2-3), 245-269.

- Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*, 660–679.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*(4), 679-703.
- Oostrom, J. K., Köbis, N. C., Ronay, R., & Cremers, M. (2017). False consensus in situational judgment tests: What would others do? *Journal of Research in Personality, 71*, 33-45.
- O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. *Academy of Management Journal, 34*(3), 487-516.
- Orrick, W. D. (2008). *Recruitment, retention, and turnover of police personnel: Reliable, practical, and effective solutions*. Springfield, IL: Charles C. Thomas.
- Ostroff, C. (1992). The relationship between satisfaction, attitudes, and performance: An organizational level analysis. *Journal of Applied Psychology, 77*, 963-974.

- Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, 89(2), 187.
- Pallant, J. (2013). *SPSS survival guide: The step by step guide to data analysis using IBM SPSS* (5th ed.). New York: McGraw Hill.
- Peeters, H., & Lievens, F. (2005). Situational judgment tests and their predictiveness of college students' success: The influence of faking. *Educational and Psychological Measurement*, 65(1), 70-89.
- Peterson, M., Griffith, R., & Converse, P. (2009). Examining the role of applicant faking in hiring decisions: Percentage of fakers hired and hiring discrepancies in single- and multiple-predictor selection. *Journal of Business and Psychology*, 24, 1–14.
- Piasentin, K.A. and Chapman, D.S. (2006). Subjective person – organization fit: bridging the gap between conceptualization and measurement. *Journal of Vocational Behavior*, 69, 202-221.
- Pinder, C. C. (2008). *Work motivation in organizational behavior* (2nd ed.). New York, NY: Psychology Press.
- Ployhart, R. E. (2006). Staffing in the 21st century: New challenges and strategic opportunities. *Journal of Management*, 32, 868–897.

- Ployhart, R. E., & Holtz, B. C. (2008). The diversity–validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153-172.
- Ployhart, R. E., Nyberg, A. J., Reilly, G., & Maltarich, M. A. (2014). Human capital is dead: Long live human capital resources. *Journal of Management*, 40, 483-508
- Ployhart, R. E., & Ryan, A. M. (1998). The relative importance of procedural and distributive justice in determining applicants' reactions. *Journal of Applied Psychology*, 83, 3-16.
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory* (3rd ed.). Mahwah, NJ: Erlbaum.
- Ployhart, R. E., & Weekley, J. A. (2010) Strategy, selection, and sustained competitive advantage. In J. L. Farr & N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 195-212). New York, NY: Routledge.
- Police Executive Research Forum (PERF). (2013). *Policing and the economic downturn: Striving for efficiency is the new normal*. Washington, DC: PERF.
- Popovich, M. G. (1998). *Creating high-performance government organizations*. San Francisco, CA: Jossey-Bass.
- Price, J. L. (1977). *The study of turnover*. Ames: Iowa State University Press.

- Price, J.L. (2001). Reflections on the determinants of voluntary turnover. *International Journal of Manpower*, 22, 600-624.
- Pulakos, E. D., & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, 9, 241–258.
- Public and Nonprofit Division of the Academy of Management. (nd). About us. Retrieved from: <https://pnp.aom.org/about-us/>
- Quinones, M.A. (1995). Pretraining context effects: Training assignment as feedback. *Journal of Applied Psychology*, 80, 226-38.
- Raganella, A.J., & White, M.D. (2004). Race, gender, and motivation for becoming a police officer: Implications for building a representative police department. *Journal of Criminal Justice*, 32, 501-513.
- Raja, U., Bouckennooghe, D., Syed, F., & Naseer, S. (2018). Interplay between P-O fit, transformational leadership and organizational social capital. *Personnel Review*, 47(4), 913-930.
- Reaves, B.A. (2012). *Hiring and retention of state and local law enforcement officers, 2008-statistical tables*. Washington, DC: US Department of Justice.

- Reynolds, D. H., Winter, J. L. & Scott, D. R. (1999, May). Development, Validation, and Translation of a Professional-Level Situational Judgment Inventory. Paper presented at the 14th Annual Conference of the Society for Industrial and Organizational Psychology, Atlanta, GA.
- Rheem, J. (1998). *Problem based learning: An introduction*. Retrieved on February 12, 2004 from: <http://www.ntlf.com/html/pi/9812/pbl.htm>
- Ricci v. DeStefano*, 557 U.S. 557 (2009)
- Ridgeway, G., Lim, N., Gifford, B., Koper, C., Matthies, C.F., Hajiamiri, S., & Huynh, A.K. (2008). *Strategies for improving officer recruitment in the San Diego Police Department*. Santa Monica, CA.: RAND Corporation.
- Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 13, 279–301.
- Rosse, J.G., Stecher, M.D., Miller, J.L. & Levin, R.A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83, 634-644
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer III, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta- analysis. *Personnel Psychology*, 54(2), 297-330.

- Roth, P. L., Bobko, P., & Buster, M. A. (2013). Situational judgment tests: The influence and importance of applicant status and targeted constructs on estimates of Black–White subgroup differences. *Journal of Occupational and Organizational Psychology*, 86(3), 394-409.
- Rubenstein, A.L., Eberly, M.B., Lee, T.W., & Mitchell, T.R. (2017). Surveying the forest: A meta-analysis, moderator investigation, and future-oriented discussion of the antecedents of voluntary employee turnover, *Personnel Psychology*, 71(1), 23-65.
- Rynes, S. L., & Connerley, M. L. (1993). Applicant reactions to alternative selection procedures. *Journal of Business and Psychology*, 7(3), 261-277.
- Rynes, S.L. & Gerhart, B. (1990). Interviewer assessments of applicant ‘fit’: an exploratory investigation. *Personnel Psychology*, 43(1), 13-35.
- Sackett, P. R., & Ellingson, J. E. (1997). The effects of forming multi-predictor composites on group differences and adverse impact. *Personnel Psychology*, 50(3), 707-721.
- Saks, A.M. & Belcourt, M. (2006). An investigation of training activities and transfer of training in organizations. *Human Resource Management*, 45, 629-648.
- Saks, A. M. & Burke-Smalley, L. A. (2014), Transfer and firm performance. *International Journal of Training and Development*, 18, 104-115.

- Saks, A. M., Salas, E. & Lewis, P. (2014), Editorial. *International Journal of Training and Development*, 18, 81-83.
- Salas, E., Tannenbaum, S.I., Kraiger, K. & Smith-Jentsch, K.A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest*, 13(2), 74 – 101
- San Jose Police Department. (2018, June 23). *Field training officer (FTO) program*. Retrieved from <http://www.sjpd.org/bfo/fieldtraining/>
- Schaubroeck, J., Ganster, D. C., & Jones, J. R. (1998). Organization and occupation influences in the attraction-selection-attrition process. *Journal of Applied Psychology*, 83(6), 869-891.
- Schmidt, D. B., & Wolf, P. P. (2003, April). Susceptibility of SJTs to applicant faking: An examination of applicant and incumbent samples. Paper presented at the 18th annual conference of the Society for Industrial and Organizational Psychology, Orlando, FL.
- Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). The impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, 71, 432–439.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel selection: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262–274.

- Schmidt, F.L., Oh, I.S. and Shaffer, J.A. (2016). *The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings*. [Working paper]. Available from www.researchgate.net/profile/Frank_Schmidt10/publication/309203898_The_Validity_and_Utility_of_Selection_Methods_in_Personnel_Psychology_Practical_and_Theoretical_Implications_of_100_Years_of_Research_Findings/links/58051e2608aef87fbf3bbb52.pdf
- Schmitt, N., & Ott-Holland, C. (2012). Theoretical and practical issues: Research needs. In *The Oxford Handbook of Personnel Assessment and Selection*.
- Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40, 437–453.
- Schneider, B., Goldstein, H. W., & Smith, D. B. (1995). The ASA framework: An update. *Personnel Psychology*, 48(4), 747–773.
- Schneider, B., Smith, D. B. & Goldstein, H. W. (2000). Attraction-selection-attrition: Toward a person-environment psychology of organizations. In W. B. Walsh, D. B. Craik, R. H. Price (Eds.) *Person-environment psychology: New directions and perspectives* (2nd ed.) (pp. 61-85).
- Schneiderman, J., & Sachau, D. (2000). The effect of person–organization fit on job satisfaction, performance, and intent to leave. Presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

- Scott, A.J., & Jones, M.A. (2003). Florida law enforcement recruiting and retention series part 10: Traditional recruiting in the technology age. *The Florida Police Chief*, 29, (3), 24-26.
- Scott, M.S. (2000). *Problem-oriented policing: Reflections on the first 20 years*, Washington, DC: US Department of Justice.
- Scott, J. (2010). Survey: The status of field training, *The Police Chief*, 77, 66.
- Selden, S., Schimmoeller, L. & Thompson, R. (2013). The influence of high performance work systems on voluntary turnover of new hires in US state governments, *Personnel Review*, 42(3), 300-323.
- Shantz, A. & Latham, G. P. (2012), Transfer of training: Written self-guidance to increase self-efficacy and interviewing performance of job seekers. *Human Resource Management*, 51, 733-746.
- Shapiro, D., Hom, P., Shen, W., & Agarwal, R. (2016). How do leader departures affect subordinates' organizational attachment? A 360-degree relational perspective. *Academy of Management Review*, 41, 479–502.
- Sharp, S. E. (1899). Individual psychology: A study in psychological method. *The American Journal of Psychology*, 10 (3), 329-391.
- Shaw, J. D. (2011). Turnover rates and organizational performance: Review, critique, and research agenda. *Organizational Psychology Review*, 1, 187–213.

- Shaw, J. D., Duffy, M. K., Johnson, J. L., & Lockhart, D. E. (2005). Turnover, social capital losses, and performance. *Academy of Management Journal*, 48, 594 – 606.
- Singley, M. K., & Anderson, J. R. (1989). *Transfer of cognitive skill*. Cambridge, MA: Harvard University Press.
- Sleezer, C. M., Russ-Eft, D. F. and Gupta, K. (2015). Learning, training, and performance timeline: A walk through history. *Performance Improvement*, 54, 7-15.
- Smith, D. B. (2008). *The people make the place: Exploring dynamic linkages between individuals and organizations*. New York, NY: Psychology Press.
- Smith, S. (2016, June). A crisis facing law enforcement: Recruiting in the 21st century. *The Police Chief*, 83, Retrieved from: <http://www.policechiefmagazine.org/a-crisis-facing-law-enforcement-recruiting-in-the-21st-century/>
- Smith, S. & Aamodt, M. (1997). The relationship between education, experience, and police performance. *Journal of Police and Criminal Psychology*, 12(2), 7-14
- Smith-Jentsch, K. A., Salas, E., & Brannick, M. T. (2001). To transfer or not to transfer? Investigating the combined effects of trainee characteristics, team leader support, and team climate. *Journal of Applied Psychology*, 86, 279-292.

- Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, 46(1), 49- 76.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th Ed.). College Park, MD: SIOP.
- Stajkovic, A. D., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin*, 124(2), 240-261.
- Surrette, M.A., Aamodt, M.G. & Serafino, G. (2004). Using MMPI special scale configurations to predict performance ratings of police officers in New Mexico. *Applied HRM Research*, 9(2), 71-72.
- Tabachnick, B.G., & Fidell, L. S. (2007). *Using multivariate statistics*. (5th ed.). New Jersey: Pearson.
- Tai, W.T. (2006). Effects of training framing, general self-efficacy and training motivation on trainees' training effectiveness. *Personnel Review*, 35(1), 51–65.
- Tannenbaum, S. I., Mathieu, J. E., Salas, E., & Cannon-Bowers, J. A. (1991). Meeting trainees' expectations: The influence of training fulfilment on the development of commitment, self-efficacy, and motivation. *Journal of Applied Psychology*, 76, 759–769.

- Taylor, P. J., Russ-Eft, D. F., & Chan, D. W. L. (2005). A meta-analytic review of behavior modeling training. *Journal of Applied Psychology, 90*, 692-709.
- Tett, R. P., & Meyer, J. P. (1993). Job satisfaction, organizational commitment, turnover intention, and turnover: Path analyses based on metanalytic findings. *Personnel Psychology, 46*, 259–293.
- Tharenou, P., Saks, A. M., & Moore, C. (2007). A review and critique of research on training and organizational-level outcomes. *Human Resource Management Review, 17*, 251-273.
- Thorndike, E. (1914). The measurement of ability in reading: Preliminary scales and tests: Introduction. *The Teachers College Record, 15*(4), 1-2.
- Title 42 U.S.C. Section 1983 (1978). The public health and welfare: Civil rights. United States Code.
- Torres, W.J., & Beir, M.E. (2016). It's time to examine the nomological net of job knowledge. *Industrial and Organizational Psychology: Perspectives on Science and Practice, 1*, 51-55.
- Tyler, M.A. & Mckenzie, W.E. (2014). Training by the dashboard lights: Police training officers' perspectives, *Studies in the Education of Adults, 46*(1), 74-90.
- Tziner, A., Haccoun, R. R., & Kadish, A. (1991). Personal and situational characteristics influencing the effectiveness of transfer of training improvement strategies. *Journal of Occupational Psychology, 64*, 167-177.

UCLA Statistical Consulting Group (n.d.). FAQ: What are the differences between one-tailed and two-tailed tests? Retrieved June 1, 2019, from

<https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-what-are-the-differences-between-one-tailed-and-two-tailed-tests/>

U.S. Department of Justice. (2011, October). The Impact of the economic downturn on American police agencies: A Report of the U.S. Department of Justice Office of Community Oriented Policing Services (the COPS Office).

Retrieved from

http://www.ncdsv.org/images/COPS_ImpactOfTheEconomicDownturnOnAmericanPoliceAgencies_10-2011.pdf

Van Iddekinge, C. H., Roth, P. L., Putka, D. J., & Lanivich, S. E. (2011). Are you interested? A meta-analysis of relations between vocational interests and employee performance and turnover. *Journal of Applied Psychology*, 96(6), 1167.

Vaughn, M. S., Cooper, T. W., & del Carmen, R. V. (2001). Assessing legal liabilities in law enforcement: Police chiefs' views. *Crime and Delinquency*, 47(1), 3-27.

Verquer, M. L., Beehr, T. A., & Wagner, S. H. (2003). A meta-analysis of relations between person–organization fit and work attitudes. *Journal of Vocational Behavior*, 63(3), 473-489.

- Vittinghoff, E., & McCulloch, C.E. (2006). Relaxing the rule of ten events per variables in logistic and Cox regression. *American Journal of Epidemiology*, 165, 710-718.
- Vroom, V. H. (1964). *Work and motivation*. New York, NY: Wiley.
- Walker, J.R. (2005). Law enforcement field training models: Is it time for a change? *Campus Law Enforcement Journal*, 35(5), 23-30.
- Walker, S., & Katz, C. (2002). *The police in America: An introduction*. New York: McGraw-Hill.
- Wallin, J. E. (1911). Experimental studies of rhythm and time. *Psychological Review*, 18(2), 100.
- Warners, R.H. (November 2010). The field training experience: Perspectives of field training officers and trainees. *The Police Chief*, 77, 58-64.
- Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. *Journal of Occupational and Organizational Psychology*, 72, 351-375.
- Wayne, S. J., Shore, L. M., Bommer, W. H., & Tetrick, L. E. (2002). The role of fair treatment and rewards in perceptions of organizational support and leader-member exchange. *Journal of Applied Psychology*, 87, 590-598.
- Wayne, S. J., Shore, L. M., & Linden, R. C. (1997). Perceived organizational support and leader-member exchange: A social exchange perspective. *Academy of Management Journal*, 40, 82-111.

- Weekley, J. A., & Jones, C. (1997). Video-based situational testing. *Personnel Psychology*, 50(1), 25-49.
- Weekley, J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology*, 52, 679–700.
- Weekley, J. A., & Ployhart, R. E. (2005). Situational judgment: Antecedents and relationships with performance. *Human Performance*, 18, 81–104.
- Weekley, J. A., & Ployhart, R. E. (2006). An introduction to situational judgment testing. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* Mahwah, NJ: Lawrence Erlbaum Associates.
- Weekley, J. A., Ployhart, R. E. & Harold, C. M. (2003). *Personality and situational judgment tests across applicant and incumbent settings*. Paper presented at the 19th annual conference of the Society for Industrial and Organizational Psychology, Orlando, Fl.
- Weiss, P. A., & Inwald, R. (2018). A brief history of personality assessment in police psychology: 1916–2008. *Journal of Police and Criminal Psychology*, 33, 1-12.
- Weissbein, D.A., Huang, J.L., Ford, J.K., & Schmidt, A.M. (2011). Influencing learning states to enhance trainee motivation and improve training transfer. *Journal of Business and Psychology*, 26, 423-435.

- Werbel, J.D. and Gilliland, S.W. (1999). Person-environment fit in the selection process, In Ferris, G.R. (Ed.), *Research in Personnel and Human Resource Management* (pp. 209-243, Vol. 17). Stamford, CT: JAI Press.
- Westring, A.J.F., Oswald, F.L., Schmitt, N., Drzakowski, S., & Imus, A., Kim, B., & Shuvpuri, S. (2009). Estimating trait and situational variance in a situational judgment test. *Human Performance*, 22(1), 44–63.
- Wexley, K. N., & Baldwin, T. T. (1986). Posttraining strategies for facilitating positive transfer: An empirical exploration. *Academy of Management Journal*, 29, 503-520.
- Whetzel, D.L., & McDonald, M.A. (2009). Situational judgement tests: An overview of current research. *Human Resource Management Review*, 19, 188-202.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21(3), 291-309.
- Whetzel, D.L., & Reeder, M.C. (2016). Why some situational judgement tests fail to predict job performance (and others succeed). *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 1, 71-77.
- White, M.D. (2008). Identifying good cops early: Predicting recruit performance in the academy. *Police Quarterly*, 1(1), 27-49.

- Williams DM (2010) Outcome expectancy and self-efficacy: Theoretical implications of an unresolved contradiction. *Personality and Social Psychology Review*, 14, 17–425.
- Wilson, J.M. (2014). Strategies for police recruitment: A review of trends, contemporary issues, and existing approaches. *Law Enforcement Executive Forum*, 14(1), 78-97.
- Wilson, J.M. (2012) Articulating the dynamic police staffing challenge: An examination of supply and demand. *Policing: An International Journal of Police Strategies & Management*, 35(2), 327-355.
- Wilson, J. M., & Grammich, C. A. (2009). *Police recruitment and retention in the contemporary urban environment: A national discussion of personnel experiences and promising practices from the front lines*. Santa Monica, CA: RAND Corporation.
- Winters, C.A. (1990). Problems of variance in the utility of the MMPI in the selection of metropolitan police. *Police Journal*, 63(2), 121-128.
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, 19(3), 188-202.
- Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance*, 21(3), 291-309.

- Wolfe, S.E., & Nix, J. (2016). The alleged “Ferguson effect” and police willingness to engage in community partnership. *Law and Human Behavior*, 40(1), 1-10.
- Woska, W.J. (2016, April). Police officer recruitment: A decade later. Retrieved from: <http://www.policechiefmagazine.org/police-officer-recruitment/>
- Woska, W.J. (2006, October). Police officer recruitment: A public-sector crisis. *Police Chief*, 73(10), 52-59.
- Wright, B. E., & Christensen, R. K. (2010). Public service motivation: A test of the job attraction-selection-attrition model. *International Public Management Journal*, 13(2), 155-76.
- Wright, B., Dai, M., & Greenbeck, K. (2011). Correlates of police academy success. *Policing: An International Journal of Police Strategies and Management*, 34(4), 625-637.
- Xiao, J. (1996). The relationship between organizational factors and the transfer of training in the electronics industry in Shenzhen, China. *Human Resource Development Quarterly*, 7, 55-73.
- Ziegler, M., MacCann, C., & Roberts, R. (Eds.). (2011). *New perspectives on faking in personality assessment*. Oxford University Press, USA.
- Zickar, M., Rosse, J., & Levin, R. (1996, April). Modeling faking on a personality test. Paper presented at the 11th annual conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

Zuber, A. (2001). A career in food service cons: High turnover. *Nation's Restaurant News*, 35(21), 147-148.

Appendix

Critical Performance Deficiencies leading to Remedial Training

Performance Domain	Deficiencies
1. Field Performance: Stress (FP Strs 1)	8
2. Field Performance: Non-Stress	7
3. Officer Safety: General	17
4. Officer Safety: Suspects, 13P's, & 10-15s	14
5. Control of Conflict: Verbal	8
6. Control of Conflict: Physical	5
7. Perception/Judgement	15
8. Verbal Communications	5
9. Interpersonal Skills	3
10. Directing Others	8
11. Problem Solving/Decision Making	15
12. Adaptability	6
13. Self-Initiated-Activity	7
14. Driving Skills: Normal Conditions	2
15. Driving Skills: Moderate/High Stress	4
16. City Geography	4
17. Response Time	5
18. Radio: Articulation & Procedures	5
19. Radio: Listens & Comprehends	10
20. Radio: Articulation	4
21. Investigative Skills/Interviewing	10
22. Routine Forms: Accurate/Complete	10
23. Report Writing: Organization/Details	10
24. Report Writing: Grammar/Spelling/Neat	11
25. Report Writing: Appropriate Time Used	13
26. General Appearance	0
27. Department Policies and Procedures: Verbal/Written/Simulated Testing	7
28. Department Policies and Procedures: Reflected Field Performance.	9
29. Criminal Statutes and City Ordinances: Verbal/Written/Simulated Testing	2
30. Criminal Statutes and City Ordinances: Reflected Field Performance.	3
31. Traffic Laws: Verbal/Written/Simulated Testing	0
32. Traffic Laws: Reflected Field Performance.	1
33. Traffic Crash Investigation: Verbal/Written/Simulated Testing	1
34. Traffic Crash Investigation: Reflected Field Performance.	4
35. Arrest Procedures: Verbal/Written/Simulated Testing	1
36. Arrest Procedures: Reflected Field Performance.	2
37. Detention Procedures: Verbal/Written/Simulated Testing	2
38. Detention Procedures: Reflected Field Performance.	2
39. Acceptance of Criticism	4
40. Toward Police Work	1
41. Dependability	3