

Investigating factors related to score change at retest: Examining promotional assessments

by

Sydney L. Reichin

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Masters of Arts in Industrial and Organizational Psychology

Middle Tennessee State University

May 2018

Thesis Committee:

Dr. Mark C. Frame, Chair

Dr. Alexander T. Jackson

Dr. Michael Hein

ACKNOWLEDGEMENTS

First and foremost, I'd like to thank my thesis advisor, Dr. Mark Frame. Without your guidance and support, I do not know how I would have successfully completed the thesis process. You were always available when I really needed you, whether that was to continue progress on my thesis and other research opportunities, discuss the future of my career or just talk about life. While I wish I could take my Grad Dad to Penn State with me in the fall, my future success in that program is largely thanks to your impact on my development as a researcher and a student.

Next, this process would not have been possible without my incredible thesis committee, Dr. Alexander Jackson and Dr. Michael Hein. You both provided great insights and perspectives that really pushed my thesis to the next level. I have learned so much, especially from your combined methodological and statistical knowledge, and am grateful for your mentorship. Similarly, I would also like to thank Tennessee Highway Patrol for allowing me to use data from your promotional processes. Without your cooperation, as well as the work done by many volunteers to score this data, my thesis would not have been possible.

Finally, I would like to thank my family and friends. Without the support of my family to accomplish my goals, I would not have ended up in Tennessee in this great program. Thank you for your continued encouragement on this road to becoming Dr. Sydney Reichin. I would also like to thank my amazing friends. Thank you to my friends from far away for never losing touch while I am off avoiding adulthood. Thank you to my wonderful cohort family for making the move to Tennessee easier than I could have ever imagined. Thank you to the perfect pup, Harper, for always making me smile after a long day. Lastly, thank you Justin for always being there for me and for encouraging me to follow my dreams. I do not know where I would be without you.

ABSTRACT

While research regarding retesting in the selection process is vast, there is limited research regarding retesting in promotion. Additionally, retesting on in-basket simulations and situational judgement tests is also lacking. This study sought to bridge these gaps in the literature, while also attempting to provide an explanation for the majority of findings that scores do increase at retest. Specifically, the current study assessed the effect of anxiety, motivation and information known on subsequent test scores. Data was collected from Tennessee Highway Patrol's (THP) annual promotional process. Participants were approximately 451 THP officers who were eligible for promotion over three years. Results indicated that for those applying to the sergeant position, that anxiety had a stronger effect on those without previous experience than those with previous experience. The other moderating variables showed no moderating effect of previous experience on performance. It is the hope that this study acts as a catalyst for the continued research on the explanation for score increases in the retesting domain.

TABLE OF CONTENTS

LIST OF FIGURES	vi
CHAPTER I: INTRODUCTION AND REVIEW OF THE LITERATURE.....	1
Introduction.....	1
Background.....	1
Review of the Literature	2
Retesting in Personnel Selection.....	2
Scores Increase with Experience	3
Why Do Scores Increase?	5
Assessment Centers	9
Situational Judgement Tests	10
Situational Judgement Tests and Retesting.....	11
In-Baskets	11
In-Baskets and Retesting.....	12
Research Questions and Hypotheses	13
In-Basket Simulation	13
Situational Judgment Test.....	14
CHAPTER II: METHODS	16
Participants.....	16
Measures	16
In-Basket Simulation	16
Situational Judgement Test (SJT).....	18
Information Known.....	19

Motivation.....	20
Anxiety.....	20
Previous Experience.....	21
Procedure	21
2015.....	21
2016.....	21
2017.....	21
CHAPTER III: RESULTS.....	23
Statistical Analyses	23
Test of In-Basket Simulation Hypotheses.....	23
Test of Situational Judgement Test Hypotheses	26
CHAPTER IV: DISCUSSION	31
Practical Implications.....	33
Limitations	34
Future Research	35
Conclusion	36
REFERENCES	37
APPENDICES	43
APPENDIX A: FIGURES	44

LIST OF FIGURES

Figure 1. Year predicting in-basket simulation performance for those applying for Sergeant...44

Figure 2. Year predicting in-basket simulation performance for those applying for Lieutenant..45

Figure 3. Anxiety moderating the effect of previous experience on situational judgement test performance.....46

Figure 4. Year predicting situational judgement test performance for those applying for Sergeant.....47

Figure 5. Year predicting situational judgement test performance for those applying for Lieutenant.....48

CHAPTER I: INTRODUCTION AND REVIEW OF THE LITERATURE

Introduction

Background

Personnel selection is the process used to determine the people who will join an organization. Organizations use a variety of selection processes for a variety of reasons. Some of these selection processes include screening résumés, conducting background checks and administering various selection tests. Selection tests are often used to assess a variety of job relevant constructs such as cognitive ability, personality, physical ability and job knowledge. These tests can be computerized tests, paper-and-pencil tests or work samples (EEOC, 2007). Tests are not only used for hiring job applicants, but also for other employment decisions, such as training and development, compensation and promotion and demotion (EEOC, 1978; U.S. Department of Labor Employment and Training Administration, 1999). Organizations rely heavily on tests to help predict job performance, leadership potential, career success, job satisfaction and tenure (U.S. Department of Labor Employment and Training Administration, 1999).

The use of tests may require organizations to spend considerable time and money when tests and assessment data are used in decision-making. For example, the expense for assessing one employee alone can be extremely expensive. The Centre for Economics and Business Research (CEBR) estimates that for a typical business with 20-49 employees, it costs about \$40,000 to hire one employee (CEBR, 2014). Because of the large expense that often comes with administering these selection tests, organizations need to ensure that these tests are valid. Differentiating between good and bad

assessments is another critical aspect of using data from tests and assessments effectively. Organizations need to ensure that they are assessing applicants with reliable and valid tests, or the organization could be wasting time, money and other resources assessing these applicants with bad instruments. This is also crucial for organizations because if applicants do not perceive the assessments they are taking as valid and reliable, then they may likely view the tests and the entire process as unfair, unethical and immoral (Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). While all of these factors may add to the expense of administering tests, organizations should also be mindful of the expenses associated with administering tests to applicants and candidates more than once. Not only are expenses important to keep in mind, but also the implications of retesting. Because the literature is not complete and/or conflicting, organizations probably do not fully understand the implications of allowing candidates to retest.

Review of the Literature

Retesting in Personnel Selection

Implementing proper retesting policies may require time, money and other resources in addition to the costs associated with initial testing. SIOP's Principles for the Validation and Use of Personnel Selection Procedures (2017) states that "employers should provide opportunities for reassessment and reconsidering candidates whenever technically and administratively feasible." Further, the EEOC's Uniform Guidelines on Employee Selection Procedures (1978) state that "users should provide a reasonable opportunity for retesting and reconsideration. Where examinations are administered periodically with public notice, such reasonable opportunity exists, unless persons who have previously been tested are precluded from retesting." Organizations that allow

retesting spend additional time testing and assessing those applicants that decide to retest. Organizations may allow candidates to retest to increase perceptions of fairness in the application process or retesting may be a means to expand the applicant pool (Brogden, 1949; Gilliland, 1993).

There are many additional reasons why an applicant might repeat the selection process. Applicants whose first test may not have been truly reflective of their ability for reasons outside of their control are good candidates for retesting. These are candidates who may have had the capability to perform better if it were for different circumstances and this score may better predict job performance (U.S. Department of Labor Employment and Training Administration 1999). Some of these reasons include the candidate's state of mind or health at the time of the test, testing conditions and familiarity with questions on the test (U.S. Department of Labor Employment and Training Administration 1999). Further, job applicants who were not selected may also be able to repeat the selection process (Hausknecht, 2010). Because of these possibilities for retesting, organizations have the responsibility to manage this process. One key concern in managing such a process is understanding how test and assessment scores may change over time (Martin, 2014).

Scores Increase with Experience

Research has investigated retesting in a variety of jobs (Kelley, Jacobs, & Farr, 1994; Kolk, Born, & der Flier, 2003) and more specifically in law enforcement (Hausknecht, Trevor, & Farr, 2002; Maurer, Solamon, & Troxtel, 1998), military contexts (Carretta, 1992), for fire fighters (Dunlop, Morrison, & Cordery, 2011) and in school admissions contexts (Lievens, Buyse, & Sackett, 2005; Lievens, Reeve, &

Heggestad, 2007; Puddey, Mercer, Andrich, & Styles, 2014). A variety of tests have been examined in the retest literature, including those that assess cognitive ability (Bartels, Wegrzyn, Wiedl, Ackermann, & Ehrenreich, 2010; Hausknecht, Halpert, Dipaolo, & Gerrard, 2007; Hausknecht et al., 2002; Kulik, Kulik, & Bangert, 1984; Lievens et al., 2005; Lievens et al., 2007), personality (Kelley et al., 1994; Walmsley & Sackett, 2013), and knowledge, skills and abilities (Carretta, 1992; Dunlop et al., 2011; Lievens et al., 2005; Van Iddekinge, Morgeson, Schleicher, & Campion, 2011). Various types of tests have also been assessed, including assessment centers (Brummel, Rupp, & Spain, 2009; Kolk et al., 2003), situational judgement tests (Lievens et al., 2005; Lievens & Sackett, 2007; Maurer et al., 1998) and in-basket exercises (Brannick, Michaels, & Baker, 1989).

The general consensus from the research is that scores tend to increase with experience. This is true for cognitive ability tests (Bartels, et. al., 2010; Hausknecht et al., 2007; Hausknecht et al., 2002; Puddey et. al., 2014) and knowledge tests (Puddey et. al., 2014; Van Iddekinge et al., 2011), among others. Further, it has been found that for personality tests, scores can become more positive (i.e. a participant becoming more cautious) due to previous experience with the test (Hausknecht, 2010; Kelley et al., 1994; Landers, Sackett, & Tuzinski, 2011). This is often deemed as a demonstration of faking on personality tests (Landers et. al., 2011).

This research has been explored with both within person designs (e.g., Dunlop, Morrison & Cordery, 2011) and between person designs (e.g., Dunlop et al., 2011). This is important to distinguish because researchers have explored how those retaking a test may have an advantage over those taking a test for the first time (Dunlop et al., 2011). As previously discussed, mean scores are generally higher at retake versus a participant's

initial attempt. This research is crucial for organizations to understand because while applicants who retest are potentially competing with other retesting applicants, their scores are also being compared to applicants who have never taken the test before. This is a practical implication for organizations, as this may show an unfair advantage to those retaking a test.

Why Do Scores Increase?

The body of research on scores increasing at retest is vast, but why this happens remains unknown. Researchers have postulated a few test-related explanations for why this may happen as suggestions for future research, but still, minimal research has examined these reasons. Some reasons include decreased anxiety (Hausknecht et. al., 2002; Maurer et. al., 1998; Van Iddekinge et. al., 2011), decreased test confusion or increased knowledge about the test format and processes (Hausknecht et. al., 2002; Maurer et. al., 1998; Van Iddekinge et. al., 2011) and increased test motivation (Carretta, 1992; Hausknecht, 2010; Van Iddekinge et. al., 2010). There is a call for research on this topic, because although these explanations may make intuitive sense, they have not been empirically tested (Van Iddekinge & Arnold, 2017).

Test Anxiety. The U.S. Department of Labor Employment and Training Administration (1999) posit that many individuals may feel anxiety about taking a test, which can affect performance on the test. When performance on a test is affected, a true prediction of job performance may not be obtained from the applicant. One of the factors that Randall and Villado (2016) suggest that can cause construct irrelevant score change is individual difference states. They argue that these states may influence a participant's score rather than some other construct relevant reason. Specifically, test anxiety is one

state that may change within a participant. Previous research which has found that as anxiety increases, test performance decreases, supports the assertion that anxiety has a negative relationship with test performance (Arvey, Strickland, Drauden, & Martin, 1990; McCarthy & Goffin, 2005; McCarthy, Van Iddekinge, Lievens, Kung, Sinar & Campion, 2013; Proost, Derous, Schreurs, Hagtvet, & De Witte, 2008). Similarly, anxiety has been shown to be a significant, yet negatively related predictor of test performance (Chenge, Klinger, Fox, Doe, Jin & Wu, 2014). This finding is also true in interviews, such that as anxiety increases, interview performance decreases (McCarthy & Goffin, 2004). More specifically, the effect of test anxiety on test performance is especially present in high stakes settings (Wrightsmann, 1962).

Test Familiarity/Knowledge of Format and Processes. Researchers have also speculated that those who have experience with a particular assessment improve scores the second time because they know what to expect (Hausknecht et. al., 2002; Maurer et. al., 1998; Van Iddekinge et. al., 2011). Not only do test takers know what to expect when it comes to the format of the test or assessment, they may remember or understand what to expect when it comes to the test process as a whole. Previous research supports the notion that when participants know what to expect on a test, they will adjust their study strategies and subsequently perform better on the test (Abd-El-Fattah, 2011). Therefore, it makes sense that Dutke, Barenberg and Leopold (2010) also found that when the test format is known, (i.e. information is known about a test), study strategies are adjusted, which may lead to score changes.

Motivation. Motivation of the applicant or candidate is another factor that has been proposed as a reason that those with experience on a test have higher scores than

first time test takers (Carretta, 1992; Hausknecht, 2010; Van Iddekinge et. al., 2010). Previous research supports that motivation has a positive relationship with test performance, such that as motivation increases, test performance increases (Arvey et al., 1990; Cheng et al., 2014; Wrightsman, 1962). Previous research has also shown that motivation only affected test performance in a condition in which there were no consequences (Sundre, 1999; Sundre & Kitsantas, 2004). This seems counterintuitive, but the authors suggest that this was probably due to low variability in levels of motivation. In high-stakes settings, everyone's motivation is likely to be high resulting in little variability or differences in motivation. However, in low stakes settings, variability in motivation is more probable and thus more likely to be positively associated with test performance.

It is possible that there may be between subject differences in motivation during a retest based on the candidate's scores the first time the assessment was taken. Specifically, research shows that people with lower scores on an initial test are more likely to have a higher magnitude of score increase at retest (Carretta, 1992; Hausknecht, 2010; Puddey et al., 2014). Further, Randall and Villado (2016) outline that this effect may stem from a motivation for candidates to redeem themselves. Because their scores were so low on the first attempt – so low that they could have even failed a test – they may feel motivated to work harder between the two attempts in order to prove that they can do well.

Number of Retest Attempts. The number of times an applicant decides to retest may affect the magnitude of score gains; however, the literature is conflicting. Some research shows that there are score improvements from test one to test two and from test

two to test three (Hausknecht et al., 2002). Conversely, research also provides evidence that while there are score improvements from test one to test two, there are no improvements from test two to test three (Bartels et al., 2010; Dunlop et al., 2011; Kelley et al., 1994; Puddey et al., 2014; Randall & Villado, 2016). All studies used high stakes settings (selection/admission) and all studies used similar constructs (ability/knowledge). While it is difficult to determine the underpinning reasons for the conflicting findings, it could be that there is an upper limit on the improvement one might yield in ability. However, knowledge gain can be built upon more easily and may not have the same upper limit. This idea may be related to the fact that there is less time between tests if a candidate tests three years in a row. Perhaps, for scores to increase, candidates need more time on the job to learn the relevant KSAs and perform better.

Time between Tests. The amount of time between the initial test and retest is another factor that one must take into consideration and again, the literature is conflicting. Previous research provides evidence that a longer amount of time between test attempts results in greater score improvements (Van Iddekinge et al., 2011). However, other previous research provides evidence that less time between initial test and retake results in greater score improvements (Bartels et. al., 2010; Dunlop et al., 2011; Hausknecht et al., 2007; Randall & Villado, 2016).

Research that showed a longer amount of time between test attempts resulted in greater score improvement used a job knowledge test to assess participants (Van Iddekinge et al., 2011). Research that showed a shorter amount of time between test attempts resulted in greater score improvement used some aspect of cognitive ability to assess participants (Bartels et al., 2010; Dunlop et al., 2011; Hausknecht et al., 2007).

Therefore, for job knowledge tests, longer time between tests results in greater score improvement, but cognitive ability tests are more stable, so there is more improvement with smaller amounts of time between tests.

Randall & Villado (2016) note that score gains in cognitive ability tests may be due to construct-irrelevant factors. In this case, it is possible that with shorter amounts of time between tests, participants are more able to recall information about their initial assessment and improve scores because of memory. Construct-relevant score change, however, is true improvement in the relevant construct that could happen through formal training, self-study, practice, or coaching (Randall & Villado, 2016). In the case of a job knowledge test, longer periods allow participants to actually become better on the job and score gains may be due to true ability change (Randall & Villado, 2016).

Assessment Centers

As previously noted, there are many types of personnel assessments. An assessment center is a method of measurement that uses an array of activities or simulations in an effort to predict how candidates may behave in a particular setting (often in a supervisory or managerial setting; Brannick et. al., 1989). These activities may include various low and high fidelity behavioral simulations in which candidates act in a way similar to if they were on the job (Brummel et al., 2009). Assessment Centers are most commonly used to predict job performance in higher-level positions because of the candidates required use of problem-solving and decision-making skills (U.S. Department of Labor Employment and Training Administration, 1999). Situational judgement Tests (SJTs) and in-basket exercises are types of tests frequently used as part of an assessment center (Guidelines and Ethical Considerations for Assessment Center Operations, 2015).

Situational Judgement Tests

Situational judgement tests (SJTs) attempt to measure how a candidate would respond to a scenario in a work setting (McDaniel & Nguyen, 2001). SJTs are developed most commonly by using information from critical incidents interviews to create situations or scenarios and potential response options to those situations (Campion, Ployhart, & MacKenzie, 2014). These critical incident interviews may be provided by supervisors with no guidance on content, but just simply by remembering important events that have happened to them. Conversely, the subject matter experts (SMEs) may be given certain competencies derived from a job analysis in which to frame the critical incidents around (McDaniel & Nguyen, 2001). Developing these critical incidents is the first step in developing SJTs, the next step is to shorten these incidents into stems that can be used as part of the test and then develop possible responses to each of the incidents. SMEs then review these stems and response options (McDaniel & Nguyen, 2001). The types of response options can vary. Candidates may respond by indicating which response they would be most likely to perform, least likely to perform, or both. They might indicate which response option is the best, or worst, or both. They may also be asked to rate the effectiveness of each response (McDaniel & Nguyen, 2001). SJTs may be used for selection or for training and development and can measure as few as two dimensions or as many as 12 dimensions (Campion et al., 2014). Some of these constructs may include job knowledge, job skills, teamwork, leadership, interpersonal skills, conscientiousness and cognitive ability (Campion et al., 2014; McDaniel & Nguyen, 2001).

Situational Judgement Tests and Retesting

In an attempt to address the issue of score changes in testing and retesting, Lievens and Sackett (2007) examined whether it was possible to create alternate forms of SJTs. As part of this research, they tested the various methods of creating alternate forms and examined retest effects depending on which approach they took. They found that retest effects were larger when only some words were changed and the same situations were used. More modest gains resulted when the situations were completely changed between test administrations (Lievens & Sackett, 2007). Further, Lievens and colleagues (2005) examined retest effects with video SJTs in a high stakes (admissions) setting. Within-person analyses of the results found that participants improved by one-third of a standard deviation on the second attempt at a parallel form of the same SJT.

While these two studies contribute tremendously to the retesting research, they still leave gaps in the literature. Both studies were evaluated in a high stakes admissions context. Further, the first study aimed at creating parallel forms and the main purpose was not to examine retest effects, but rather to use findings to help support validity of the parallel forms.

In-Baskets

A previously mentioned simulation that can also be used as part of an assessment center is an in-basket exercise. The in-basket exercise presents applicants and candidates with items and communications that might be found in the in-box or in-basket of someone in a given role, often a manager or supervisor (Brannick et al., 1989). The in-basket may include items that require the candidate to respond in some way, such as phone messages, complaints from customers, schedule conflicts and emails from clients

and coworkers (Brannick et al., 1989; U.S. Department of Labor Employment and Training Administration, 1999). Some constructs commonly examined through in-basket exercises include organization and planning, perceptiveness, delegation, leadership and decision making (Brannick et al., 1989).

In-Baskets and Retesting

As previously mentioned, there have been a variety of types of tests examined in the retest literature. However, very little research has examined in-baskets and retesting. One study found that implementing training specifically designed to increase in-basket scores resulted in significantly higher in-basket scores at retest (Brannick et. al., 1989). However, this research was purposefully trying to increase in-basket scores on retest. In the selection process, it is unlikely that employees will have access to this type of training before retaking an in-basket assessment. Further, this may not apply to those who retest for reasons outside of their control (i.e. technological malfunctions or family emergencies) versus those who retest after an allotted period of time (i.e. allowing employees to retest once a year).

The Current Study

The current study will examine how previous experience with in-basket exercises and situational judgement tests affect test scores. These two types of assessments are being examined specifically because there is a lack of research about promotional decisions in the retest domain and a lack of retest research on these two methods. Further, the current study is specifically interested in exploring the extent to which prior experience with a promotional test may result in lower levels of state anxiety, resulting in higher test scores. Similarly, it could be expected that officers who have previous

experience with the promotional process will prepare differently than those who have not seen the test before and therefore will perform better compared to first time test takers. Additionally, it is the hope that motivation might help explain score changes depending on how much previous experience candidates have with the promotional tests. Finally, it is of interest to researchers to examine how time between tests and number of retest attempts affects score improvement.

Research Questions and Hypotheses

In-Basket Simulation

Hypothesis 1: Participants with previous experience with an in-basket at Year 1 will have an increase in in-basket scores from Year 1 to Year 2.

Hypothesis 2a: Anxiety will moderate the relationship between having previous experience with an in-basket and increased in-basket scores, such that as anxiety decreases, in-basket scores increase.

Hypothesis 2b: Given the ambiguous nature of in-basket simulations, participants with previous experience with an in-basket will have lower levels of anxiety than those participants without previous experience with an in-basket.

Hypothesis 3a: Information known will moderate the relationship between having previous experience with an in-basket and increased in-basket scores, such that as information known increases, in-basket scores increase.

Hypothesis 3b: Participants with previous experience with an in-basket will have higher levels of information known than those participants without previous experience with an in-basket.

Hypothesis 4a: Test motivation will moderate the relationship between having previous experience with an in-basket and increased in-basket scores, such that as test motivation increases, in-basket scores increase.

Hypothesis 4b: Participants with previous experience with an in-basket will have higher levels of motivation than those participants without previous experience with an in-basket.

Research Question 1: Will third-time (Year 3) scores on an in-basket be significantly different from second-time (Year 2) scores on an in-basket for those that completed the in-basket three times?

Research Question 2: Will longer time intervals between in-basket attempts (between Year 1 and Year 3) show greater differences in scores than shorter time intervals between in-basket attempts (between Year 1 and Year 2 and between Year 2 and Year 3) for those that completed the in-basket at least two times?

Situational Judgment Test

Hypothesis 7: Participants with previous experience with an SJT at Year 1 will have an increase in SJT scores from Year 1 to Year 2.

Hypothesis 8a: Anxiety will moderate the relationship between having previous experience with an SJT and increased SJT scores, such that as anxiety decreases, SJT scores increase.

Hypothesis 8b: Participants with previous experience with an SJT will have lower levels of anxiety than those participants without previous experience with SJT.

Hypothesis 9a: Information known moderates the relationship between having previous experience with an SJT and increased SJT scores, such that as information known increases, SJT scores increase.

Hypothesis 9b: Participants with previous experience with an SJT will have higher levels of information known than those participants without previous experience with an SJT.

Hypothesis 10a: Test motivation moderates the relationship between having previous experience with an SJT and increased SJT scores, such that as test motivation increases, SJT scores increase.

Hypothesis 10b: Participants with previous experience with an SJT will have higher levels of motivation than those participants without previous experience with an SJT.

Research Question 3: Will third-time (Year 3) scores on an SJT be significantly different from second-time (Year 2) scores on an SJT for those that completed the SJT three times?

Research Question 4: Will longer time intervals between SJT attempts (between Year 1 and Year 3) show greater differences in scores than shorter time intervals between SJT attempts (between Year 1 and Year 2 and between Year 2 and Year 3) for those that completed the SJT at least two times?

CHAPTER II: METHODS

Participants

Participants were Tennessee Highway Patrol (THP) officers who were eligible for promotion. Data was collected over three years (2015, 2016, 2017). From a total of 451 candidates over three years, there were 323 current troopers applying for the Sergeant position and 126 current sergeants applying for the lieutenant position. A total of 284 candidates completed the promotion process in 2015, 155 completed the process in 2016 and 232 completed the process in 2017. Of the 451 candidates, 246 candidates were taking these tests for the first time, 184 were taking these tests for the second time, and 19 were taking these tests for the third time. There was a total of 353 men, 12 women, and 86 did not disclose their gender. Ages ranged from 24 to 59 ($M = 41.58$, $SD = 7.12$). Finally, about 71% of candidates were White, 6% were Black, about 1% made up other races and 20% did not disclose.

Measures

In-Basket Simulation

In-Basket Task. The in-basket was used to assess how well candidates solve problems and handle various situations while on the job. The in-basket work sample items were developed using critical incidents interviews. The consultants who collected the interviews served as our SMEs. A fictional district was created within the highway patrol using a command structure similar to that of THP for the purposes of the in-basket. Materials given to the candidates included in-basket instructions, an organizational chart, a calendar of important dates, in-basket response instructions and in-basket response sheets. Various response forms that candidates could use include a Complaint Against

Department Member, Highway Patrol Pursuit Report, Department of Human Resources Request for Mediation, Tennessee Department of Labor and Workforce Development Employer's First Report of Work Injury or Illness, Occupational Safety and Health Administration Form 301 Injuries and Illnesses Incident Report and the Tennessee Department of Safety and Homeland Security Use of Force Report. The final in-basket simulation for potential sergeants consisted of 33 items (26 emails, 5 voicemails and 2 radio messages). The final in-basket simulation for potential lieutenants consisted of 32 items (23 emails, 5 voicemails, 2 radio messages and 1 memo). Participants were given three hours to respond.

Scoring Process. The in-basket simulation requires open ended responses from candidates. The qualitative data provided by the candidates was coded and scored by a new team of researchers. A five-point behaviorally anchored rating scale (BARS) was created to assess the ability of candidates to handle situations and problems that might occur on the job. Rating options ranged from one to five (1 = *very ineffective*, 5 = *very effective*). At least two raters were assigned to score each in-basket simulation. Rater's scores were evaluated, and if they were not within one point of the other rater's scores, they were required to come to consensus. When consensus was required, the two raters met, discussed their scores and agreed on one score for the exercise.

Frame of Reference Training. Those chosen to score in-baskets were trained to use the BARS using frame of reference (FoR) training. First established by Bernardin and Buckley (1981), FoR training is meant to give raters a common standard to compare performance and has been shown to increase the validity of assessment center ratings (Schleicher, Mayes, Day & Riggio, 2002). A sample of in-baskets was de-identified and

used to conduct FoR training. After training, the sample in-baskets were included in the in-baskets to be scored and were rescored. Scores provided (under blind rating conditions) on the sample in-baskets were similar to those provided during the training. Raters scored each in-basket individually and then came together to discuss ratings.

Differences by Year. This is a yearly promotional process. There are some differences worth noting by year. There is a location shift from Year 1 to Year 2 and Year 3. In Year 1, the promotional process was conducted at the THP training center, but in Year 2 and Year 3, the promotional process was conducted at Middle Tennessee State University. Finally, the administration method changed by year. In Year 1 and Year 2 the candidates responded via paper and pencil. In Year 3, candidates responded via a computer using Qualtrics Survey Software.

Situational Judgement Test (SJT)

SJT Task. This test was used to assess candidates' preferred course of action in specified situations. One-hundred and fifty situations were generated and brought to the THP command staff for evaluation. These situations represent actual scenarios that the candidates would likely encounter on the job and were developed using critical incidents of performance interviews. A team of two project associate consultants conducted each interview. One interviewer was in charge of asking questions and taking brief notes and the other interview was in charge of taking detailed notes. The command staff served as our SMEs. The SMEs were able to get rid of items and they were asked to rank the response options given after each situation. Four response options were given and SMEs were asked to rank the various courses of action (1 = *best option*, 4 = *worst option*). They were also able to alter some response options. If command staff reached a threshold of

agreement, items were included. The final SJT for both potential sergeants and lieutenants consisted of 100 items (each with four response options). The content of the two exams was different for sergeant and lieutenant candidates. Participants were given three and a half hours to respond.

Scoring Process. Candidates were asked to rank order each of the four response options based on how likely they would be to perform each course of action. This method was used because four data points could be collected from candidates. SJTs were assessed based on absolute difference scores. If candidates rank ordered the four response options in the same order as the SMEs, they earned a score of zero. If candidates ordered the four response options the exact opposite as the SMEs, they scored an eight. The difference scores from each situation are summed to get a final difference score. The final score was subtracted from 50 to get an actual score that was more understandable to the candidate, since there is a tendency for individuals to associate higher numbers with better performance. Final scores ranged from 0 to 50.

Differences by Year. The same differences by year for the in-basket exercise apply to the SJTs (location and administration method).

Information Known

In order to measure the perceived information known about each test, the information known subscale on the selection procedural justice scale was used. The selection procedural justice scale is a ten-dimension inventory measuring subject's procedural justice perceptions to selection and was developed by Bauer, Truxillo, Sanchez, Craig, Ferrara and Campion (2001). The current study used the three-item information known dimension of this scale (e.g., "I knew what to expect on the test").

Participants rated items using a 5-point Likert scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*).

Motivation

In order to measure the test motivation of each promotional candidate, the motivation subscale of the test attitude survey was used. The motivation subscale is a nine-dimension inventory measuring subject's attitudes about taking a test and was developed by Arvey and colleagues (1990). The current study used the ten-item motivation dimension of this scale (i.e., "Doing well on this test is important to me").

Participants rated items using a 5-point Likert scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*).

Anxiety

In order to measure the state anxiety of each promotional candidate, the six-item short form of the state-trait anxiety inventory was used. The original two-dimension, 40-item inventory was developed by Spielberger, Gorsuch, & Lushene (1970) to assess both state (able to fluctuate) and trait (more stable) anxiety of subjects. The current study used a six-item short form of the state anxiety dimension of the original STAI that was developed by Marteau & Bekker (1992). Similar to Jackson, Howes, Kausel and Young (2017) and Jackson, Howes, Kausel, Young and Loftis (under review), additional items relating to feeling happy were added to avoid priming participants on anxiety. The final scale consisted of 12 items. Participants rated items using a 5-point sliding scale (1 = *Not at all like me*, 5 = *Very much like me*).

Previous Experience

Previous experience of each candidate was collected. Candidates whose only test score was from 2017 were coded as not having previous experience with the promotional process. However, candidate's that had a test score from 2017 and at least one other year (either 2015 or 2016) were coded as having previous experience with the promotional process.

Procedure

2015

Candidates completed the promotional testing process using paper and pencil at the highway patrol agency training facility. Candidates were prompted to take the in-basket simulation and given three hours. Candidates then took a lunch break. Following lunch, candidates were prompted to take the SJT and were given three and a half hours to complete the assessment.

2016

Candidates completed the promotional testing process using paper and pencil at Middle Tennessee State University. Candidates were prompted to take the in-basket simulation and given three hours. Following the in-basket simulation, candidates took a break for lunch. When they returned, they were prompted to take the SJT and were given three and a half hours.

2017

Candidates completed the promotional testing process using Qualtrics Survey Software at Middle Tennessee State University. Candidates opened their surveys to a welcome page. When they were instructed to do so, they continued to the SPJS-IKS,

TAS-M and the STAI-6. Candidates were then prompted to take the in-basket simulation and were given three hours. Following the in-basket simulation, candidates repeated the SPJS-IKS, TAS-M and STAI-6. Candidates then took a break for lunch. When they returned, they were prompted with a welcome back screen. When they were instructed to do so, they continued to the SPJS-IKS, TAS-M and the STAI-6. Participants were then prompted to take the SJT and were given three and a half hours. The SJT was again followed by the SPJS-IKS, TAS-M and the STAI-6.

CHAPTER III: RESULTS

Statistical Analyses

Test of In-Basket Simulation Hypotheses

Hypothesis 1. A repeated measures regression was conducted to determine whether previous experience with an in-basket predicted increased performance on a subsequent in-basket. Previous experience on an in-basket was not a significant predictor of subsequent in-basket performance for those applying to be sergeants ($\beta = 0.66$, $SE = 0.40$, $t = 1.66$, $p > .05$) or lieutenants ($\beta = -0.47$, $SE = 0.98$, $t = -0.48$, $p > .05$). Therefore, hypothesis 1a was not supported.

Hypothesis 2a. Regression analyses were conducted to determine whether anxiety moderated the effect of previous experience on in-basket performance. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. For those applying to the sergeant position, the main effects of previous experience ($t(165) = -0.42$, $p = .68$) and anxiety ($t(165) = 1.11$, $p = .27$) were not significant. Further, the interaction was not significant ($t(165) = -1.04$, $p = .30$). For those applying to the lieutenant position, the main effect of previous experience was not significant ($t(59) = 1.38$, $p = .17$), but the main effect of anxiety on performance was significant ($t(59) = 2.44$, $p = .02$). However, the interaction was not significant ($t(59) = -1.71$, $p = .09$). Therefore, hypothesis 2a was not supported.

Hypothesis 2b. A univariate ANOVA was conducted to determine whether participants with previous experience with an in-basket will have lower levels of anxiety than those without previous experience. Results showed no significant differences in

levels of anxiety for those with previous experience ($M = 10.32$, $SD = 5.14$) and without previous experience ($M = 10.83$, $SD = 4.40$) for those applying for sergeant ($F(1, 167) = 0.41$, $p = .52$). Further, the results showed no significant differences in levels of anxiety for those with previous experience ($M = 11.48$, $SD = 5.53$) and without previous experience ($M = 11.87$, $SD = 5.68$) for those applying for lieutenant ($F(1, 61) = 0.07$, $p = .79$). Therefore, hypothesis 2b was not supported.

Hypothesis 3a. Regression analyses were conducted to determine whether information known moderated the effect of previous experience on in-basket performance. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. For those applying to the sergeant position, the main effects of previous experience ($t(164) = -1.59$, $p = .12$) and information known ($t(164) = -0.97$, $p = .33$) were not significant. Further, the interaction ($t(164) = 1.47$, $p = .14$) was not significant. For those applying to the lieutenant position, the main effects of previous experience ($t(59) = 0.79$, $p = .43$) and information known ($t(59) = 0.74$, $p = .46$) were not significant. Further, the interaction ($t(59) = -0.66$, $p = .51$) was not significant. Therefore, hypothesis 3a was not supported.

Hypothesis 3b. A univariate ANOVA was conducted to determine whether participants with previous experience with an in-basket will have higher levels of information known than those without previous experience. Results showed, for those applying to the sergeant position, that those with previous experience had greater levels of information known ($M = 4.20$, $SD = 0.56$) than those without previous experience ($M = 3.88$, $SD = 0.84$), $F(1, 166) = 8.38$, $p < .01$. There were no differences in levels of

information known for those with previous experience ($M = 4.18$, $SD = 0.74$) and those without previous experience ($M = 4.07$, $SD = 0.59$) applying for lieutenant ($F(1, 61) = 0.38$, $p = .54$). Therefore, hypothesis 3b was partially supported.

Hypothesis 4a: Regression analyses were conducted to determine whether test motivation moderated the effect of previous experience on in-basket performance. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. For those applying to the sergeant position, while the main effect of test motivation was not significant ($t(164) = 0.68$, $p = .50$) the main effect of previous experience on in-basket performance was significant ($t(164) = -2.26$, $p = .03$). Further, the interaction of previous experience and test motivation was significant ($t(164) = 2.13$, $p = .04$). Because of the significant interaction, simple effects were tested. The test of simple effects revealed that there was no effect of motivation on in-basket performance for those with previous experience ($t(3,164) = -0.07$, $p = .94$) or without previous experience ($t(3,164) = -1.18$, $p = .24$). For those applying to the lieutenant position, the main effects of previous experience ($t(59) = 0.02$, $p = .99$) and test motivation ($t(59) = 0.34$, $p = .74$) were not significant. Further, the interaction was not significant ($t(59) = 0.46$, $p = .65$). Therefore, hypothesis 4a was not supported.

Hypothesis 4b. A univariate ANOVA was conducted to determine whether participants with previous experience with an in-basket would have higher levels of test motivation than those without previous experience. Results showed no significant differences in test motivation for those with previous experience ($M = 4.64$, $SD = 0.58$) and without previous experience ($M = 4.63$, $SD = 0.44$) for those applying for sergeant

($F(1, 167) = 0.03, p = .87$). Further, the results showed no significant differences in test motivation for those with previous experience ($M = 4.52, SD = 0.67$) and without previous experience ($M = 4.81, SD = 0.48$) for those applying for lieutenant ($F(1, 61) = 3.27, p = .08$). Therefore, hypothesis 4b was not supported.

Research Question 1. A repeated measures regression was conducted to determine how in-basket performance changed over time. Results showed that in-basket scores increased over three years for both sergeants ($\beta = 2.78, SE = 0.08, t = 33.88, p < .05$) and lieutenants ($\beta = 7.31, SE = 0.21, t = 34.54, p < .05$). See Figure 1 and Figure 2.

Research Question 2. A univariate ANOVA was conducted to determine whether longer time between in-baskets or less time between in-baskets accounted for greater score change. No significant differences were found in mean score change with one year between tests ($M = 6.34, SD = 7.92$) and two years between tests ($M = 6.65, SD = 6.42, F(1, 182) = .09, p = .77$).

Test of Situational Judgement Test Hypotheses

Hypothesis 7. A repeated measures regression was conducted to determine whether previous experience with an SJT predicted increased performance on a subsequent SJT. Previous experience on an SJT was not a significant predictor of subsequent in-basket performance for those applying to be sergeants ($\beta = 0.14, SE = 0.47, t = -0.30, p > .05$) or lieutenants ($\beta = -0.62, SE = 0.85, t = -0.73, p > .05$).

Hypothesis 8a. Regression analyses were conducted to determine whether anxiety moderated the effect of previous experience on SJT performance. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. For those applying to the

sergeant position, while the main effect of previous experience was not significant ($t(165) = -0.11, p = .91$) the main effect of anxiety on SJT performance was significant ($t(165) = 2.87, p < .01$). Further, the interaction of previous experience and anxiety was significant ($t(165) = -2.363, p = .02$). Because of the significant interaction, simple effects were tested. Results indicated that the effect of anxiety on SJT performance was stronger for those without previous experience ($\beta = 1.28, t(3, 165) = 2.71, p = .01$) than for those with previous experience ($\beta = 0.47, t(3, 165) = 3.24, p = .001$). However, this is not the expected relationship as there is a positive relationship between anxiety and performance. See Figure 3. For those applying to the lieutenant position, the main effects of previous experience ($t(59) = -0.22, p = .83$) and anxiety ($t(59) = -0.02, p = .99$) were not significant. Further, the interaction ($t(59) = 0.10, p = .92$) was not significant. Therefore, hypothesis 8a was not supported.

Hypothesis 8b. A univariate ANOVA was conducted to determine whether participants with previous experience with an SJT would have lower levels of anxiety than those without previous experience. Results showed no significant differences in levels of anxiety for those with previous experience ($M = 9.79, SD = 5.40$) and without previous experience ($M = 9.00, SD = 4.72$) for those applying for sergeant ($F(1, 167) = 0.85, p = .36$). Further, the results showed no significant differences in levels of anxiety for those with previous experience ($M = 10.98, SD = 5.85$) and without previous experience ($M = 11.50, SD = 5.07$) for those applying for lieutenant ($F(1, 61) = 0.13, p = .72$). Therefore, hypothesis 8b was not supported.

Hypothesis 9a. Regression analyses were conducted to determine whether information known moderated the effect of previous experience on SJT performance. In

order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. For those applying to the sergeant position, the main effects of previous experience ($t(164) = 1.19$, $p = .24$) and information known ($t(164) = 1.33$, $p = .19$) were not significant. Further, the interaction of previous experience and information known was not significant ($t(164) = -1.13$, $p = .26$). For those applying to the lieutenant position, the main effects of previous experience ($t(59) = 0.12$, $p = .91$) and information known ($t(59) = -0.09$, $p = .93$) were not significant. Further, the interaction ($t(59) = -0.14$, $p = .89$) was not significant. Therefore, hypothesis 9a was not supported.

Hypothesis 9b. A univariate ANOVA was conducted to determine whether participants with previous experience with an SJT will have higher levels of information known than those without previous experience. Results showed no significant differences in information known for those with previous experience ($M = 4.06$, $SD = 0.69$) and without previous experience ($M = 4.03$, $SD = 0.63$) for those applying for sergeant ($F(1, 167) = 0.92$, $p = .76$). Further, the results showed no significant differences in levels of anxiety for those with previous experience ($M = 4.31$, $SD = 0.60$) and without previous experience ($M = 4.13$, $SD = 0.86$) for those applying for lieutenant ($F(1, 61) = 0.91$, $p = .34$). Therefore, hypothesis 9b was not supported.

Hypothesis 10a. Regression analyses were conducted to determine whether test motivation moderated the effect of previous experience on SJT performance. In order to reduce the effects of multicollinearity between the main effects and the interactions, the predictors were centered before being entered into the model. For those applying to the sergeant position, the main effects of previous experience ($t(164) = 1.41$, $p = .16$) test

motivation ($t(164) = 1.77, p = .08$) were not significant. Further, the interaction of previous experience and test motivation was not significant ($t(164) = -1.38, p = .17$). For those applying to the lieutenant position, the main effects of previous experience ($t(59) = 0.64, p = .53$) and test motivation ($t(59) = -0.88, p = .39$) were not significant. Further, the interaction ($t(59) = -0.63, p = .53$) was not significant. Therefore, hypothesis 10a was not supported.

Hypothesis 10b. A univariate ANOVA was conducted to determine whether participants with previous experience with an SJT will have higher levels of test motivation than those without previous experience. There were no differences in levels of test motivation for those with previous experience ($M = 4.63, SD = 0.51$) and without previous experience ($M = 4.58, SD = 0.49$) for those applying for the sergeant position ($F(1, 166) = 0.42, p = .52$). However, for those applying to the lieutenant position, those without previous experience had greater levels of test motivation ($M = 4.85, SD = 0.30$) than those with previous experience ($M = 4.57, SD = 0.49$), $F(1, 61) = 6.01, p = .02$. This is not in the intended direction. Therefore, hypothesis 10b was not supported.

Research Question 3. A repeated measures regression was conducted to determine how SJT performance was predicted over time. Results showed that SJT scores were predicted to decrease over three years for both sergeants ($\beta = -1.00, SE = 0.11, t = -9.10, p < .05$) and lieutenants ($\beta = -5.27, SE = 0.15, t = 35.45, p < .05$). See Figure 4 and Figure 5.

Research Question 4. A univariate ANOVA was conducted to determine whether longer time between SJT or less time between SJT accounted for greater score change. Results showed that those with two year between tests had greater mean score change ($M = -5.17, SD = 5.66$) than those with one year between tests ($M = 0.18, SD = 8.83$), $F(1, 82)$

= 24.72, $p < .001$.

CHAPTER IV: DISCUSSION

The goal of the current study was to evaluate retesting in a promotional process while exploring tests that have been under-researched in the retesting domain. The current research was aimed at identifying the underlying factors accounting for score change at retest. In particular, level of anxiety, level of motivation and amount of information known were predicted to account for these changes. Specifically, most literature has found that scores increase with previous experience on a test (Bartels, et. al., 2010; Hausknecht et al., 2007; Hausknecht et al., 2002; Puddey et. al., 2014; Van Iddekinge et al., 2011). However, the current study did not find that previous experience on the in-basket or SJT predicted better performance on these tests. These results conflict with the vast majority of literature. One reason for these conflicting findings may be that, when looking at actual scores, in-basket scores were much lower than SJT scores in 2015. Therefore, it is possible that retesters focused their attention on improving in-basket performance and not SJT performance. This may have, in turn, diverged from the normal retesting relationships.

While the relationship between previous experience and test performance was not as expected, the predicted moderating effect of anxiety, motivation and information known was still assessed. These interactions were still evaluated because the relevant data for the moderation hypotheses only involved the data collected from 2017, while the previous hypothesis included data from all three years. Analyses were parsed out into data from in-basket performance and data from SJT performance.

First, anxiety did not moderate the relationship between previous experience and in-basket performance. This is likely because those with previous experience and first

time test takers did not differ on level of anxiety. Further, for those applying to the sergeant position, those with previous experience did know significantly more information about the test than those without previous experience. However, this also did not result in a moderation between previous experience and in-basket performance. Finally, for those applying to the sergeant position, test motivation did not moderate the relationship between having previous experience and test scores, indicating motivation has the same effect on those with previous experience than those without previous experience.

Alternatively, for SJT performance, moderators emerged in a different way. Anxiety was found to moderate the relationship between previous experience and SJT performance, such that anxiety had a larger effect on those without previous experience and specifically for those applying to the sergeant position. This is interesting considering levels of anxiety did not differ for those with and without previous experience. However, it does make sense that anxiety would still affect those without previous experience more because perhaps they cannot cope as well with anxiety as easily as those with previous experience on an SJT. Interestingly, there is a positive relationship between anxiety and SJT performance for those without previous experience. In the current study, it seems as though anxiety actually predicts higher performance on the SJT. When evaluating scores on anxiety, they are actually quite low. So, while anxiety is higher for those without previous experience, levels of anxiety are not actually high. This may be attributed to the inverted-U theory, which states that some arousal is actually beneficial for performance (Martens & Landers, 1970).

Conversely, information known and motivation did not moderate this relationship. Interestingly, for those applying to the lieutenant position, those without previous experience were found to know more information regarding the SJT than those with previous experience. Perhaps this relationship was blurry because all candidates are currently sergeants (higher ranking) at the agency, so they may know more about the promotional process even when they have not participated in this process before.

The first research question was aimed at determining whether scores would not only increase from year one to year two, but also year two to year three. Results indicated that scores increased across all three years for in-basket performance, but decreased over three years for SJT performance. This may be due to the finding that in-basket scores were so low at year one, so officers may have compensated by solely focusing on in-basket performance and not SJT performance in an attempt to raise their previous in-basket score.

The second research question was aimed at determining whether more time or less time between tests accounted for greater score change on retest. For in-baskets, amount of time between tests did not make a difference on magnitude of score change. For SJTs, however, more time between tests accounted for greater score change, although this change was a decrease after two years. These results can likely be accounted for again by the difference in in-basket and SJT scores at time one.

Practical Implications

While the findings of the current study were unexpected, they still have important implications. Because previous experience on an in-basket and an SJT did not predict subsequent in-basket and SJT performance, organizations may not need to worry about

those who retest having an advantage over those who are testing for the first time. Further, it may be of importance to organizations what the underlying reasons are for score change at retest. Instead of making the assumption that those who retest are learning how to take the test smarter or simply memorizing material, differences in anxiety may actually account for these changes.

Additionally, this research helps bridge some gaps in the literature. Very little research has explored in-basket simulations and situational judgement tests in the retest domain. Further, very little research explores retesting in a promotional context. Perhaps the current study can be a catalyst for continued research on these unexplored aspects of retesting research. Further, the current study attempted to reconcile conflicting findings in the literature. Specifically, while there was conflicting research regarding score changes at time three, the current study did find that scores can increase across all three years, specifically for an in-basket simulation. Moreover, there is conflicting research regarding time between tests accounting for more or less score change at retest. The current research, however, may help to close this gap as evidence points towards more time between tests accounting for more score change. This may be due to construct relevant score change. An in-basket and SJT are a form of a job knowledge test (Brannick et. al., 1989; McDaniel & Nguyen, 2001) and therefore, more time between tests would provide employees more time to learn relevant KSAs that would allow them to perform better on a subsequent test.

Limitations

There are some limitations to the current study. While having the ability to obtain data from a real high-stakes promotional setting was invaluable, there were also factors

that could not be controlled since data was collected as part of normal human resources functions. For example, there was no control over testing conditions. Therefore, the conditions that candidates tested in may have differed slightly. Additionally, while there are three years of data collection, which allows for an evaluation of scores over three years, not every candidate actually tested all three times. Over time, it would be possible to collect more data on each individual so that each candidate has three years of data and a within person design can be used with a sample of more than 19 candidates. Similarly, the anxiety, motivation and information known measures were only collected for those who tested in 2017. Therefore, while we can compare first time test takers versus retesters, it was not possible to compare within person a change in any of these individual factors.

Future Research

Future research will be collected in subsequent years, so within person changes in anxiety, motivation, and information known can be assessed. Further, future research should continue to evaluate retesting in the promotional process because the literature is crowded by retesting in initial selection. Similarly, research should continue to assess in-baskets and SJTs as the literature on these tests is lacking. Additionally, new selection methods should continue to be researched within the retest literature. For instance, it may be of interest to examine the effect of re-interviewing for the same organization. Finally, researchers should continue to investigate various reasons for score change at retest. For example, perhaps there are non-testing related issues, such as maturation, cognitive development, life events, that may account for these changes.

Conclusion

In conclusion, the current study sought to explore retesting in a promotional context, particularly for in-basket simulations and situational judgement tests. Specifically, the study was interested in uncovering the underpinning reasons for score change at retest and proposed anxiety, test motivation and information known of the candidate as possible explanations. Although the results are limited, it is the hope that this study will act as a stepping stone for future research on this topic.

REFERENCES

- Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology, 43*(4), 695-716.
- Bartels, C., Wegrzyn, M., Wiedl, A., Ackermann, V., & Ehrenreich, H. (2010). Practice effects in healthy adults: A longitudinal study on frequent repetitive cognitive testing. *BMC Neuroscience, 11*(1), 118-130. doi:10.1186/1471-2202-11-118
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54*(2), 388-420. doi:10.1111/j.1744-6570.2001.tb00097.x
- Bernardin, H., & Buckley, M. (1981). Strategies in rater training. *Academy of Management Review, 6*(2), 205-212.
- Brannick, M. T., Michaels, C. E., & Baker, D. P. (1989). Construct validity of in-basket scores. *Journal of Applied Psychology, 74*(6), 957-963. doi:10.1037/0021-9010.74.6.957
- Brogden, H. E. (1949). When testing pays off. *Personnel Psychology, 2*(2), 171.
- Brummel, B. J., Rupp, D. E., & Spain, S. M. (2009). Constructing parallel simulation exercises for assessment centers and other forms of behavioral assessment. *Personnel Psychology, 62*(1), 137-170. doi:10.1111/j.1744-6570.2008.01132.x
- Campion, M., Ployhart, R., & MacKenzie, W. (2014). The state of research on situational judgment tests: A content analysis and directions for future research. *Human Performance, 27*(4), 283-310. doi:10.1080/08959285.

- Carretta, T. R. (1992). Short-term retest reliability of an experimental U.S. Air Force pilot candidate selection test battery. *The International Journal of Aviation Psychology*, 2(3), 161-173. doi:10.1207/s15327108ijap0203_12014.929693
- Centre for Economics and Business Research. (2014). Cost of small business employment. Retrieved from <https://cebr.com/reports/cost-of-small-business-employment/>
- Dunlop, P. D., Morrison, D. L., & Cordery, J. L. (2011). Investigating retesting effects in a personnel selection context. *International Journal of Selection & Assessment*, 19(2), 217. doi:10.1111/j.1468-2389.2011.00549.x
- Dutke, S., Barenberg, J., & Leopold, C. (2010). Learning from text: Knowing the test format enhanced metacognitive monitoring. *Metacognition & Learning*, 5(2), 195-206.
- Equal Employment Opportunity Commission. (1978). The uniform guidelines on employee selection procedures. Retrieved from <https://www.gpo.gov/fdsys/pkg/CFR-2016-title29-vol4/xml/CFR-2016-title29-vol4-part1607.xml>
- Equal Employment Opportunity Commission. (2007). *Employment Tests and Selection Procedures* [Fact sheet]. Retrieved from https://www.eeoc.gov/policy/docs/factemployment_procedures.html
- Gilliland, W. S. (1993). The perceived fairness of selection systems: An organizational justice perspective. *The Academy of Management Review*, 18(4), 694-734.

- International Task Force on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, 41(4), 1244-1273.
- Hausknecht, J. j. (2010). Candidate persistence and personality test practice effects: Implications for staffing system management. *Personnel Psychology*, 63(2), 299-324. doi:10.1111/j.1744-6570.2010.01171.x
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Gerrard, M. M. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92(2), 373-385.
- Hausknecht, J. P., Trevor, C. O., & Farr, J. L. (2002). Retaking ability tests in a selection setting: implications for practice effects, training performance, and turnover. *Journal of Applied Psychology*, 87(2), 243-255.
- Jackson, A. T., Culbertson, S. S., Kausel, E. E., Young, M. E. (May 2017). The impact of escalation decisions on investments, anger, and confidence over time. Poster presented at the 2017 Congress of the European Association of Work and Organizational Psychology, Dublin, Ireland.
- Jackson, A. T., Howes, S. S., Kausel, E. E., Young, M. E., & Loftis, M. E. (under review). The impact of escalation decisions on investments, anger, and confidence over time.
- Kelley, P. L., Jacobs, R. R., & Farr, J. L. (1994). Effects of multiple administrations of the MMPI for employee screening. *Personnel Psychology*, 47(3), 575-592.

- Kolk, N. J., Born, M. P., & der Flier, H. v. (2003). The transparent assessment centre: The effects of revealing dimensions to candidates. *Applied Psychology: An International Review*, 52(4), 648-668. doi:10.1111/1464-0597.0015
- Kulik, J. A., Kulik, C. C., & Bangert, R. L. (1984). Effects of practice on aptitude and achievement test scores. *American Educational Research Journal*, 21(2), 435-447.
- Landers, R. N., Sackett, P. R., & Tuzinski, K. A. (2011). Retesting after initial failure, coaching rumors, and warnings against faking in online personality measures for selection. *Journal of Applied Psychology*, 96(1), 202-210. doi:10.1037/a0020375
- Lievens, F. f., Buyse, T., & Sackett, P. R. (2005). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology*, 58(4), 981-1007. doi:10.1111/j.1744-6570.2005.00713.x
- Lievens, F., Reeve, C. L., & Heggestad, E. D. (2007). An examination of psychometric bias due to retesting on cognitive ability tests in selection settings. *Journal of Applied Psychology*, 92(6), 1672-1682. doi:10.1037/0021-9010.92.6.1672
- 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), 1043-1055. doi:10.1037/0021-9010.92.4.1043
- Martin, W. (2011). The problem with using personality tests for hiring. Retrieved from <https://hbr.org/2014/08/the-problem-with-using-personality-tests-for-hiring>
- Martens, R., & Landers, D. M. (1970). Motor performance under stress: A test of the inverted-U hypothesis. *Journal Of Personality And Social Psychology*, 16(1), 29-37.

- Maurer, T., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. *Journal of Applied Psychology, 83*(1), 128-136.
- McDaniel, M., & Nguyen, N. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection And Assessment, 9*(1-2), 103-113.
- Puddey, I. B., Mercer, A., Andrich, D., & Styles, I. (2014). Practice effects in medical school entrance testing with the undergraduate medicine and health sciences admission test (UMAT). *BMC Medical Education, 14*(1), 48-63.
doi:10.1186/1472-6920-14-48
- Randall, J., & Villado, A. (2017). Take two: Sources and deterrents of score change in employment retesting. *Human Resource Management Review, 27*(3), 536-553.
doi:10.1016/j.hrmr.2016.10.002
- Saad, S., Carter, G., Rothenberg, M., & Israelson, E. (1999). *Testing and Assessments: An Employer's Guide to Good Practices*, Washington, D.C.: U.S. Department of Labor Employment and Training Administration.
- Schleicher, D. J., Mayes, B. T., Day, D. V., & Riggio, R. E. (2002). A new frame for frame-of-reference training: enhancing the construct validity of assessment centers. *Journal of Applied Psychology, 87*(4), 735-746.
- Schleicher, D. J., Van Iddekinge, C. H., Morgeson, F. P., & Campion, M. A. (2010). If at first you don't succeed, try, try again: Understanding race, age, and gender differences in retesting score improvement. *Journal of Applied Psychology, 95*(4), 603-617. doi:10.1037/a0018920

- 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. (2017). Principles for the validation and use of personnel selection procedures. Retrieved from http://www.siop.org/_principles/principles.pdf
- Spielberger, C. D., Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Van Iddekinge, C. H., Morgeson, F. P., Schleicher, D. J., & Campion, M. A. (2011). Can I retake it? Exploring subgroup differences and criterion-related validity in promotion retesting. *Journal of Applied Psychology*, *96*(5), 941-955.
doi:10.1037/a0023562
- Walmsley, P., & Sackett, P. (2013). Factors affecting potential personality retest improvement after initial failure. *Human Performance*, *26*(5), 390-408.

APPENDICES

APPENDIX A: FIGURES

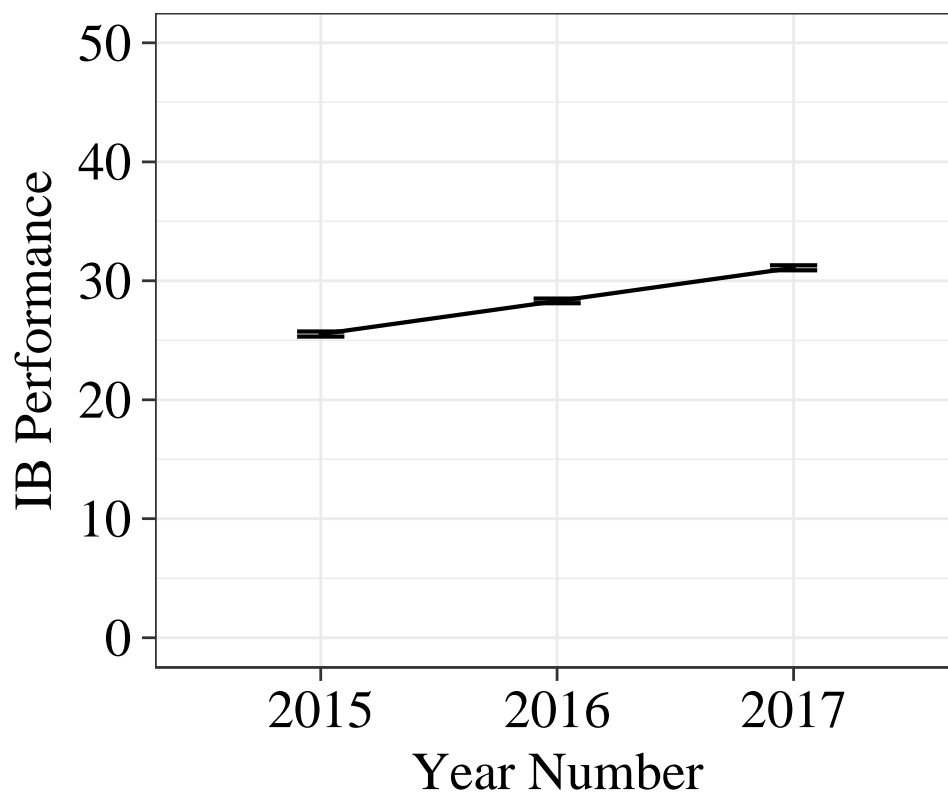


Figure 1. Year predicting in-basket simulation performance for those applying for Sergeant. Error bars represent +/- 1 standard error.

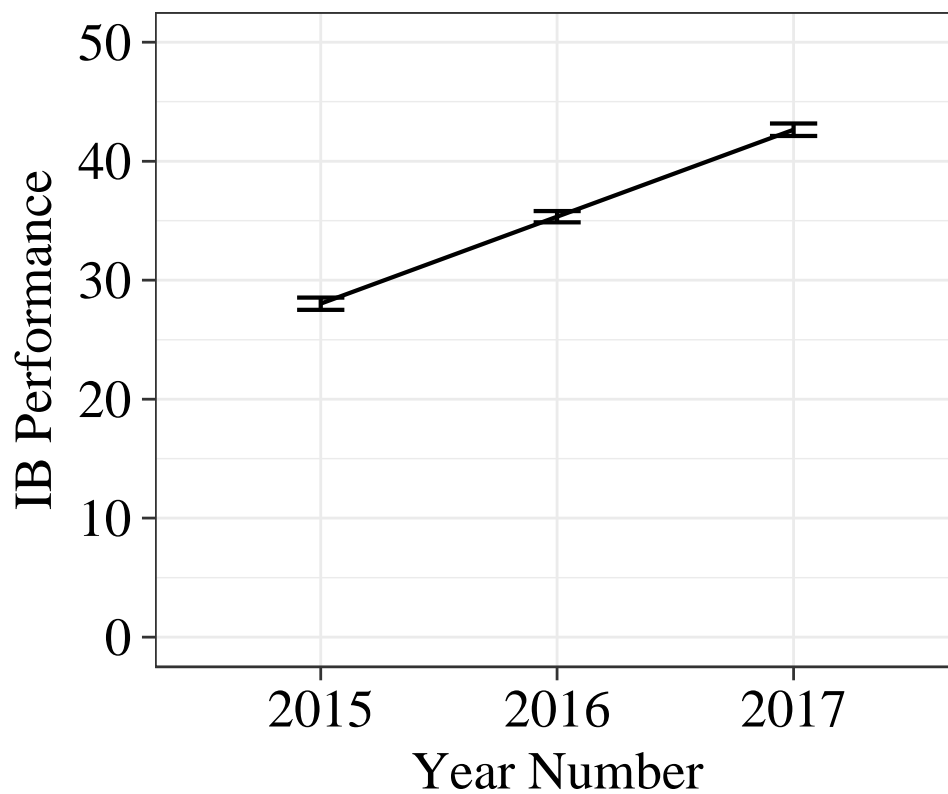


Figure 2. Year predicting in-basket simulation performance for those applying for Lieutenant. Error bars represent +/- 1 standard error.

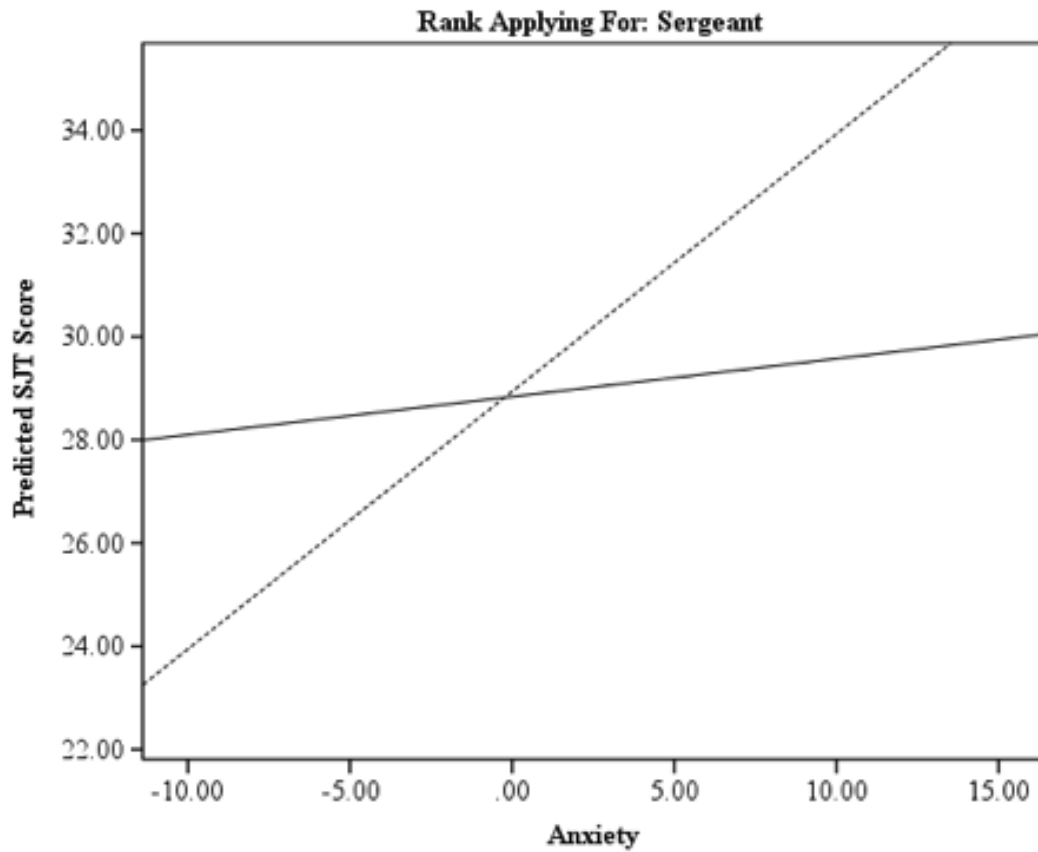


Figure 3. Anxiety moderating the effect of previous experience on situational judgement test performance. The solid line indicates participants who have previous experience. The dotted line indicates participants without previous experience.

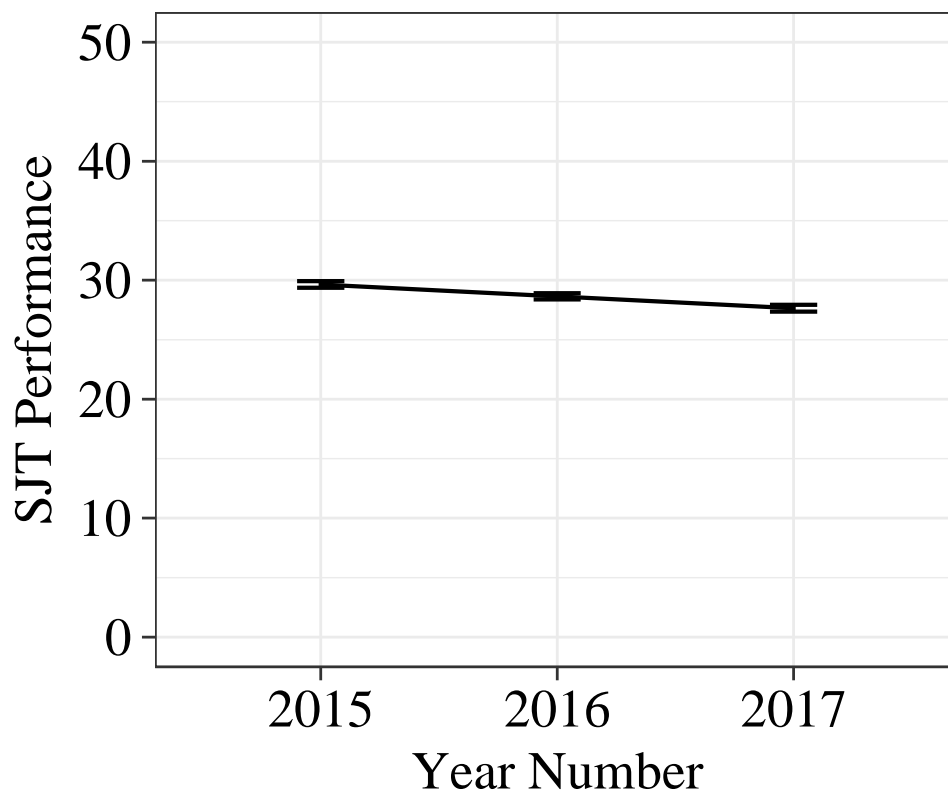


Figure 4. Year predicting situational judgement test performance for those applying for Sergeant. Error bars represent +/- 1 standard error.

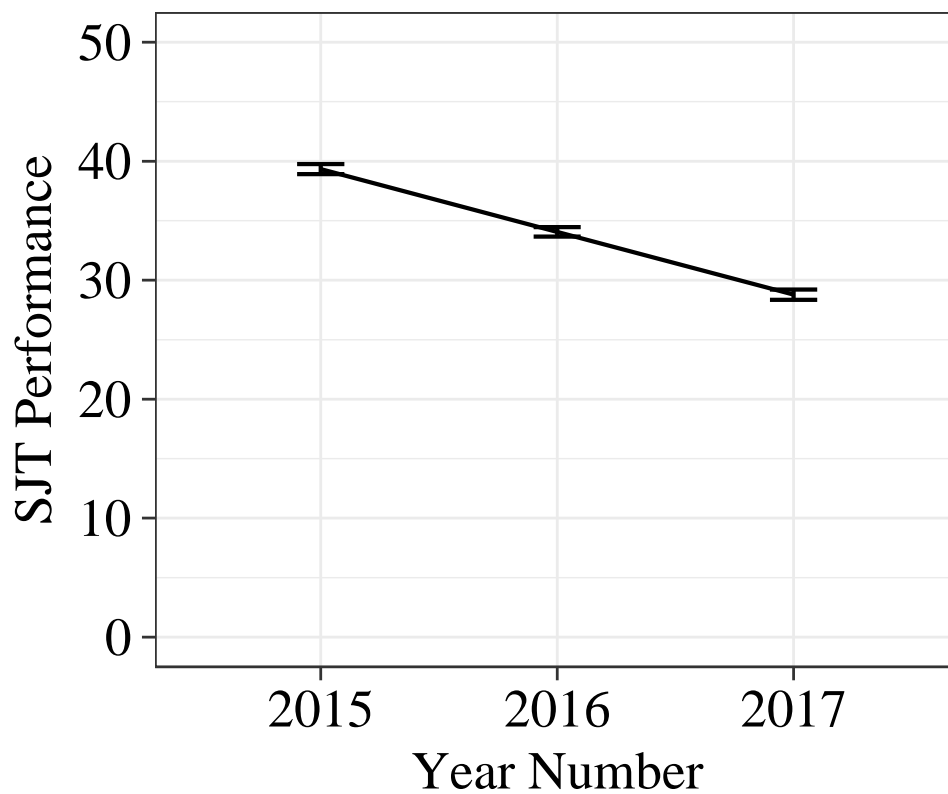


Figure 5. Year predicting situational judgement test performance for those applying for Lieutenant. Error bars represent +/- 1 standard error.