Fall 11-17-2023

Implicit Trait Policies and Situational Judgment Tests: How Personality Shapes Judgments of Effective Behavior

Alexander Edward Johnson
Seattle Pacific University

Recommended Citation

Implicit Trait Policies and Situational Judgment Tests: How Personality Shapes Judgments of Effective Behavior

Alexander Edward Johnson

A dissertation proposal submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

In

Industrial Organizational Psychology

Seattle Pacific University

School of Psychology, Family, & Community

November 17, 2023

Approved by:

Helen Chung, Ph.D.
Assistant Professor of Industrial-
Organizational Psychology
Dissertation Chair

Jorge Lumbreras, Ph.D.
Assistant Professor of Industrial-
Organizational Psychology
Committee Member

Annie Kato Ph.D.
Assistant Professor of Management
Committee Member

Reviewed by:

Paul Yost, Ph.D.
Chair & Associate Professor of Industrial-
Organizational Psychology

Katy Tangenberg, Ph.D
Dean and Professor
School of Psychology, Family, & Community

Dedication

To my Mom and Dad.

Your unconditional love, support, and encouragement have gone before me all the days of my

life. I love you both with all of my heart.

To Josiah, Sofia, and Madelyn.

Being your dad is the greatest joy in my life. I love each of you in a way I never imagined is

possible. Thank you for showing me endless grace throughout this process.

To Katey.

"Some people want diamond rings,

Some just want everything,

But everything means nothing,

If I don't have you."

Table of Contents

List of Tables

List of Figures

List of Appendices

Alexander Johnson
275

Abstract

The Situational Judgment Test (SJT) is a popular selection tool used by employers to make hiring decisions due to their strong predictive validity. SJTs present job candidates with hypothetical scenarios, asking them to choose the responses that best fit those situations. SJTs have been used to measure a range of knowledge, skills, and abilities, but what they measure and why they predict performance remains unclear. Lievens and Motowidlo (2016) called for a reframing of the SJT as a measure of general domain knowledge. According to their theory, SJTs measure procedural knowledge, which is composed of general domain knowledge—operationalized as implicit trait policies (ITPs)—and specific job knowledge. ITPs are beliefs about the effectiveness of expressing trait-related behaviors in various situations that theoretically mediate the relationship between personality and SJT performance.

This study utilized archival data from two samples of firefighters (novice and incumbent) to test the hypothesized relationships within Lievens and Motowidlo's (2016) model, which included personality, implicit trait policies, tenure, and situational judgment test (SJT) performance. The results supported several hypothesized relationships within the model, where ITPs were found to partially mediate the relationship between International Personality Item Pool (IPIP) traits and SJT performance, and ITPs were found to significantly predict SJT performance. Results failed to support the hypothesis that tenure, a proxy for specific job knowledge, moderated the relationship between ITPs and SJT performance. Overall, these findings support the call to reframe SJTs as measures of general domain knowledge, which has implications for both theory and practice. These implications are discussed along with proposed directions for future research.

CHAPTER I

Introduction

Situational judgment tests (SJTs) are a type of psychological assessment that presents job candidates with simulated but realistic on-the-job situations and asks them to make judgments about the most effective course of action (Gatewood et al., 2016). The literature on SJTs dates back to the 1920s, when they were first used by the military to identify suitable candidates for various roles but were largely forgotten until Motowidlo et al. (1990) reintroduced them to the field of selection research in their quest to find more cost-effective selection instruments. SJTs are considered low-fidelity simulations that require fewer resources to develop and administer when compared to high-fidelity simulations such as assessment centers, role plays, or work samples, making them potentially more accessible and affordable to organizations. Low-fidelity simulations subject candidates to written or video-based descriptions of work situations (Lievens et al., 2020). While there were concerns that the lower fidelity of SJTs would compromise their predictive validity, Motowidlo et al. (1990) found that SJTs were as predictive as and cheaper than their high-fidelity counterparts. These findings have led to a flurry of new research on this selection tool.

Motowidlo et al. (1990) saw their findings replicated and expanded upon in the years that followed, providing consistent support for the ability of SJTs to effectively predict job performance across organizational contexts (Christian et al., 2010; McDaniel et al., 2001; McDaniel et al., 2007; Sackett et al., 2021). Further, SJT performance was found to correlate with constructs that were already known to predict job performance, namely cognitive ability and personality dimensions, which are constructs of great value to selection practitioners because of their ability to predict a number of job-related outcomes (Arthur et al., 2014; McDaniel &

Nguyen, 2001; McDaniel et al., 2007). Given the strength of their predictive power and ability to measure a wide range of characteristics, SJTs have become preferred selection tools for organizations seeking to make critical hiring decisions.

However, there is still a lack of consensus regarding what SJTs measure (Motowidlo et al., 2006a; Whetzel & McDaniel, 2009), leading some scholars to call the state of SJT construct validity "a hot mess, without much theoretical or empirical guidance" (McDaniel et al., 2016, p. 50) and "psychometric alchemy" (Landy, 2007, p. 418). Without construct validity, test results may misrepresent candidates' abilities or traits, leading to the hiring of candidates who are not suited for the job and consequently perform poorly, which can lead to several negative organizational outcomes, not the least of which is litigation (Gatewood et al., 2016).

The process of evaluating the construct validity of a measure starts with the creation of a clear theoretical framework that defines the constructs of interest and their theoretical relationships with one another (Shadish et al., 2006). To that end, Motowidlo et al. (2006a) proposed a theoretical model to explain SJT performance. In their model, they posited that SJTs capture a type of general procedural knowledge about which behaviors are most effective in work situations. According to their theory, the antecedents of this knowledge include general life experience and individual personality traits.

In order to test the relationships within their theoretical model, Motowidlo et al. (2006b) operationalized general domain knowledge as *implicit trait policies* (ITPs), which are general beliefs about what type of behavior is most effective in particular situations. These beliefs (ITPs) are shaped by an individual's personality and life experience, and ultimately mediate the relationship between personality and SJT performance. In the years that followed, several studies demonstrated initial support for several relationships within the model, specifically personality

(IPIP traits of conscientiousness, agreeableness, and extraversion) as an antecedent of ITPs, and ITPs as both antecedents of SJT performance and as overall predictors of job performance (Motowidlo et al., 2006b; Miller et al., 2008; Motowidlo & Beier, 2010; Oostrom et al., 2012). However, these studies did not include all IPIP traits (namely emotional stability), or a measure of specific job knowledge and were conducted on samples drawn from undergraduate students in academic settings, limiting the generalizability of the findings. The limitations of these initial findings suggest that further investigation into the relationships within the model is warranted.

In 2016, Lievens and Motowidlo updated and expanded upon the Motowidlo et al. (2006a) model to include additional antecedents of SJT performance, including cognitive ability, emotional intelligence, values, and interests and added specific job knowledge as a component of general domain knowledge. This present study drew upon the expanded Lievens and Motowidlo (2016) model and evaluated the role of general domain knowledge (operationalized as ITPs), or the beliefs about what personality traits should be expressed in hypothetical scenarios, as a mediator of SJT performance within a previously untested population of firefighters. Additionally, it examined whether job tenure, a proxy for specific job knowledge, moderates the relationship between ITPs and SJT performance. In doing so, this study aimed to provide clarity about what SJTs measure and why they are predictive of job performance.

Purpose of this Study

This present study was a replication study and as such, it aimed to contribute to the viability of the Lievens and Motowidlo (2016) model by confirming the validity of previous findings. It attempted to test whether earlier findings are generalizable to different populations by analyzing components of the model on a previously untested population of firefighters. By refining the construct domains of the SJTs, evaluating the potential validity of ITPs as a

predictive construct, and providing a modified method for ITP calculation, this present study hopes to clarify the boundaries of SJTs, discern the applicability of previous findings, and inform practitioners on the design and development of more cost-effective SJTs for use in employee selection. This study also hoped to advance the understanding of the way ITPs enable some candidates to perform well on SJTs.

This study utilized archival data provided by the Company[1], a public service testing agency, gathered from a proprietary assessment tool called the Firefighter Assessment[2] (FA). The FA is used by fire departments across the country to select entry-level firefighter candidates. The FA consists of two sections: a video-based assessment with 60 items and a 260 item self-assessment component that measures individual traits related to job performance. The video-based assessment contains 40 situational judgment scenarios designed to measure interpersonal skills (see Appendix A for a sample SJT item). Two comprehensive datasets were used, which included scores from novice firefighters who took the FA during their application process, along with scores from incumbents gathered as part of the FA validation study.

This paper begins with a review of the literature about SJTs and the related challenges of construct validity, including an overview of implicit trait policies (ITPs), the role they play in Lievens and Motowidlo's (2016) theory of SJT knowledge determinants and how they are shaped by Five Factor Model (FFM) personality traits. Following a review of relevant literature, the methods used to evaluate the hypotheses are discussed, a description of the sample drawn from an archival dataset of entry-level firefighters is provided, and the SJT and IPIP FFM personality measures are defined. The procedures used to conduct the analyses are outlined,

---

[1] A pseudonym, Company, is used by request of the organization for confidentiality purposes.
[2] Firefighter Assessment is a pseudonym used in place of the instrument to protect confidentiality at the request of the Company.

concluding with a discussion of the results, limitations of the study, and implications for future research.

Literature Review

Situational Judgment Tests

The situational judgment test (SJT) is a simulation-based selection tool used to predict future job performance by assessing applicants' responses to simulated work situations (Matthews et al., 2018). Selection tools that utilize simulations are developed on the foundation of behavioral consistency logic, expressed in the adage that the best predictor of future behavior is past behavior (Wernimont & Campbell, 1968). To capture relevant behavioral samples that can be used to predict job performance, candidates are evaluated based on how they behave in response to simulated work scenarios that mimic actual job conditions (Motowidlo et al., 1990). These behavioral samples are evaluated and used to make inferences about how a candidate will perform on the job (Lievens & De Soete, 2012).

Selection tools vary in the fidelity to which their simulations approximate actual job conditions. According to Motowidlo et al. (1990), high-fidelity simulations, such as assessment centers and work samples, create realistic simulations using actual work materials and settings to create a situation in which candidates are asked to respond as if they were in the job situation. Low-fidelity simulations use hypothetical situations, typically presented verbally or in writing, and ask candidates to identify how they would respond to a series of response options (Lievens et al., 2020). Selection tools that utilize high-fidelity simulations have demonstrated strong predictive validity (Hunter & Hunter, 1984; Sacket et al., 2021) and tend to be received well by candidates because of their fidelity to the job, creating a sense of fairness in the measure (Weekley & Ployhart, 2006). However, developing and administering these high-fidelity simulations requires significant investment of time and resources.

SJTs were reintroduced to the field of selection as part of an effort to determine whether low-fidelity selection tools, which are cheaper to develop and administer, predict job performance as effectively as their high-fidelity counterparts (Motowidlo et al., 1990). Unlike assessment centers that assess actual candidate behavior in a simulated work environment, SJTs assess how candidates believe they would behave in a hypothetical situation (Weekley & Ployhart, 2006). If past behavior is the best predictor of future behavior and high-fidelity simulations produce better behavioral samples, then high-fidelity simulations should predict job performance better than low-fidelity simulations. However, Motowidlo et al. (1990) found that low-fidelity SJTs predicted supervisory ratings of job performance ($r = .30$), as well as their high-fidelity counterparts ($r = .32$). These findings have led to a renewed interest in SJTs by both researchers and practitioners.

Since their resurgence in the early 1990s, research has consistently demonstrated that SJTs reliably predict job performance across organizational contexts (Christian et al., 2010; McDaniel et al., 2001; McDaniel & Nguyen, 2007; Sackett et al., 2021) and have smaller mean group score differences ($d = .4$) than cognitive ability tests ($d = .8$) (Motowidlo & Tippins, 1993; Sackett et al., 2021; Weekley & Jones, 1999). SJTs demonstrate similar validity to cognitive ability tests, with an estimated corrected mean validity of .26 (compared to .31 for cognitive ability tests) and provide incremental validity above and beyond cognitive ability and personality assessments, suggesting that SJTs measure something beyond conventional predictors of job performance (Clevenger et al., 2001; Weekley & Ployhart 2006).

Situation Descriptions

Traditional SJT theory views the situation description as the foundation of an effective SJT measure (Campion & Ployhart, 2013). Situation descriptions are designed to mimic the

psychological, intellectual, and physical aspects of the target role in such a way that candidates can easily imagine themselves in the situation. Candidates are then asked to respond by indicating how they would handle the situation if they were in it. Their responses are then scored against a key created by subject matter experts (SMEs) who determine what course of action is most effective (Motowidlo & Beier, 2010). When a candidate's judgment about the most effective course of action aligns with the SMEs' judgments, the candidate is thought to possess specific job knowledge about effective behavior on the job, will consequently score higher on the SJTs, and determined to be more suitable for the role.

Similar to simulation fidelity, situation descriptions exist on a spectrum, often described as "richness," which can be affected by the testing medium (verbal, written, or video) (Lievens & DeSoete, 2012). Video-based SJTs, which allow for "rich" situation descriptions, have demonstrated the ability to predict interpersonal constructs over and above written versions that typically cannot provide a rich description of a situation (Lievens & Sackett, 2006). Candidates typically prefer SJTs with richer situation descriptions because of their fidelity to the job to which they are applying (Chan & Schmitt, 2002).

However, recent studies have challenged the importance of situation descriptions for SJT performance. Studies comparing candidate performance on SJTs with and without situation descriptions (see Appendix B for example SJT items with and without situation descriptions) found that situational information was not required to answer SJT items correctly. These studies revealed that the characteristics of the candidates played a more significant role in the predictive capabilities of SJTs than the characteristics of the SJTs themselves (Jackson et al., 2017; Krumm et al., 2015). These findings, based on written SJTs, also extended to video-based SJTs, where the presence or absence of situation descriptions has been shown to have no significant effect on

the construct- and criterion-related validity of SJTs (Schaepers et al., 2020). The results of these studies (Krumm et al., 2015; Jackson et al., 2017) indicate that while SJTs are predictive of job performance, there continues to be lack of clarity about which constructs are actually being measured.

If SJTs predict job performance with and without situational information, this suggests that SJTs may not measure situation-specific constructs but rather context-independent constructs such as personality (Freudenstein et al., 2023) or general knowledge about effective behavior in work settings (Lievens & Motowidlo, 2016). Both situation-specific and context-independent constructs have unique value as predictors; however, the greater question at hand is whether SJTs *intentionally* measure those constructs. By understanding the role of context-independent constructs such as personality and general knowledge, we can potentially develop more effective SJTs that do not include unnecessary detail or development costs.

Personality and Situational Judgment Test Performance

If situation-specific factors are not as important as once thought, context-independent constructs such as personality may play a bigger role in SJT performance than previously thought. Five Factor Model (FFM) traits and SJT performance have consistently been linked to one another (McDaniel et al., 2001; McDaniel & Nguyen, 2007; Lievens et al., 2020; Lievens & Motowidlo, 2016). This relationship is so strong that some researchers have argued that SJTs can be used as measures of personality (Judge et al., 2013; Lievens & Schaepers et al., 2020). FFM dimensions of emotional stability ($r = .31$), conscientiousness ($r = .26$), and agreeableness ($r = .25$) demonstrated positive relationships with SJT performance, while extraversion ($r = .06$) and openness ($r = .09$) demonstrated weak relationships (McDaniel & Nguyen, 2001).

Initial proposals to explain the relationship between personality and SJT performance argued that SJT performance is a product of the interaction between the trait characteristics of the respondent and the trait expressiveness of the SJT (Weekley & Ployhart, 2006). Social judgment research has provided evidence regarding how people judge the trait characteristics of others in accordance with their own standing on that characteristic (Dunning & Hayes, 1996; Hovland & Sherif, 1952; Lambert & Wedell, 1991). More recently, research has extended these findings to show that different personality traits influence candidate perceptions of situations and predict judgment of behaviors (Serfass & Shermann, 2013). Given that SJTs are typically used to measure constructs related to interpersonal relationships (Christian et al., 2010), this would explain why those who score higher on traits related to interpersonal skills would perform better on SJTs.

Trait activation theory has also been proposed to explain the relationship between personality and SJT performance. FFM research has shown that even though personality traits are relatively stable across an individual's lifespan and across situations (Saucier & Goldberg, 2001), the expression of these traits is contingent on the situation. Variance in trait expression across contexts can be attributed to contextual cues that activate or deactivate the expression of particular traits (Tett et al., 2003). This explains why an individual who scores high on trait extraversion may appear very extraverted in one situation and less extraverted in another. When extraversion is important for success in a role (e.g., a job in sales), it is critical for organizations to find people whose extraversion is activated by their organizational context.

SJTs seek to replicate this phenomenon by exposing candidates to hypothetical work situations designed to mimic job conditions. These job conditions present candidates with cues to activate traits that inform how they perceive and judge which response options are most

effective. Thus, attempts to explain the relationship between personality and SJT performance based on trait activation theory depend on the interaction between the personality of the respondent and trait expressiveness of the situation description. The viability of this theory is called into question by the aforementioned findings (Jackson et al., 2017; Krumm et al., 2015). Without a situation description with which to interact, there are no situationally activated traits, suggesting that the way personality impacts SJT performance is not through interaction with the situation itself.

In contrast, Motowidlo et al. (2006a) explained the relationship between personality and SJT performance in terms of *dispositional fit.* According to *dispositional fit*, the more strongly an individual expressed a given personality trait, the more likely they were to believe that a behavior which aligns with that trait will be more effective. For example, when a candidate is presented with an SJT scenario and asked to select the response option they deem to be most effective, those who score higher on a trait will be more inclined to select the response option which most strongly expresses that trait, regardless of its effectiveness. What *dispositional fit* could not account for, however, were those who could select the correct response option when the response option strongly expressed a trait on which their standing was low. Motowidlo et al. (2006a) believed this indicated a distinction between one's standing on a trait and one's knowledge about when a behavior that expresses that trait is most effective. For example, someone might be highly extraverted, but lack the knowledge of when it is most effective to act in an extraverted manner. This knowledge is a type of procedural knowledge that, they argued, is captured by SJT scores.

SJTs as Measures of Procedural Knowledge

In 2006, a theory of procedural knowledge was published to explain how personality traits influence STJ performance (Motowidlo et al., 2006a). According to this theory (see Figure 2), personality shapes SJT performance *through* implicit trait policies (ITPs), which are "implicit beliefs about the effectiveness of different levels of trait expression" (Motowidlo et al. 2006a, p. 57) and an individual's experiences. According to this theory, both individual personality traits and experiences are causal antecedents of their related ITPs, and these beliefs (ITPs) mediate the relationship between personality and SJT performance. Consequently, those who scored higher on a trait were more likely to judge the behavior expressing the trait to be most effective. Implicit trait policies (ITPs), along with an individual's experience, interact to create procedural knowledge, which is knowledge about effective and ineffective behavior for a given situation. They argue that procedural knowledge is measured by SJTs.

Figure 1

Theory of Procedural Knowledge (Motowidlo et al., 2006)
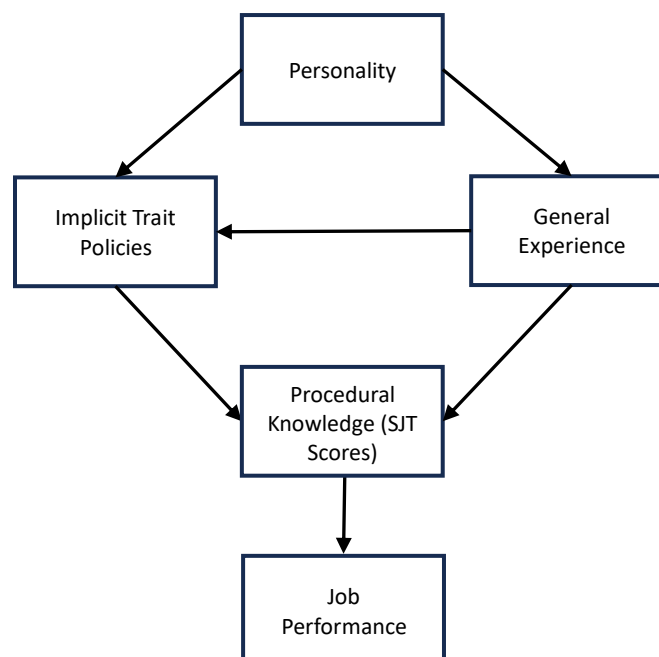
SJTs as Measures of General Domain Knowledge

As evidence calling into question the importance of situation descriptions on SJT performance began to mount (Krumm et al., 2015; Jackson et al., 2017), Lievens and Motowidlo (2016) called for the reconceptualization of SJTs as measures of general domain knowledge and offered an updated model of the knowledge determinants underlying SJT performance (see Figure 2). This model was expanded in two significant ways from the Motowidlo et al. (2006a) model. First, additional trait constructs were added as antecedents of ITPs (i.e., emotional intelligence, interests, values, and ability). Second, they updated procedural knowledge to include a component of specific job knowledge alongside general domain knowledge (operationalized as ITPs).

Figure 2

Lievens and Motowidlo (2016) Expanded Model
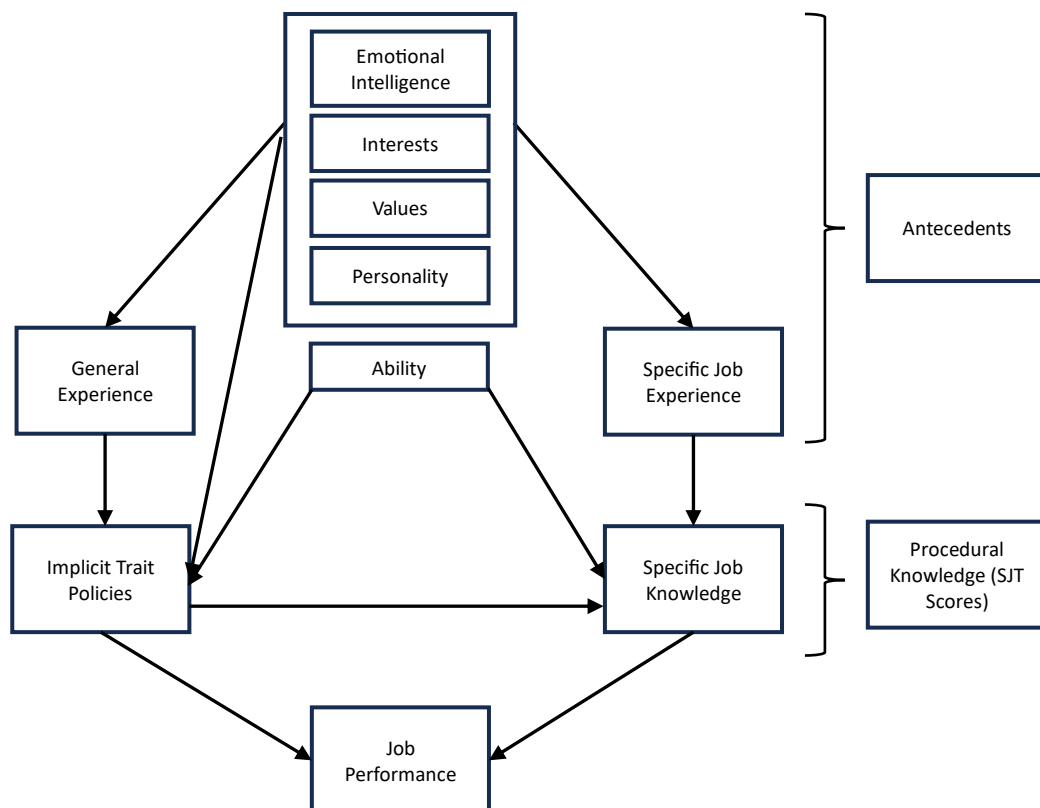
Within this updated model, Lievens and Motowidlo (2016) defined general domain knowledge as "knowledge about the utility or importance of traits such as these for effectiveness in a job that actually requires expressions of those traits for effective performance" (Lievens & Motowidlo, 2016, p. 2). Candidates who know that conscientious behavior leads to better job performance than unconscientious behavior possess greater knowledge of the utility of conscientious behavior at work. General domain knowledge is in essence a broad, context-independent knowledge about which personality trait-related behavior is most effective, whereas specific job knowledge is context-dependent knowledge of which trait-related behavior is most effective in a particular job. General domain knowledge can be learned through typical socialization processes that take place at school or in the home, while specific job knowledge must be gained from the experience of doing the job or through explicit training (Lievens & Motowidlo, 2016). According to conventional SJT theory, those with specific job knowledge will be better able to identify the most effective course of action in response to a highly contextualized SJT scenario. As such, SJTs have traditionally been designed to measure this specific job knowledge.

Research has provided some evidence to support this updated model in which both general domain knowledge and specific job knowledge make unique contributions to SJT performance. One such study by Motowidlo and Beier (2010) parsed the unique contributions of specific job knowledge and general domain knowledge (ITPs) to SJT performance by comparing the validity of SJT scoring keys developed by two different groups: SMEs with specific job knowledge about effective behavior on the job and undergraduate students with no specific job knowledge. SJT scores produced by both SME and undergraduate student keys demonstrated significant correlations with job performance ($r = .37$ and $r = .29$, respectively). Another study

by Lievens and Patterson (2011) compared the incremental validity of assessment centers (high-fidelity simulation) and SJTs (low-fidelity simulation), over job knowledge tests. They found that both assessment centers and SJTs explained additional variance in job performance above and beyond the knowledge test (5.9% and 5.7%, respectively). These studies introduce important findings about the distinction between general domain knowledge and specific job knowledge in SJTs that warrant further research.

Despite both general and specific job knowledge making unique contributions, Lievens and Motowidlo (2016) argued that moving into the future, researchers should focus on the measurement of general domain knowledge. As previous research has demonstrated, SJTs can measure general domain knowledge without situation descriptions that are saturated with specific job knowledge, which means that test development can be more cost-effective and less labor-intensive.

The Five Factor Model of Personality

The Five Factor Model of Personality (FFM), or the "Big Five," is used in research and applied settings to describe individual personality in terms of five primary personality traits that exist along a spectrum: extraversion, agreeableness, openness, conscientiousness, and neuroticism (Neal et al., 2012). These traits are relatively stable across the lifespan and culture (Saucier & Goldberg, 2001) and predict a number of life outcomes, such as mental and physical health, life satisfaction, divorce, and many others (Barrick & Mount, 1991; Malouff et al., 2005; Bagby et al., 1997). Additionally, these traits are predictive of work outcomes relevant to selection researchers, including job performance (Barrick et al., 2001; Salgado, 2003; Judge et al., 2013; Sacket et al., 2021), job satisfaction (Judge, Heller, & Klinger, 2008), and leadership effectiveness (Judge & Piccolo, 2004). These predictive powers also hold true for SJT

performance. The FFM dimensions of emotional stability ($r = .31$), conscientiousness ($r = .26$), and agreeableness ($r = .25$) demonstrate positive relationships with SJT performance and are frequently considered in hiring decisions, while extraversion ($r = .06$) and openness ($r = .09$) demonstrate weak relationships and are typically only considered for specific roles (McDaniel & Nguyen, 2001).

Emotional Stability

Previously labelled neuroticism, the term emotional stability, is now preferred by a growing number of researchers and clinicians as a framing that avoids the potential for pathologizing this trait (Costa & McCrae, 2008; Goldberg, 1990; Lamers et al., 2012). Emotional stability is defined as an individual's ability to remain calm and balanced under stress (Judge et al., 2004), Individuals who score high on this trait are usually calm, composed, less likely to feel anxious or depressed, and more able to handle stress (Gardner & Dunkin, 2018). Those who score low in emotional stability tend to experience more mental and physical health problems, relationship difficulties, and impulsivity (Stasielowics, 2020; Takacs et al., 2022; Trull & Widiger, 2022). Emotional stability tends to increase with age, leveling out after age 40 (Debast et al., 2014; Lucas & Donnellan, 2009; McAdams & Olson, 2010; McCrae & Costa, 2003), and is positively related to several work-related outcomes, including job performance, job satisfaction, turnover intention, and leader effectiveness (Barrick & Mount, 1991; Judge et al., 2014). Because firefighters experience periods of extreme stress, it is thought that those who score high in emotional stability will potentially manage stress more effectively and perform better on the job. Thus, this trait is included in the FA.

Conscientiousness

Conscientiousness is defined as an individual's tendency to be dependable, responsible, organized (Costa & McCrae, 1985), achievement-oriented (Motowidlo et al., 2018); self-disciplined and competency-oriented (Takacs et al., 2022). Conscientious individuals are typically willing to conform to group norms and organizational rules (Smithikrai, 2008) and contribute to team functioning (Hofmann & Jones, 2005). Because of its ability to predict several job-related criteria, such as job performance, job satisfaction, and salary (Ng et al., 2005), employers frequently consider conscientiousness when selecting personnel (Gardner & Dunkin, 2018). Among firefighters, who are expected to be thorough and adhere to group norms within a hierarchy, conscientiousness is a valuable trait; consequently, the FA was designed to measure this trait.

Agreeableness

Agreeableness is defined as an individual's tendency to be cooperative, considerate, and understanding (Costa & McCrae, 1985). Individuals who score high on this trait are perceived as empathetic, kind, compassionate, pleasant, easy to get along with (Costa & McCrae, 2008) and tend to become increasingly agreeable with age (Debast et al., 2014; Lucas & Donnellan, 2009; McAdams & Olson, 2010; McCrae & Costa, 2003). Firefighters typically work 24-to-48-hour shifts during which they work and live alongside their team doing everything from routine tasks of daily living (e.g., eating, cleaning, etc.) to completing high stress job-specific tasks. The ability to effectively navigate interpersonal situations is critical to success on the job. Because of this, the Company included items for agreeableness in the FA.

Implicit Trait Policies

Within the Lievens and Motowidlo (2016) model, general domain knowledge has been operationalized as implicit trait policies (ITPs). ITPs are the extent to which people know when it is most appropriate to express personality trait related behavior in work settings. Research into ITPs began to appear in the early 2000s (Hooper et al., 2004; Motowidlo et al., 2003), but lacked a cohesive theoretical model until Motowidlo et al. (2006a) published the theory of procedural knowledge to explain the causal antecedents of SJT performance (See Figures 1 and 2). Based on the foundational assumption that SJTs measure procedural knowledge, they introduced the concept of *implicit trait policies* to explain the relationship between personality traits and SJT performance. They defined ITPs as "implicit beliefs about the effectiveness of different levels of trait expression" (Motowidlo et al., 2006a; p. 57). According to their theory, both individual personality traits and experiences are causal antecedents of related ITPs, and these beliefs (ITPs) mediate the relationship between personality and SJT performance. They theorized that those who score higher on a trait, such as conscientiousness, will weigh it more heavily when judging the effectiveness of SJT response options.

Later that year, Motowidlo et al. (2006b) expanded on the theoretical foundations of ITP theory, drawing from the vast field of social judgment research, which has consistently furnished evidence that the "self appears to play a pervasive role in people's judgments of others" (Dunning & Hayes, 1996; p. 213). Motowidlo et al. (2006a, 2006b) drew upon well-established effects in social judgment research (such as contrast, assimilation, and accentuation) to derive the foundational principle of ITP theory, that "individual differences in personality traits affect judgments of the effectiveness of behavioral episodes that express those personality traits" (Motowidlo et al., 2006, p. 749). They introduced the concept of dispositional fit, which suggests

that people tend to evaluate behavior that expresses their own personality traits as more effective

than behaviors that do not express their own traits (Motowidlo & Beier, 2010; Motowidlo, 2003).

Applied to ITP theory and SJT research, this suggests that as respondents score higher on a given

trait, such as agreeableness, they are more likely to judge the response options that express

agreeableness as effective.

ITPs are considered *implicit* measures because they are not directly measured. With

implicit measures, the respondents are unaware of the constructs being evaluated. A candidate's

rating on an implicit construct is the byproduct of their responses to items that may appear to

measure another construct (Fazio & Olson, 2003). Implicit measures are valued by selection

practitioners, as respondents tend to be less governed by social desirability (Fazio & Olson,

2003), faking, and self-presentation biases (Bornstein, 2002), which are documented issues in the

use of self-report personality measures in selection contexts (Birkeland et al., 2006; Morgeson et

al., 2007).

Calculating Implicit Trait Policies

Motowidlo et al. (2006b) offered four different methods for calculating ITPs. They used

two of these methods in their study and provided two additional alternatives. The first three of

the methods described utilized respondent ratings of the effectiveness of each SJT response

option. In the first method, ITPs were then calculated by computing the correlation between

these respondent effectiveness ratings with the intended level of the trait expressed (high or low).

In the second method, Motowidlo et al. (2006b) subtracted the sum of the effectiveness ratings

for options designed to express low levels of a trait from the sum of the effectiveness ratings for

options designed to express high levels of a trait. These first two methods were found to correlate

strongly with one another ($r = .95$, $p < .01$) and exhibited the same pattern of relationships with

the self-reported scores for each personality trait (Motowidlo et al., 2006b), indicating each as a viable method. These methods were utilized in the Motowidlo et al. (2006b) study.

The third method offered, but not used, was to formally calculate ITPs as an effect size. Conceptually, ITPs are effect sizes, or the extent to which individual personality traits impact respondents' judgments about the most effective response options (Freudenstein et al., 2023; Motowidlo et al., 2006a, 2006b; Motowidlo & Beier, 2010; Oostrom et al., 2012). Like the second method, this method would also utilize the respondents' ratings for the effectiveness of each response option and subtract the sum of the effectiveness ratings for options designed to express low levels of a trait from the sum of the effectiveness ratings for options designed to express high levels of a trait. This difference would then be divided by the average standard deviation around each mean.

The final method offered, and utilized in this current study, was a way of *estimating* the effect sizes of ITPs. In this method, these estimates would be calculated by computing the difference between the number of response options selected with low-trait expressiveness from the number of response options selected with high trait-expressiveness. In contrast to the first three methods, this final method did not depend on the respondent ratings of the effectiveness of each SJT response option. The current study utilized this final method in large part due to the nature of the archival data that did not include respondent ratings of effectiveness.

Further support for use of this method comes from the concept of the accentuation effect from social judgment research, a theory relied upon heavily by Motowidlo et al.'s (2006a) ITP theory. The concept of the accentuation effect is a psychological phenomenon where contextual information exaggerates the perceived differences between options or stimuli, leading to more polarized judgments (Eiser & van der Plight; Hovland & Sherif, 1952; Tajfel, 1957). Lambert

and Wedell (1990) applied this to their research about how personality traits influence judgments of behavioral episodes. After having participants complete a self-report assessment of sociability, they were then asked to rate the level of sociability expressed by written descriptions of behavioral episodes. They computed an index of evaluative extremity by subtracting favorability rating that express low sociability from those statements that expressed high sociability and found this index to correlate with the self-report measure of personality ($r = .44$; Lambert & Wedell, 1991).

ITP scores for this present study were calculated as estimates of ITP effect sizes in alignment with the method Lambert and Wedell (1991) used to calculate the index for evaluative extremity. This method was chosen, in part, due to the nature of the archival data provided by the Company which did not ask respondents to rate the effectiveness of each response option. It was also chosen as it aligns with the design of the SJTs in the FA. When developing an SJT, the developer must decide whether to create response options with only one correct response, or several response options which vary in their level of effectiveness. This decision is typically dependent upon whether there is one correct way to respond or if there are several reasonable responses (Weekley et al., 2006). The FA is designed with one correct response option for each SJT stem. Scoring ITPs in a similar dichotomous method aligns with the design of the SJT.

This Study

This study tested the components of the Lievens and Motowidlo (2016) model (see Figure 3 for dissertation model being tested) on a population of firefighters using archival data. The Company developed the Firefighter Assessment (FA) to identify candidates with the requisite knowledge, skills, and abilities to successfully perform the job of a firefighter. Firefighters are embedded in heavily regulated and scrutinized work environments, where there

are defined expectations for interpersonal interactions that are thought to be unique to the fire

service (Fernandez et al., 2020). Given the importance of the ability to navigate interpersonal

situations and the strength of SJTs in measuring interpersonal skills (Christian et al., 2010), the

Company included 40 SJT scenarios in the FA. Each item provided candidates with a situation

via a short video containing a brief but rich situation description of a common interpersonal

situation faced by firefighters. Based on traditional SJT theory that underscores the importance

of rich situation descriptions, candidates who possess firefighter-specific knowledge would be

expected to perform better on the SJTs and ultimately better on the job.

   The Firefighter Assessment (FA) was also designed to measure personality traits that

predict job success. According to their internal research and findings, the Company found

emotional stability, conscientiousness, and agreeableness to predict job performance, and

consequently included these three traits in the FA assessment. Because this study utilized the

Company's archival data, only these three traits were considered.

Personality and Implicit Trait Policies

   Early ITP research has focused on the relationship between FFM personality traits and

their associated ITPs. Several studies have found significant correlations between ITPs for

agreeableness and NEO FFM personality scores (Miller et al., 2008; Motowidlo et al., 2003;

Motowidlo et al., 2006a, 2006b). Additionally, ITPs for agreeableness were found to correlate

with agreeableness scores calculated by assessors in role-play scenarios (Motowidlo et al.

2006b). The consistent relationship between ITPs for agreeableness and NEO agreeableness

scores may be explained by the majority of SJTs being designed to measure interpersonal skills

(Christian et al., 2010), which are correlated with agreeableness (Judge et al., 2008).

Additionally, multiple studies found significant correlations between ITPs for conscientiousness

and NEO personality scores (Freudenstein et al. 2023; Miller et al. 2008; Motowidlo et al. 2006a; Oostrom et al. 2012).

As stated previously, the Company included conscientiousness, agreeableness, and emotional stability within the FA. The Company excluded extraversion and openness to experience as those personality traits did not demonstrate relevance to the job of a firefighter. Previous research has demonstrated that extraversion and openness to experience are not effective predictors of job performance across occupations and are only useful predictors of job performance for a small number of jobs (Judge et al., 2008). This, in addition to the fact that they demonstrate weak relationships with SJT performance ($r = .06$ and $r = .09$, respectively) (McDaniel &Nguyen, 2001), is the rationale for their exclusion from this study.

In contrast to extraversion and openness to experience, the personality trait of emotional stability does demonstrate significant correlations with SJT performance ($r = .31$; McDaniel & Nguyen, 2001). Given these previous findings, this study expected that the relationship between IPIP emotional stability and its corresponding ITP will mirror the significant relationships between agreeableness and conscientiousness, leading to the first hypothesis:

Hypothesis 1: Individual trait levels for conscientiousness, agreeableness, and emotional stability will be positively related to their associated ITPs, such that the greater the levels of these traits, the greater the associated ITPs.

Implicit Trait Policies and Situational Judgment Test Performance

Currently, SJTs and ITPs are inseparable because SJTs are the only known measurement tools capable of effectively assessing individual differences in ITPs (Freudenstein et al., 2023). As described above, several studies have found significant relationships between ITPs and SJT performance (Motowidlo & Beier, 2010; Oostrom et al., 2012). ITPs also demonstrate a stronger

relationship with SJT performance than either personality traits or experience alone (Oostrom et al., 2012), leading to the second hypothesis:

Hypothesis 2: ITPs for conscientiousness, agreeableness, and emotional stability will predict SJT performance.

Implicit Trait Policies as Mediators

ITP theory holds that personality traits have a direct effect on SJT performance and an indirect effect on SJT performance through ITPs (Lievens & Motowidlo, 2016; Motowidlo et al., 2006a, 2006b). This means a portion of how personality traits impact SJT performance is determined by an individual's general beliefs about what behavior will be most effective. An individual may score highly on a given trait but have learned through life experience that expressing that trait has negative consequences in certain situations. Consequently, scoring high on a personality trait and knowing when it is best to express that trait are two related, but distinct concepts (Lievens & Motowidlo, 2016). Despite the existence of this relationship in each ITP model, no study was found that has directly evaluated ITPs as mediators of the relationship between their associated personality traits and SJT performance. Oostrom et al. (2012) found that ITPs for conscientiousness and extraversion correlated more strongly with SJT performance ($r = .29$ and $r = .22$, respectively) than the traits alone (extraversion, $r = .17$; conscientiousness, $r = .17$), providing some support for partial mediation. This, along with findings that support the direct effect of personality on situational judgment test (SJT) performance (McDaniel et al., 2007; McDaniel & Nguyen, 2001) and personality on ITPs (Freudenstein et al., 2023; Miller et al., 2008; Motowidlo et al., 2006a, 2006b; Oostrom et al., 2012), leads to the following hypothesis:

Hypothesis 3: ITPs for conscientiousness, agreeableness, and emotional stability will

partially mediate the relationship between their associated traits and SJT performance.

Specific Job Experience, Specific Job Knowledge, and Situational Judgment Tests

Lievens and Motowidlo's (2016) model argues that SJTs capture procedural knowledge which includes components of both general domain knowledge and specific job knowledge. In this study, general domain knowledge was operationalized as ITPs. The archival dataset does not include a measure of specific job knowledge, but it does include job tenure, which was used as a proxy for specific job knowledge. Within their model, job experience (tenure) is a direct antecedent to specific job knowledge. Previous research provides support for the direct causal relationship between specific job experience and specific job knowledge, suggesting that tenure may be a reasonable proxy for specific job knowledge (Schmidt et al. 1986; Dokko et al. 2009; Wolcott et al. 2021). The type of specific job knowledge being considered in these studies (Schmidt et al. 1986; Dokko et al. 2009; Wolcott et al. 2021) is task or skill-related knowledge. Task or skill knowledge is related to one's ability to complete a specific task, such as putting together a type of report (Dokko et al., 2009). Given that other types of job knowledge demonstrate a positive relationship with job experience, it is reasonable to believe that specific job experience (or tenure) is related to specific job knowledge in a similar way to task or skill knowledge.

Additionally, previous research supports the relationship between SJT performance and prior job experience (Clevenger et al. 2001; McDaniel & Nguyen 2001; Motowidlo & Beier, 2010). Motowidlo and Beier (2010) created an SJT with two scoring keys: one developed by experts with extensive job knowledge, and another created by novices with no job knowledge. After tabulating the scores according to each scoring key, they found that while both expert and novice knowledge correlated with SJT performance ($r = .37$ and $r = .29$, respectively), expert

knowledge demonstrated a stronger relationship. Thus, they concluded that both general and specific job knowledge contribute independently to the variance in ITPs and SJT performance, but that specific job knowledge predicted SJT performance above and beyond general knowledge (Motowidlo & Beier, 2010). This suggests that when confronted with a fictional situation, as described in an SJT item, novice respondents may lean more heavily on their ITPs, whereas experts may tend to rely on their specific job knowledge when assessing the effectiveness of response options (Freudenstein et al. 2023).

The FA contains 40 video-based SJT items that are "saturated" with job-related information about firefighting. SJT theory has traditionally held that the more saturated a situation description is with job-specific information, the more candidates will rely upon their specific job knowledge to answer (Freudenstein et al., 2023), which should provide an advantage to those with more specific job knowledge. On SJTs that are saturated with job-specific information, candidates with greater specific job knowledge may increasingly rely on that knowledge when selecting SJT responses. With that in mind, this study anticipates that as reliance on specific job knowledge increases, dependency on general domain knowledge, or ITPs, will decrease.

Hypothesis 4: Among incumbents, specific job knowledge (tenure) will moderate the relationship between ITPs and SJT performance such that at higher levels of tenure, the strength of the relationship between ITPs and SJT performance will be diminished.

Figure 3

Dissertation Research Model
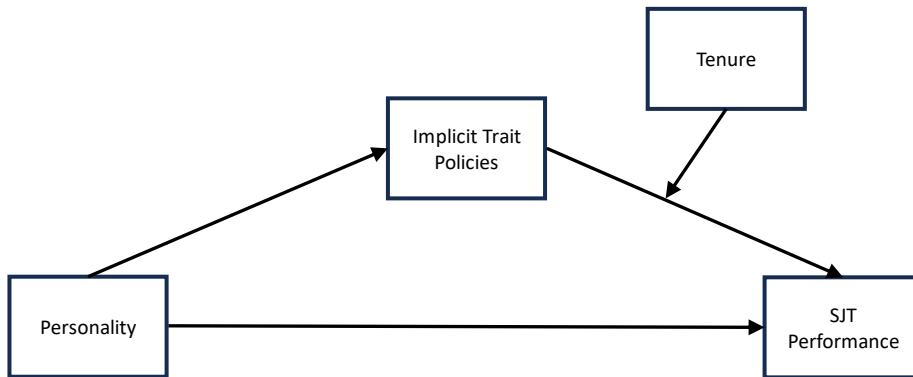
CHAPTER II

Method

This study utilized archival data gathered by the Company that developed and owns the Firefighter Assessment (FA). In the following section, information on the sampling procedure, participant characteristics, data collection, research design, methods, and analyses used to evaluate the hypotheses for this study is presented.

Sampling Procedure

Archival data were provided by the Company, which developed and validated the FA for use in the selection of entry-level firefighters. Two separate datasets were provided by the Company for this study. The first dataset, consisting of approximately 42,000 cases, included information gathered from firefighter candidates (or novices) who took the FA as part of their initial testing process to become firefighters. The second dataset included information gathered from incumbents who took the FA as part of the validation study for the tool. The incumbent dataset included 198 cases.

Participant Characteristics

The data provided by the Company was de-identified prior to being released for this study. Demographic information such as race and gender were provided and are presented in Table 1 for the combined dataset and broken down by novice and incumbent samples.

Table 1

Sample Demographics

Variable	Total Sample		Incumbent		Novice	
	n	%	n	%	n	%
Total Participants	32,122	100.00	197	100.00	31,925	100.00
Experience (Tenure)						
0 – 6 months	31,925	99.39	0	0.00	31,925	100.00
6 months – 1 year	10	< 1.00	10	5.08	0	0.00
1 – 3 years	32	< 1.00	32	16.24	0	0.00
3 – 5 years	30	< 1.00	30	15.23	0	0.00
5 – 10 years	42	< 1.00	42	21.32	0	0.00
10 years or more	83	< 1.00	83	42.13	0	0.00
Gender						
Female	3,100	9.58	15	7.60	3,085	9.66
Male	28,456	88.58	168	85.30	28,288	88.60
Did not disclose	568	1.82	15	7.61	553	1.74
Race/Ethnicity						
White	21,218	66.05	156	79.18	21,062	65.97
Hispanic	5,403	16.82	11	5.58	5,392	16.89
Black / African American	1,924	5.98	3	1.53	1,921	6.02
Other	1,763	5.49	18	9.14	1,745	5.47
Asian	1459	4.54	6	3.05	1,453	3.47
Native American	323	1.00	3	1.52	320	1.00
Hawaiian/Pacific Islander	379	1.18	0	0	379	1.19

Sample Size and Power

Consideration of the required sample size is important in this quantitative study to avoid

sampling errors. Therefore, four factors were considered in the power analysis: the significance

level, effect size, test power, and statistical technique. The significance level, also known as

Type I error, refers to the chance of rejecting a null hypothesis given that it is true (Field, 2018).

Most quantitative studies use a 95% confidence level, because this provides adequate statistical

evidence for testing (Field, 2012). The effect size refers to the estimated measurement of the

relationship between the variables considered (Cohen, 2013). The power of a test refers to the

probability of correctly rejecting the null hypothesis. In most quantitative studies, 80% power is

used (Shadish et al., 2002). This current study is a regression study that uses simple and hierarchical regressions to evaluate the hypotheses. In order to conduct hierarchical multiple regression that detects a medium effect size at the 5% level of significance with 80% power at least 92 participants were required for this study.

This study utilized archival data and did not require Institutional Review Board (IRB) approval. Although informed consent was not obtained from the participants, all data were de-identified. The Director of Research and Development employed by the Company provided the relevant de-identified data from the firm's database.

Measures

IPIP Big Five

The International Personality Item Pool (IPIP) is a public domain collection of personality items that can be used to develop personality assessment scales. The IPIP was developed as an alternative to proprietary personality measures with the goal of making personality assessment tools more accessible to researchers and practitioners. The IPIP contains a large number of items that measure various personality traits or dimensions.

The Company's Firefighter Assessment (FA) consists of two sections. The first section is comprised of 60 video-based items, 40 of which are SJTs. Candidates are required to complete all items in this section. The second section of the FA is a self-assessment comprised of 240 items that measure traits related to integrity, a dimension they developed and validated. 180 of those items remain consistent for all candidates, while the other 60 items vary from candidate to candidate. The IPIP scales for conscientiousness, agreeableness, and emotional stability were included in the section of varied items, meaning that each candidate has only has one IPIP dimension score. This did impact the sample size (n) for the analyses of hypothesis 1 and 3 as

both tested the relationship between IPIP traits and their associated ITPs. To test hypothesis 1,
that IPIP personality traits predict their associated ITPs, the sample size used to conduct the
analyses for each (IPIP-conscientiousness and ITP-conscientiousness, IPIP-agreeableness and
ITP-agreeableness, and IPIP-emotional stability and ITP-emotional stability) was n = 10,942. To
test hypothesis 3, that ITPs mediated the relationship between their associated IPIP personality
traits and SJT performance, the sample size was n = 10,942 for each trait. For the incumbent
dataset, IPIP conscientiousness and emotional stability were the included traits, and each
respondent was asked to complete items for both traits.

These variables are based on a 5-point Likert scale ranging from 1 (strongly disagree) to
5 (strongly agree). Greater values on these measures correspond to greater levels of
conscientiousness, agreeableness, and emotional stability. The scales for these three dimensions
can be found in Appendix D. According to the validation report provided by the Company, the
self-assessment integrity section of the FA demonstrated strong internal consistency (Cronbach's
α = .98). The incumbent dataset, used for the validation of the FA, only included scales for
conscientiousness and emotional stability, as the agreeableness dimension was added after the
validation. All three dimensions were included in the novice dataset.

Implicit Trait Policies

As discussed in the literature review, there are several methods for measuring and
calculating ITPs. Because this study operates within the confines of an archival dataset (i.e.,
there were no respondent effectiveness ratings), ITPs were estimated by calculating the
difference between the number of response options chosen to express high levels of a trait and
those chosen to express low levels of a trait (Motowidlo et al., 2006b).

To calculate the trait expressiveness of each response option, three test development SMEs from the Company rated all SJT response options according to their levels of trait expressiveness for conscientiousness, agreeableness, and emotional stability. The three raters received a brief frame of reference training (Melchers et al., 2011) where FFM personality traits were defined and examples of strong and weak trait expressiveness for each factor were provided. After viewing the situation description, the SMEs rated each response option for trait expression of conscientiousness, agreeableness, and emotional stability on a scale of -4 to 4, with -4 representing the lowest level of trait expression and 4 representing the highest level of trait expression. The reliability of the ratings was evaluated by calculating Intraclass Correlations (ICCs) using *ICC(3, k)*, which is similar to Cronbach's alpha (Schrout & Fleiss 1979). The ICC of the ratings was .85, demonstrating sufficient rater agreement.

Only response options that exceeded | 2 |, demonstrating high or low trait expressiveness, were included in the ITP calculations. The trait with the highest rating was then assigned to the response option. To reduce ambiguity, response options with two or more trait expressiveness ratings exceeding | 2 | were excluded from the calculations. This method was applied in a manner consistent with how Lambert and Wedell (1991) calculated scores for evaluative extremity. In their study, response options were eliminated if they ambiguously expressed a targeted trait. In the present study, each response option included in the ITP calculations was coded as a high or low expression of one of the focal traits. These ratings were used to create a trait-scoring key for each FFM dimension. Overall, 31 response options were included for conscientiousness (16 high, 15 low), 40 response options for agreeableness (19 high, 20 low), and 14 response options for emotional stability (7 high, 7 low). 85 response options were excluded for not clearly expressing a particular trait. The ITPs for each personality dimension were calculated as the

difference in the number of response choices that expressed high levels of a trait and the number of response options selected that expressed low levels of the same trait.

Situational Judgment Test

The dependent variable of the study was performance on 40 video SJT items from the Firefighters Assessment (FA). All SJT items provided short, videotaped vignettes of common interpersonal situations that firefighters encounter on the job (see Appendix A for an example video SJT). At the end of each video, candidates were asked how they would handle the hypothetical situation and were given four response options. Each response option provided a different course of action, with one correct answer for each situation. The scoring key was created by a panel of 66 subject matter experts (SMEs) who reviewed every question and response option for content validity, appropriateness of answers for applicants without experience, and the degree of good or poor judgment demonstrated by the answer. The video-based section of the test was scored as a percentage of the total possible points. This variable was measured at the interval level. The FA has 20 subscales such as (a) adaptability, (b) supervisory skills, and (c) industriousness (see Appendix C for the full list).

The FA was validated using a group of incumbent firefighters (n = 198) from four fire departments throughout the country. Demographics of the incumbents reasonably approximated the novice sample demographics (79% Caucasian, 85% male, 70% under the age of 40). The group was highly experienced, with 42% having more than 10 years of experience and 21% having between 5-10 years of experience. The measure demonstrated test/retest reliability ($r =$ .70) and the video SJT portion demonstrated strong internal consistency (Cronbach's $\alpha =$ .81). Performance on the video SJTs demonstrated a significant correlation with overall supervisory ratings ($r =$ .36), suggesting strong criterion-related validity.

Tenure (Specific Job Knowledge)

According to Lievens and Motowidlo's (2016) model, procedural knowledge is composed of an element of general domain knowledge (ITPs) and a component of specific job knowledge. In this model, job experience is an antecedent of job knowledge, but a measure of specific job knowledge for firefighters at this point does not exist. For this study, job tenure, or the number of years of job experience, was used as a proxy for specific job knowledge. Data on job tenure were collected for each participant in the validation study. Incumbents were asked to provide their years of experience by selecting from the answer choices that included (a) six months to one year, (b) one year to three years, (c) three years to five years, (d) five to ten years, and e) ten + years. For the analysis, tenure was coded as a factor with five levels. See Table 1 for levels of tenure among incumbents.

Research Design

This study employed a non-experimental regression design because there was no treatment, and the independent variables were not manipulated to establish cause-and-effect. Archival data were used to evaluate the relationships between constructs within the Lievens and Motowidlo (2016) model, which includes (a) the strength and nature of the relationship between ITPs and their associated personality traits, (b) the relationship between ITPs and SJT performance, and (c) the relationship between ITPs, tenure, and SJT performance. As such, a regression design is an appropriate choice for evaluating the relationship between IV and DVs and the mediating and moderating variables (Field, 2018).

Statistical Analyses

This study aimed to evaluate the relationships within the Lievens and Motowidlo (2016) model for knowledge determinants of procedural knowledge, including ITPs, personality traits, job tenure, and SJT scores. The Lavaan package in R Studio was used to conduct several regression analyses to evaluate the relationship between personality traits and their associated ITPs (Hypothesis 1) and the relationship between ITPs and SJT performance (Hypothesis 2). The Mediate package in R studio was used to evaluate ITPs as mediators of personality and SJT performance (Hypothesis 3). Finally, the Interactions package in R studio was used to evaluate tenure as a moderator of the relationship between ITPs and SJT performance (Hypothesis 4).

Sample Combination

Prior to hypothesis testing, several statistical tests were conducted to determine if combining the two samples (novice and incumbent) would be appropriate. First, a comparison of descriptive statistics and correlations between study variables was conducted. This was followed by chi-squares and t-tests to compare measures across samples.

The two samples were comparable in terms of gender but differed significantly in race. The incumbent sample was comprised of more White/Caucasian and Other/Undisclosed participants, and fewer Hispanic, Black/African American, Asian, and Hawaiian/Pacific Islander participants than the novice sample $\chi_2(4) = 50.80$, $p < .001$. Significant differences existed on the following study variables (See Table 2 below): SJT scores, ITP-conscientiousness, ITP-emotional stability, and ITP-agreeableness. The two samples were comparable in terms of IPIP-conscientiousness and IPIP-emotional stability. The incumbent sample did not include IPIP-agreeableness.

Considering these differences between the samples, supplemental analyses were conducted to test hypotheses 1-3 on the novice and incumbent samples separately. The results were consistent with those obtained on the combined sample, indicating there was no significant impact on the results. Additionally, the diversity in demographics (race) may increase the generalizability of the findings. Consequently, the samples were combined in order to test hypotheses 1-3. Testing for hypothesis 4 was conducted on only the incumbent sample as the novice dataset did not have a measure of tenure due to the fact that all novices have no tenure.

Table 2

Study Variable Differences

Variable	t-score	p value
SJT Scores	-6.07	<0.000
IPIP-conscientiousness	2.37	0.019
ITP-conscientiousness	-8.88	<0.000
IPIP-emotional stability	0.43	0.669
ITP-emotional stability	-8.29	<0.000
IPIP-agreeableness	-	-
ITP-agreeableness	-30.77	<0.000

*Novice sample is reference

Data Preparation

Two datasets were provided by the company for this study. One dataset contained a sample of incumbent firefighters (n = 198), gathered as part of the validation study for the Firefighter Assessment (FA). The other dataset contained a sample of novice firefighters (n = 42,334) who took the FA as a component of their entry-level testing process. Before hypothesis testing commenced, preliminary analyses were conducted for each dataset to evaluate missingness, and parametric assumptions were tested to assess the normality of the data (e.g., identifying outliers, skewness, and kurtosis). ITP scales and descriptive statistics, such as means and standard deviations, were calculated.

Missingness

The incumbent dataset contained 198 rows with 100% complete data. The novice dataset was evaluated for missingness in R Studio by using the Mice package to calculate the proportion of missingness at the item level. As expected, 100% of the rows in the novice dataset contained missing data for IPIP scales. 32,119 rows (75.25%) had missing agreeableness values, 31,935 rows (74.82%) had missing conscientiousness values, 32,075 rows (75.14%) had missing emotional stability values, and 74.79% were missing all three dimensions. Cases missing all IPIP values were excluded from the novice dataset, with the final dataset at n = 31,926.

Normality

The datasets were assessed for outliers and for normality. Within the incumbent dataset, the skewness and kurtosis values fell outside the typical range for a normal distribution, highlighting the non-symmetry and differing peak characteristics of the data. Specifically, the Shapiro-Wilk test for normality was conducted for each variable, resulting in p-values less than 0.05 for all variables. These findings collectively lead to the rejection of the null hypothesis of normality for all examined variables. Consequently, the non-normal distribution of these data warrants caution in the selection of subsequent statistical analyses and may necessitate the application of non-parametric methods or data transformation techniques to satisfy the assumptions of normality.

For the novice dataset, the skewness and kurtosis values provided insights into the asymmetry and tail behavior of the distributions. Skewness ranged from -0.98 (ITP-C) to -0.49 (ITP-A), indicating a tendency toward left-skewness in all variables. Kurtosis values varied from -0.10 (IPIP-ES) to 1.58 (ITP-C), with only IPIP-ES displaying a negative kurtosis. The Shapiro-Wilk test, which was conducted to assess normality, revealed significant departures from

normality for all variables ($p < 0.001$). However, it should be noted that the size of the dataset (n

= 42,685) exceeded the recommended threshold for the Shapiro-Wilk test, which may limit the

accuracy of the p-values. Overall, these findings suggest that the variables may not strictly

adhere to a normal distribution. However, because of the large sample size, normality at the item

and scale levels is often not required to conduct parametric analyses (Field, 2012).

CHAPTER III

Results

The purpose of this study was to test some of the hypothesized relationships within Lieven and Motowidlo's (2016) model of SJTs, which included personality, implicit trait policies, tenure, and situational judgment test (SJT) performance. This study employed regressions to test the hypothesized relationships within the model. The results of these analyses provided support for several relationships within the model.

Correlations

Sample descriptives can be found in Table 3 and Table 4 presents the bivariate correlations among the study variables. Consistent with previous research (Motowidlo et al. 2006a, 2006b), small and statistically significant relationships in the expected direction were found between personality traits (conscientiousness and emotional stability) and their associated ITPs (see Table 3). All IPIP dimensions demonstrated small and statistically significant relationship with SJT performance, which is consistent with previous findings (McDaniel et al. 2001; McDaniel & Nguyen, 2007). All ITPs demonstrated a moderate and statistically significant relationship with SJT performance. In both samples, all ITPs demonstrated stronger significant relationships with SJT performance than their associated personality traits. Experience (tenure) did not demonstrate significant relationships with any IPIP dimension but did demonstrate small but significant relationships with ITPs.

Table 3

Sample Descriptives

Variable	Mean	SD	Min	Max	Skew	Kurtosis	S-W *p* value
SJT Scores	82.51	8.45	22.78	100.00	-0.68	0.40	0.97 <0.000
IPIP-conscientiousness	4.63	0.40	1.00	5.00	-0.96	0.84	0.86 <0.000
ITP-conscientiousness	9.79	2.15	-7.00	14.00	-0.98	1.58	0.93 <0.000
IPIP-emotional stability	4.49	0.43	1.00	5.00	-0.59	-0.10	0.92 <0.000
ITP-emotional stability	6.40	1.25	-1.0	8.0	-0.74	0.45	0.90 <0.000
IPIP-agreeableness	4.62	0.36	1.00	5.00	-0.94	1.49	0.89 <0.000
ITP-agreeableness	5.48	1.30	3.0	10.0	-0.49	0.62	0.93 <0.000

Table 4

Sample Correlations

Variable	1	2	3	4	5	6	7
1. IPIP-conscientiousness							
2. IPIP-emotional stability	.29**						
3. IPIP-agreeableness	.06**	.04**					
4. ITP-conscientiousness	.04**	.04*	.03**				
5. ITP-emotional stability	.02**	.02	.02**	.29**			
6. ITP-agreeableness	.01	.02**	.02**	.17**	.08**		
7. Experience (Tenure)	-.02	.00	.00	.06**	.04**	.20**	
8. SJT Scores	.13**	.14**	.12**	.51**	.31**	.22**	.03**

Note. * Indicates $p < .05$. ** indicates $p < .01$.

The Effect of Personality Traits on Implicit Trait Policies

Overall, results supported Hypothesis 1 that personality traits (conscientiousness, agreeableness, and emotional stability) are positively related to their associated ITPs. ITPs for conscientiousness, emotional stability, and agreeableness were regressed on their associated IPIP dimensions using the base package in R Studio. The results are presented in Table 4. The results supported Hypothesis 1 that IPIP-conscientiousness predicted ITP-conscientiousness ($\beta = 0.50$, SE = 0.05, $t = 9.78$, $p < 0.001$), IPIP-emotional stability predicted ITP-emotional stability ($\beta = 0.22$, SE = 0.03, t = 7.81, $p < 0.001$), and IPIP-agreeableness predicted ITP-agreeableness ($\beta = 0.28$, SE = 0.03, t = 8.09, $p = 0.01$). These results replicate previous findings for the relationship

between ITPs for conscientiousness and agreeableness and extend the findings to include ITPs

for emotional stability. Though significant, each of the IPIP dimensions explained a small

amount of the variance in each of the associated ITPs ($R^2 = .01$). According to the model,

personality is one of the antecedents of ITPs in addition to others (such as experience and

cognitive ability). These results support the model for the relationship between personality and

ITPs and suggest that other antecedents, if added to the model, might account for additional

variance in the criterion.

Table 5

Hypothesis 1: Main Effect of Personality Traits on Their Associated ITPs

Predictor	β	SE	t	F	R^2	95% CI	
All				(1, 10941)			
(Intercept)	7.49***	0.24	31.23			7.02	7.96
conscientiousness	0.50***	0.05	9.68	93.69***	0.01	0.40	0.60
(Intercept)	5.45***	0.12	43.71			5.21	5.70
emotional stability	0.22***	0.03	7.81	60.98***	0.01	0.16	0.27
(Intercept)	4.17***	0.16	26.35			3.86	4.48
agreeableness	0.28***	0.03	8.09	64.54***	0.01	0.21	0.34

Note. * Indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

The Effect of Implicit Trait Policies on Situational Judgment Test Performance

Overall results supported Hypothesis 2 that ITPs for conscientiousness, agreeableness,

and emotional stability would predict SJT performance. SJT scores were regressed upon ITPs for

conscientiousness, emotional stability, and agreeableness using the base package in R Studio.

The results are presented in Table 5. The results supported Hypothesis 2 that ITP-

conscientiousness (β = 2.01, *SE* = .02, *t* = 123.1, *p* < 0.001), ITP-emotional stability (β = 2.06,

SE = 0.03, *t* = 66.64, *p* < 0.001), and ITP-agreeableness (β = 1.40, *SE* = 0.03, *t* = 45.56, p <

0.001) predicted SJT performance. The results replicated previous studies which found support for the relationship between ITP- conscientiousness and SJT performance and ITP-agreeableness and SJT performance. The same hypothesized relationship was found with ITP-emotional stability and SJT performance. ITPs for emotional stability ($R^2 = .09$) and agreeableness ($R^2 = .05$) explained very little variance in SJT performance. However, ITP-conscientiousness ($R^2 = .26$) did explain a greater amount of variance in SJT performance. There is evidence suggesting that conscientiousness is the easiest dimension to fake on self-report personality assessment, likely due to its well-known value to employers (Griffin et al., 2004).

Table 6

Hypothesis 2: Main Effect of ITPs on SJT Performance

Predictor	β	SE	t	F	R^2	95% CI	
All				(1, 42880)			
(Intercept)	62.84***	0.16	384.1			62.57	63.31
ITP-conscientiousness	2.01***	0.02	123.1	15140.00***	0.26	1.98	2.04
(Intercept)	69.30***	0.20	343.15			68.87	69.79
ITP-emotional stability	2.06***	0.03	66.64	4442.00***	0.09	1.99	2.04
(Intercept)	74.84***	0.17	432.93			74.48	75.26
ITP-agreeableness	1.40***	0.03	45.56	2076.00***	0.05	1.33	1.47

Note. * Indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$

The Indirect Effect of Personality on Situational Judgment Test Performance through Implicit Trait Policies

Overall results supported Hypothesis 3, predicting that ITPs for conscientiousness, agreeableness, and emotional stability would partially mediate the relationship between personality traits and SJT performance. Using the Mediate package in R Studio, a causal mediation analysis utilizing a quasi-Bayesian method was used to estimate the Average Causal

Mediation Effects (ACME), Average Direct Effects (ADE), Total Effect, and the Proportion

Mediated. The results are presented in Table 6. The results supported the hypothesis that ITP-

conscientiousness ($\beta = 0.93$, $p < 0.001$, 95% CI [0.72, 1.14]), ITP-emotional stability ($\beta = 0.40$,

$p < 0.001$, 95% CI [0.30, 0.52]), and ITP-agreeableness ($\beta= 0.36$, $p < 0.001$, 95% CI [0.26,

0.46]), partially mediate the relationship between their associated IPIP dimension and SJT

performance. These findings support the relationships within the Lievens and Motowidlo (2016)

model and suggest that personality does impact SJT through ITPs.

Table 7

Hypothesis 3: The Indirect Effect of Personality on SJT Performance Through ITPs

Mediator	Average Causal Mediation Effects	Average Direct Effects	Total Effect	Proportion Mediated
ITP-conscientiousness	0.93***	7.40***	8.32***	.11***
	[0.72, 1.14]	[7.03, 7.81]	[7.90, 8.80]	[0.09, 0.13]
ITP-emotional stability	0.40***	7.80***	8.20***	.0.05***
	[0.30, 0.52]	[7.47, 8.13]	[7.85, 8.56]	[0.04, 0.06]
ITP-agreeableness	0.36***	10.18***	10.54***	0.03***
	[0.26, 0.46]	[9.73, 10.67]	[10.06, 11.04]	[0.02, 0.04]

Note. * Indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

The Moderating Effect of Tenure

Overall results failed to support Hypothesis 4, predicting that among incumbents, job

tenure moderates the relationship between ITPs and SJT performance, such that at greater levels

of experience, the relationship between ITPs and SJT performance is weaker. Specifically, results

failed to support Hypothesis 4 for ITP-conscientiousness, ITP-emotional stability, and ITP-

agreeableness (Table 7). There was one significant interaction found between ITP-

conscientiousness and 3-5 years of experience ($\beta = 1.24$, t $=2.17$, $p = .05$). This indicates that

ITP-conscientiousness moderates the relationship between IPIP-conscientiousness and SJT

performance at 3-5 years of experience. This may be attributed to the larger number of

respondents (n = 41) at that level of tenure.

Table 8

Hypothesis 4: The Moderating Effect of Tenure on SJT Performance

Predictor	β	SE	t	R^2	F
ITP-conscientiousness				.307	$F(10, 184) = 9.58$***
(Intercept)	74.31***	5.76	12.89		
ITP-conscientiousness	1.11***	0.25	4.50		
*ITP-c * 6 months – 1 year*	0.18	0.61	0.29		
*ITP-c * Between 1 – 3 years*	0.43	0.48	0.89		
*ITP-c * Between 3 – 5 years*	1.24*	0.57	2.17		
*ITP-c * Between 5 – 10 years*	0.81	0.46	1.74		
ITP-emotional stability				.108	$F(10, 184) = 3.35$***
(Intercept)	78.99***	6.82	11.58		
ITP-emotional stability	1.23*	0.52	2.37		
*ITP-es * 6 months – 1 year*	2.94	2.27	1.29		
*ITP-es * Between 1 – 3 years*	0.70	1.25	0.56		
*ITP-es * Between 3 – 5 years*	-0.86	1.51	-0.59		
*ITP-es * Between 5 – 10 years*	1.57	0.94	1.67		
ITP-agreeableness				.262	$F(10, 184) = 6.55$***
(Intercept)	75.26***	7.39	10.17		
ITP-agreeableness	1.69*	0.67	2.51		
*ITP-a * 6 months – 1 year*	0.37	2.27	1.29		
*ITP-a * Between 1 – 3 years*	0.76	1.25	0.56		
*ITP-a * Between 3 – 5 years*	1.67	1.51	-0.59		

*Note: Significance Codes: * = .05, ** = .01, *** < .0001*

Figure 4

Moderating Effect of Tenure and ITP-conscientiousness
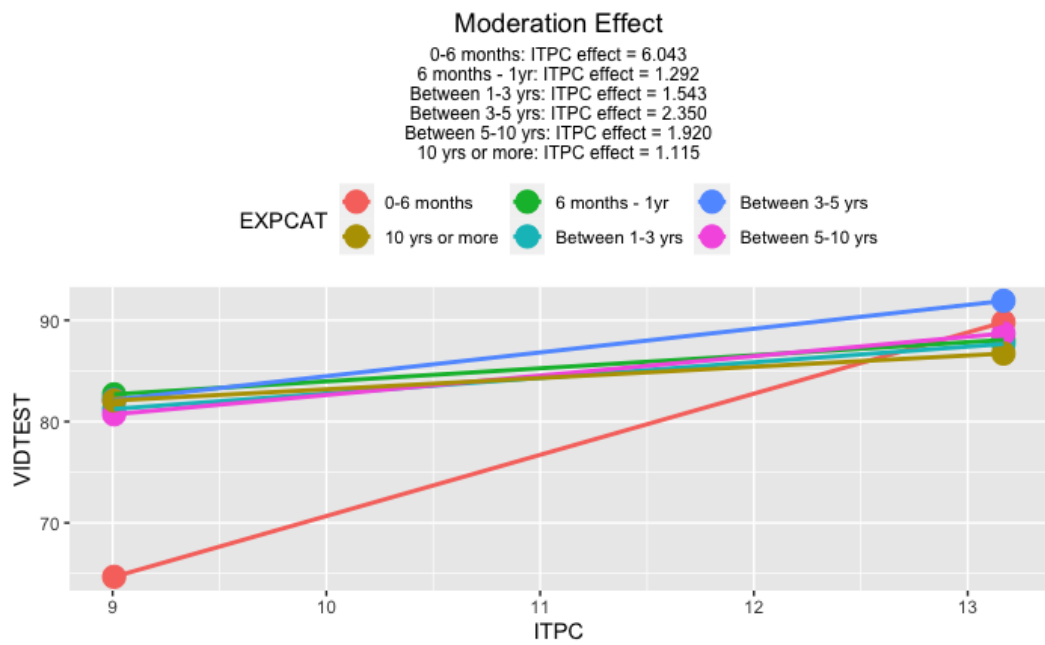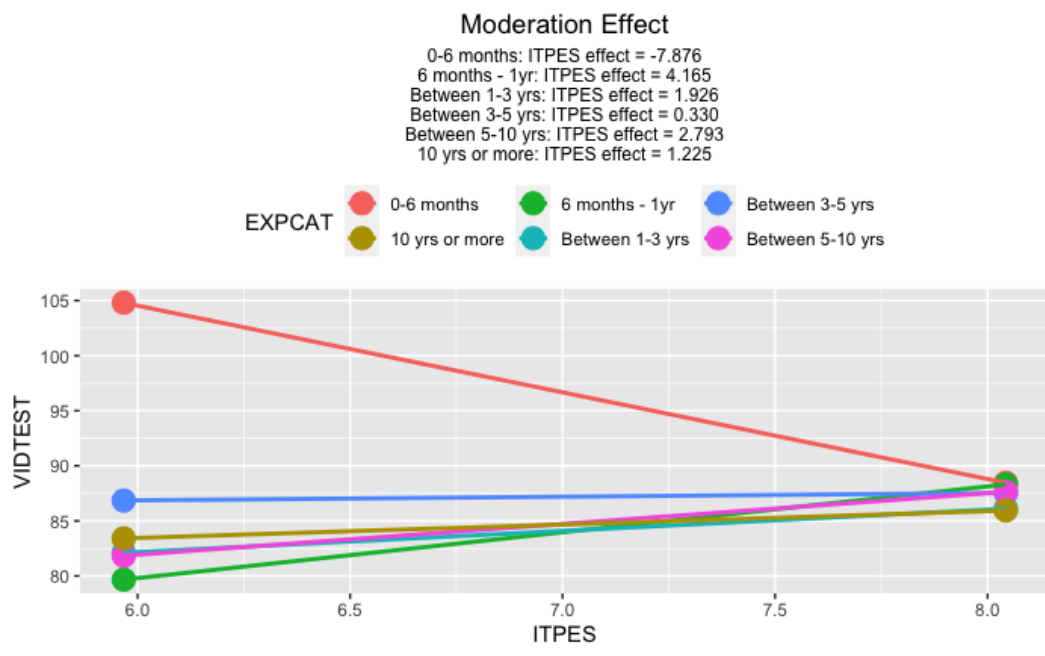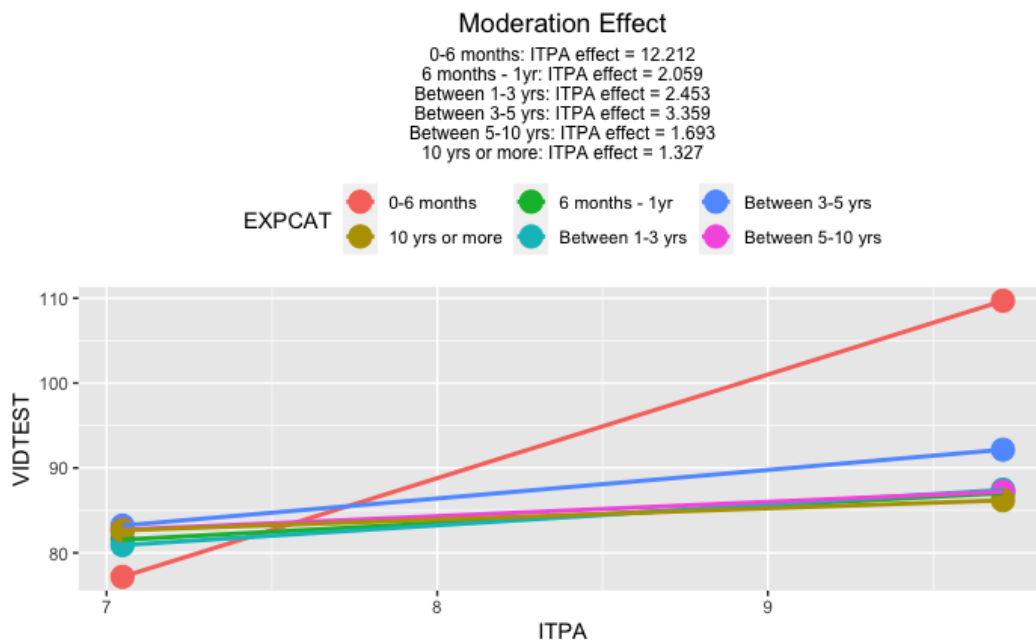

Figure 5

Moderating Effect of Tenure and ITP-emotional stability

Figure 6

Moderating Effect of Tenure and ITP-agreeableness

Moderation Effect

0-6 months: ITPA effect = 12.212
6 months - 1yr: ITPA effect = 2.059
Between 1-3 yrs: ITPA effect = 2.453
Between 3-5 yrs: ITPA effect = 3.359
Between 5-10 yrs: ITPA effect = 1.693
10 yrs or more: ITPA effect = 1.327

Conditional Effects. A conditional effects analysis was conducted to further investigate the effect of ITPs on SJT performance at the six levels of tenure. SJT scores were regressed on ITPs for conscientiousness, emotional stability, and agreeableness at each level of the moderator (tenure) to evaluate the predictive power of ITPs at different levels of tenure. The results are presented in Table 8. Significant effects were found at each level of the moderator for ITP-conscientiousness and ITP-agreeableness that indicate a curvilinear relationship between ITPs and SJT performance. As tenure increased, ITP-conscientiousness increased in its predictive power, peaking at 3-5 years of experience ($\beta = 2.35$, t = 4.58, *p < .05)* and then diminishing. ITP-agreeableness followed a similar curvilinear pattern, peaking at 3-5 years of experience ($\beta = 3.36$, t = 4.39, p *< .001),* and then diminishing. ITP-emotional stability did significantly predict SJT performance at 0-6 months of experience ($\beta = 1.23$, t = 2.37, *p = .05*), at 5-10 years of experience ($\beta = 2.79$, t = 3.56, *p < .001*), and at 10+ years of experience ($\beta = 1.23$, t = 2.37, *p =*

.05). These results indicate that general domain knowledge is most predictive of SJT

performance during the middle years of a firefighter's career.

Table 9

Conditional Effects of ITPs on SJT Performance at Each Level of Experience

	β	SE	*t*
ITP-conscientiousness			
0 – 6 months	1.11***	0.25	4.50
6 months – 1 year	1.29*	0.56	2.33
Between 1 – 3 years	1.54***	0.41	3.73
Between 3 – 5 years	2.35***	0.51	4.58
Between 5 – 10 years	1.92***	0.39	4.91
10 years or more	1.12***	0.25	4.49
ITP-emotional stability			
0 – 6 months	1.23*	0.52	2.37
6 months – 1 year	4.17	2.21	1.88
Between 1 – 3 years	1.93	1.13	1.70
Between 3 – 5 years	0.33	1.41	0.23
Between 5 – 10 years	2.79***	0.78	3.56
10 years or more	1.23*	0.52	2.37
ITP-agreeableness			
0 – 6 months	1.69*	0.68	2.51
6 months – 1 year	2.06*	0.93	2.21
Between 1 – 3 years	2.45***	0.64	3.84
Between 3 – 5 years	3.36***	0.76	4.39
Between 5 – 10 years	1.69*	0.68	2.51
10 years or more	1.33**	0.42	3.19

*Note: Significance Codes: * = .05, ** = .01, *** < .0001*

CHAPTER IV

Discussion

This study sought to examine the role of general domain knowledge (operationalized as ITPs), or the beliefs about what traits should be expressed in hypothetical scenarios, on SJT performance within a previously untested population of firefighters. Additionally, this study examined whether job tenure, a proxy for specific job knowledge, moderates the relationship between ITPs and SJT performance. This study also sought to expand on previous findings to evaluate the FFM trait of emotional stability and its associated ITP. Overall, the results from the study replicated previous findings that support the call to recast SJTs as a measure of general domain knowledge. General domain knowledge, operationalized as ITPs, for conscientiousness, emotional stability, and agreeableness each demonstrated significantly stronger relationships with SJT performance than any of their associated personality traits. This indicates that candidate personality traits may be less important to SJT performance than their knowledge of when it is most effective to express that trait.

According to ITP theory, personality is a direct antecedent of general domain knowledge. This study replicated previous findings supporting these relationships and found significant relationships between IPIP dimensions conscientiousness and agreeableness and their associated ITPs. Further, the study extended these findings to IPIP emotional stability and its associated ITP, providing additional support for one of the foundational relationships in ITP theory, that personality shapes our beliefs regarding effective workplace behavior. These findings align with previous research supporting the relationship between IPIP traits and their associated ITPs (Motowidlo et al., 2006a, 2006b; Oostrom et al., 2012).

Significant relationships were found in the expected direction between all ITPs and SJT performance aligning with previous research (Miller et al., 2008; Motowidlo et al., 2006a, 2006b; Motowidlo & Beier, 2010). By extending the relationship to ITP-emotional stability, additional support is provided for the theoretical relationship between ITPs and SJT performance. ITP-conscientiousness explained more variance in SJT scores than the other ITPs. Conscientiousness is a valued construct within personnel selection and is frequently included in the design of SJT scenarios and response options. Consequently, those who possess greater knowledge about effective expressions of conscientious behavior are likely to perform better on SJTs.

Additionally, this study found that ITPs for conscientiousness, emotional stability, and agreeableness partially mediated the relationship between their associated personality traits and SJT performance. Previous research has demonstrated the direct effects of personality and ITPs on SJT performance, but testing ITPs as mediators has not been done to the knowledge of the researcher. This supports the theoretical relationship in Lievens and Motowidlo's (2016) model, in which personality impacts SJT performance through ITPs. ITP-conscientiousness mediates a larger proportion than ITP-agreeableness and ITP-emotional stability, mirroring the pattern we find in the results from other studies and reinforcing the importance of conscientiousness in SJT performance.

The findings did not support tenure as a moderator in the relationship between ITPs and SJT performance among the incumbent sample. These findings may be the result of tenure being an ineffective proxy for job knowledge. However, the results of a conditional effects analysis indicated that ITPs for conscientiousness and agreeableness were found to increase in their predictive power as tenure increased, peaking at 3-5 years of experience, and then diminishing.

This suggests that general domain knowledge about effective expressions of conscientious and agreeable behavior plays a more significant role in predicting SJT performance as job experience increases, peaking during the midpoint of firefighter careers (i.e., when they are often promoted to frontline supervisors), and then diminishing in its predictive power during the latter part of a firefighter's careers. Internal job analyses conducted by the Company indicated that the ability to navigate interpersonal relationships is key to early career success. Specific job knowledge about effective behavior for firefighters is likely to increase during the early years of their career. As firefighters are promoted, the interpersonal demands of the job decrease dramatically, and the acquisition of task knowledge and administrative skills become the focus of their attention; consequently, they may rely upon their general knowledge to a lesser degree when selecting SJT response options.

Lievens and Motowidlo (2016) called for recasting SJTs as measures of general domain knowledge or implicit trait policies (ITPs). In their model, personality has a direct causal effect on general beliefs about what constitutes effective behavior in the workplace. The partially supported hypothesis for this relationship indicates that ITPs are a viable construct that is related to but distinct from personality and provides a viable explanation for the way personality impacts SJT performance. However, ITPs for emotional stability significantly predicted SJT performance, suggesting that one can acquire general knowledge about effectively expressing emotional stability without a high standing on the trait itself. Future research on ITPs and the unique experiences that shape them may provide additional value for explaining SJT performance.

Implications for Theory

The findings of this study make three contributions to SJT theory. First, the results provide support for relationships within the Lievens and Motowidlo (2016) model, finding significant relationships between personality, implicit trait policies, and SJT performance. The significant role that general domain knowledge plays in SJT performance supports their call for reframing SJTs as measures of general domain knowledge and indicates that further testing of their theoretical model is warranted.

Second, the results provide some support for a distinction between personality traits and general domain knowledge, meaning that one's standing on a personality trait and the knowledge of how to effectively express that trait are related but distinct. Personality traits were significant predictors of their related ITPs but did not explain a large amount of variance in ITPs. According to ITP theory, personality is one of several antecedents for ITPs and additional research may reveal that these other antecedents help explain additional variance in ITPs. This distinction may explain why ITPs demonstrate stronger relationships with SJT performance than do personality traits alone. For this reason, ITPs may actually measure something similar to emotional intelligence or adaptive behavior. The ability to adapt one's behavior as circumstances evolve is of great value to organizations, where change is constant. The mixed model of emotional intelligence conceptualizes emotional intelligence as both a set of capabilities (such as decision making, problem solving, and interpersonal relationships) and individual differences in personality traits (specifically extraversion, agreeableness, and conscientiousness) (Jun & Ryue, 2023). Future research is needed to explore a possible convergence between the measures of emotional intelligence and ITPs.

Third, the results of this study indicate that the relationship between ITPs and SJT performance may not depend on job tenure. Lievens and Motowidlo (2016) argued that SJTs capture respondents' knowledge of effective work behavior, some of which is learned through general socialization processes as well as experience on the job. For this study, specific job knowledge was operationalized as tenure based on the assumption that as job tenure increases, so should the knowledge of what constitutes effective behavior on the job. The lack of significant results may be explained in several ways. First, it could be due to tenure being an ineffective proxy of specific job knowledge. In this study, tenure was measured as a categorical rather than continuous variable. When a continuous variable is transformed to a categorical variable it can lead to a loss of information, reduce statistical power, and there is a potential for misleading interpretations (Field, 2012). Second, it could be that the knowledge required to successfully navigate interpersonal situations as a firefighter may be closely aligned with the knowledge required to successfully navigate interpersonal situations in general. Further studies conducted within organizational contexts, where unique expectations for interpersonal interactions require more distinct procedural knowledge, may provide insight into this situation.

And finally, the results may provide some insight into the way ITPs enable some candidates to perform better than others on SJTs. According to ITP theory, ITPs are general beliefs about effective behavior that reflect an individual's underlying personality traits. When an individual's beliefs about effective trait-related behavior align with the trait-related behavior required for success on the job, the candidate is thought to possesses the general domain knowledge that is required for success on the job (Motowidlo et al., 2006a). If an SJT is designed to measure this general domain knowledge, SJT scores should reflect individual differences of that knowledge. In other words, when a candidate generally believes behavior which expresses

emotional stability to be most effective (ITP-emotional stability), and emotional stability is required for success on the job, that candidate is thought to possess general domain knowledge about the effective expression of emotional stability at work and thus perform better on SJTs (Lievens & Motowidlo, 2016). The significant relationships found between ITPs and SJT performance in this study provide support for this explanation of how ITPs affect SJT performance.

Implications for Practice

This study has several implications for the practical use of SJTs for selection purposes. First, the findings of this study provide some support for Lievens and Motowidlo's (2016) call to recast SJTs as measures of general domain knowledge. ITPs account for a significant amount of variance in SJT performance and by focusing SJTs to measure ITPs, the validity may be enhanced. Consequently, developing and administering SJTs could be less resource intensive. Although SJTs are typically valued as cost-effective measures of job performance, organizations still invest significant resources in their development and administration. SJTs are developed with the help of SMEs, who provide specific job knowledge that is used to create situation descriptions. If their knowledge is not necessary for the development of valid SJTs, the costs of development may be reduced. Further, previous research has shown that scoring keys developed by panels without any experience in the job are as predictive as those developed by SMEs (Motowidlo & Beier, 2010). Instead of developing SJTs that utilize customized scenarios and scoring keys developed by organizational SMEs, generic or "off-the-shelf" SJTs could provide a viable and cost-effective alternative for organizations seeking to make hiring decisions.

Relatedly, by focusing SJTs on general domain knowledge, the lengths of selection assessments may be reduced. If ITPs predict SJT performance and do so more effectively than

personality traits alone, selection assessments may be designed with fewer self-report personality items or entire scales may be eliminated, thus reducing the length of the assessments. Shortened assessments can help reduce test-taker fatigue and the costs required for SJT development.

Second, as ITPs are implicit measures, they may help reframe the conversation about faking, a common issue with self-report personality items on employment assessments. In selection processes, candidates may be motivated by social desirability and rate themselves higher on traits they believe to be of importance to the employer making the hiring decision. The ability to identify the traits of value to an employer and then to effectively adapt one's responses to demonstrate higher levels of that trait, may be considered a type of faking. However, it may also be an indicator that a candidate has general domain knowledge about the effective expression of trait related behavior, reframing the faking conversation in positive terms. Novice firefighter scores were somewhat higher than incumbent firefighter scores on IPIP-conscientiousness ($M = 4.63$ and $M = 4.46$, respectively) and IPIP-emotional stability ($M = 4.50$ and $M = 4.47$, respectively), reflecting this pattern. However, only IPIP-conscientiousness ($t = 2.37, p = .02$) was significantly different, which may be the result of faking. This is worth noting, keeping in mind that several studies suggest faking does not have a significant impact on the predictive validity of personality traits (McFarland & Ryan, 2000; McFarland et al., 2001; Griffin et al., 2004).

Third, this study provides an alternative way to calculate ITPs utilizing SJTs that only have one correct response option and lack effectiveness ratings. This study calculated ITPs via a modified difference method. Though additional research is needed to further evaluate the validity of this method, the results of this study indicate this method may be a promising alternative for calculating ITPs.

Ultimately, Lievens and Motowidlo (2016) urged practitioners to be pragmatic and focus on designing SJTs specifically to measure general domain knowledge. They acknowledged the value of specific job knowledge, but argued that by focusing on general domain knowledge, the cost of development would decrease and potentially make for more valid SJTs. Decreased cost and increased SJT validity are of great value to organizations seeking to make hiring decisions in the most efficient manner possible.

Limitations and Future Research

This study has several limitations related to internal validity, construct validity, generalizability, and sample size. First, a primary concern is internal validity, which is the accuracy of a causal inference between variables. This study is a correlational study; consequently, no causal inferences can be made as there may be other explanations for the significant relationships found between ITPs, IPIP personality, and SJT performance. In addition, concerns about internal validity extend to the unusually high alpha (α =.98) for the self-report portion of the FA. This high of a score may be an indication of candidates in the novice sample faking good, specifically on trait conscientiousness, as it is widely known to be of great value to employers (Tett & Simonet, 2021). However, the ability to discern the importance of conscientiousness, and then knowing how to fake it on an assessment may explain why this study found ITPs to demonstrate stronger relationships with SJT performance than personality alone. In light of these concerns with internal validity, caution should be exercised when interpreting the results. Future studies that aim to identify potential confounding variables are warranted.

Second, there are issues with construct validity, which refers to the degree to which our measures accurately represent the concepts they are intended to capture. The focal construct of this study, general domain knowledge, operationalized as implicit trait policies, is very broad and

may not have fully encapsulated the construct it was meant to measure. ITPs are beliefs about effective behavior, and most studies calculate them as the correlation between candidates' effectiveness ratings and SMEs' effectiveness ratings. The SJTs used to calculate the ITPs did not ask respondents to rate the effectiveness of each response option and SMEs did not provide effectiveness ratings when developing the scoring key. In this study, ITPs were calculated as the difference between the number of response options selected that strongly expressed a trait from the number of response options selected that did not strongly express a trait. Although the difference method has been utilized previously and demonstrated more conservative ITP scores than those calculated via the correlation method (Motowidlo et al., 2006b), it may have negatively impacted the precision with which ITPs were measured.

Similar issues may arise with the operationalization of job knowledge with job tenure. This simplistic measure likely neglected the multidimensional nature of job knowledge and the quality of the experience itself. These issues with construct validity suggest caution in interpreting the results and highlight the necessity of employing more robust and multidimensional measures in future research to better capture the constructs underlying ITPs and provide more accurate insights into the relationships explored in this study.

A third limitation of this study pertains to the generalizability of the findings. The participants were drawn from a population of firefighters who were predominantly white and male, which does not reflect the broader population. These demographic constraints limit the extent to which the results can be generalized to other age groups or individuals outside the field of public safety. Furthermore, the cultural homogeneity of the sample, largely from a single racial background, raises concerns about the applicability of the findings to more diverse populations. These factors collectively suggest that caution should be exercised when attempting

to generalize the results of this study to a wider or more diverse population. Future research should replicate these findings within more diverse and representative samples to ascertain the broader applicability and generalizability of the observed relationships.

Finally, while a large sample size is generally beneficial for increasing the statistical power and reducing the risk of Type I errors (failing to detect a true effect), it also poses certain challenges. In particular, a large sample size can sometimes lead to statistically significant findings of little practical significance. Furthermore, contrary to typical scenarios, there may be instances where a large sample size could obscure meaningful differences within subgroups by diluting effects when data are aggregated. The implications of the large sample size should be considered when interpreting the findings of this study and future research may benefit from a more focused approach with well-defined, smaller samples to ensure that meaningful differences are not overlooked.

Conclusion

In the field of selection, the goal is to promote equal opportunities for applicants by creating assessment tools that accurately and fairly evaluate individuals' capabilities, minimize bias, and ensure fairness for all candidates. For a tool to be effective, it is imperative to know what construct(s) it measures and how it measures them; otherwise, the tool may unintentionally benefit some over others and companies may use them with less than satisfactory results. Situational Judgment Tests (SJTs) are widely used selection tools because of their predictive validity and small sub-group differences; however, researchers have struggled to explain why these tests are predictive. This study sought to advance the recasting of SJTs as a measure of general procedural knowledge or implicit trait policies. The results of the study provide support

for ITPs as a construct that is distinct from but related to personality traits and hold promise for their use in hiring decisions.

Personality profoundly shapes the way we experience and interpret the world. It may be tempting to think that these personality traits, for good and for ill, determine our destiny. However, the results of this study suggest that having a specific personality trait is less important to an individual's success than the individual's knowledge of how to effectively express that trait and when it is most appropriate to do so. It is not personality alone, but rather the ability to reflect, learn, and adapt, that determines one's future. ITPs are an attempt to quantify this valuable ability and it is hope that the results of this study will inspire additional research on this important topic.

References

Allport, G. W., & Odbert, H. S. (1936). Trait names: A psycho-lexical study. *Psychological Monographs, 47*, 1-171. https://doi.org/10.1037/h0093360

American Psychological Association (n.d.) Personality. In *APA Dictional of Psychology*. Retrieved February 18, 2023, from https://dictionary.apa.org/personality?_ga=2.193037293.1369040246.1676747390-815202718.1675628840

Arthur Jr, W., & Villado, A. J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, *93*(2), 435. https://doi.org/10.1037/0021-9010.93.2.435

Bagby, R. M., Bindseil, K. D., Schuller, D. R., Rector, N. A., Young, L. T., Cooke, R. G., Seeman, M. V., McCay, E. A., & T. Joffe, R. (1997). Relationship between the five-factor model of personality and unipolar, bipolar, and schizophrenic patients. *Psychiatry Research*, *70*(2), 83-94. https://doi.org/10.1016/s0165-1781(97)03096-5

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, *44*(1), 1-26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9(1)*, 9-30. https://doi.org/10.1111/1468-2389.00160

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & 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.

https://doi.org/10.1037/e518712013-343

Bledow, R., & Frese, M. (2009). A situational judgment test of personal initiative and its

relationship to performance. *Personnel Psychology*, *62*(2), 229-258.

https://doi.org/10.1111/j.1744-6570.2009.01137.x

Bornstein, R. F. (2015). Personality assessment in the diagnostic manuals: On mindfulness,

multiple methods, and test score discontinuities. *Journal of Personality*

Assessment, *97*(5), 446-455. https://doi.org/10.1080/00223891.2015.1027346

Bridgmon, K. D., & Martin, W. E. (2012). *Quantitative and statistical research methods: From*

hypothesis to results. John Wiley & Sons.

Brooner, R. K., Schmidt, C. W., Jr., & Herbst, J. H. (2002). Personality trait characteristics of

opioid abusers with and without comorbid personality disorders. In P. T. Costa, Jr. & T. A.

Widiger (Eds.), *Personality disorders and the five-factor model of personality* (pp. 249–

268). American Psychological Association. https://doi.org/10.1037/10423-015

Buss, D. M., & Barnes, M. (1986). Preferences in human mate selection. *Journal of Personality*

and Social Psychology, *50*(3), 559. https://doi.org/10.1037/0022-3514.50.3.559

Campion, M. C., & Ployhart, R. E. (2013). Assessing personality with situational judgment

measures. *Handbook of Personality at Work*, *439*.

https://doi.org/10.4324/9780203526910

Cattell, R. B. (1946). *The description and measurement of personality*. World Book.

https://doi.org/10.1037/h0057276

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,* 143-159.

https://doi.org/10.1037%2F0021-9010.82.1.143

Chan, D., & Schmitt, N. (2002). Situational judgment and job performance. *Human Performance, 15,* 233-254. https://doi.org/10.1207/s15327043hup1503_01

Christian, M. S., Edwards, B. D., & Bradley, J. C. (2010). Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology, 63*, 83-117. https://doi.org/10.1111/j.1744-6570.2009.01163.x

Crobnach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulleting, 52(4),* 281-302. https://doi.org/10.1037/h0040957

Church, A. T., Katigbak, M. S., Del Prado, A. M., Ortiz, F. A., Mastor, K. A., Harumi, Y., Tanaka-Matsumi, J., De Jesús Vargas-Flores, J., Ibáñez-Reyes, J., White, F. A., Miramontes, L. G., Reyes, J. A. S., & Cabrera, H. F.(2006). Implicit theories and self-perceptions of traitedness across cultures: Toward integration of cultural and trait psychology perspectives. *Journal of Cross-Cultural Psychology*, *37*(6), 694-716. https://doi.org/10.1177/0022022106292078

Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt-Harvey, V. (2001). Incremental validity of situational judgment tests. *Journal of Applied Psychology, 86,* 410-417. https://doi.org/10.1037/0021-9010.86.3.410

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences.* Erlbaum. https://doi.org/10.4324/9780203771587

Cook, T. D., Campbell, D. T., & Shadish, W. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin and Company.

Costa, P. T, & McCrae, R. R. (1985). *The NEO Personality Inventory manual*. Psychological

    Assessment Resources. https://doi.org/10.1007/978-1-4899-3588-5

Costa, P. T., Jr., & McCrae, R. R. (2008). The Revised NEO Personality Inventory (NEO-PI-R).

    In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of*

    Personality Theory and Assessment, Vol. 2. Personality Measurement and Testing. Sage

    Publications, Inc. https://doi.org/10.4135/9781849200479.n9

Debast, I., van Alphen, S. P., Rossi, G., Tummers, J. H., Bolwerk, N., Derksen, J. J., &

    Rosowsky, E. (2014). Personality traits and personality disorders in late middle and old

    age: Do they remain stable? A literature review. *Clinical Gerontologist*, *37*(3), 253-271.

    https://doi.org/10.1080/07317115.2014.885917

Dokko, G., Wilk, S. L., & Rothbard, N. P. (2009). Unpacking prior experience: How career

    history affects job performance. *Organization Science*, *20*(1), 51-68.

    https://doi.org/10.1287/orsc.1080.0357

Dunning, D., & Hayes, A. F. (1996). Evidence for egocentric comparison in social

    judgment. *Journal of Personality and Social psychology*, *71*(2), 213.

    https://doi.org/10.1037/0022-3514.71.2.213

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and

    purposive sampling. *American Journal of Theoretical and Applied Statistics*, *5*(1), 1-4.

    https://doi.org/10.11648/j.ajtas.20160501.11

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their

    meaning and use. *Annual Review of Psychology*, *54*(1), 297-327.

    https://doi.org/10.1146/annurev.psych.54.101601.145225

Fernandez, W. G., Benzer, J. K., Charns, M. P., & Burgess, J. F. (2020). Applying a model of

teamwork processes to emergency medical services. *Western Journal of Emergency Medicine*, *21*(6), 264-265. doi:10.5811/westjem.2020.7.47238

Field, A., Miles, J., Field, Z. (2012). Discovering Statistics Using R. SAGE Publications. https://doi.org/10.1111/insr.12011_21

Freudenstein, J. P., Mussel, P., & Krumm, S. (2023). On the Construct-Related Validity of Implicit Trait Policies. *European Journal of Personality*, *37*(1), 113-127. https://doi.org/10.1177/08902070211056901

Freudenstein, J. P., Schäpers, P., Roemer, L., Mussel, P., & Krumm, S. (2020). Is it all in the eye of the beholder? The importance of situation construal for situational judgment test performance. *Personnel Psychology*, *73*(4), 669-700. https://doi.org/10.1111/peps.12385

Gardner, A. K., & Dunkin, B. J. (2018). Evaluation of validity evidence for personality, emotional intelligence, and situational judgment tests to identify successful residents. *JAMA Surgery*, *153*(5), 409. https://doi.org/10.1001/jamasurg.2017.5013

Gatewood, R.D., Field, H., & Barrick, M. (2016) Human Resource Selection (8th edition). Thomson South-Western, Mason.

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216. https://doi.org/10.1037/0022-3514.59.6.1216

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84-96. https://doi.org/10.1016/j.jrp.2005.08.007

Heng, Y. T., Wagner, D. T., Barnes, C. M., & Guarana, C. L. (2018). Archival research: Expanding the methodological toolkit in social psychology. *Journal of Experimental Social Psychology*, *78*, 14-22. https://doi.org/10.1016/j.jesp.2018.04.012

Hofmann, D. A., & Jones, L. M. (2005). Leadership, collective personality, and performance. *Journal of Applied Psychology*, *90*(3), 509. https://doi.org/10.1037/0021-9010.90.3.509

Hovland, C. I., & Sherif, M. (1952). Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *The Journal of Abnormal and Social Psychology*, *47*(4), 822. https://doi.org/10.1037/h0056372

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96(1), 72-98. https://doi.org/10.1037/0033-2909.96.1.72

Jackson, D. J. R., LoPilato, A. C., Hughes, D., Guenole, N., & Shalfrooshan, A. (2017). The internal structure of situational judgment tests reflects candidate main effects: Not dimensions or situations. *Journal of Occupational and Organizational Psychology, 90,* 1-27. https://doi.org/10.1111/joop.12151

Judge, T. A., Heller, D., & Klinger, R. (2008). The dispositional sources of job satisfaction: A comparative test. *Applied Psychology*, *57*(3), 361-372. https://doi.org/10.1111/j.1464-0597.2007.00318.x

Judge, T. A., Klinger, R., Simon, L. S., & Yang, I. W. F. (2008). The contributions of personality to organizational behavior and psychology: Findings, criticisms, and future research directions. *Social and Personality Psychology Compass*, *2*(5), 1982-2000. https://doi.org/10.1111/j.1751-9004.2008.00136.x

Judge, T. A., & Piccolo, R. F. (2004). Transformational and transactional leadership: a meta-analytic test of their relative validity. *Journal of Applied Psychology*, *89*(5), 755. https://doi.org/10.1037/0021-9010.89.5.755

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875. https://doi.org/10.1037/a0033901

Judge, T. A., Van Vianen, A. E., & De Pater, I. E. (2004). Emotional stability, core self-evaluations, and job outcomes: A review of the evidence and an agenda for future research. *Human Performance*, *17*(3), 325-346. https://doi.org/10.1207/s15327043hup1703_4

Jun, S. K., & Hee Ryue, S. (2023). Emotional intelligence of korean students and its recent research trends. In S. Abed Hosseini (Ed.), Emotion recognition: Recent advances, new perspectives and applications. IntechOpen. doi: 10.5772/intechopen.110702

Kerr, N. L. (1998) HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review, 2*, 196–217. https://doi.org/10.1207/s15327957pspr0203_4

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology,* 50,537-567. https://doi.org/10.1146/annurev.psych.50.1.537

Krueger, R. F., & Johnson, W. (2021). Behavioral genetics and personality: Ongoing efforts to integrate nature and nurture. In O. P. John & R. W. Robins (Eds.), *Handbook of personality: Theory and research* (pp. 217–241). The Guilford Press.

Krumm, S., Lievens, F., Hüffmeier, J., Lipnevich, A. A., Bendels, H., & Hertel, G. (2015). How "situational" is judgment in situational judgment tests? *Journal of Applied*

Psychology, *100*(2), 399–417. https://doi.org/10.1037/a0037674

Lambert, A. J., & Wedell, D. H. (1991). The self and social judgment: Effects of affective

reaction and "own position" on judgments of unambiguous and ambiguous information

about others. *Journal of Personality and Social Psychology, 61*(6), 884–

897. https://doi.org/10.1037/0022-3514.61.6.884

Lamers, S., Westerhof, G., Kovacs, V., & Bohlmeijer, E. (2012). Differential relationships in the

association of the Big Five personality traits with positive mental health and

psychopathology. *Journal of Research in Personality, 46(*5), 517-524.

https://doi.org/10.1016/j.jrp.2012.05.012

Landy, F.J. (2007) The validation of personnel decisions in the twenty-first century: Back to the

future. In S.M. Phail (Ed.), *Alternate validation strategies: Developing and leveraging*

existing validity evidence, (1st ed,. pp.409-426). Wiley.

Lievens, F., & De Soete, B. (2012). Simulations. In N. Schmitt (Ed.) *Handbook of Assessment*

and Selection. Oxford University Press. 383–410.

https://doi.org/10.1093/oxfordhb/9780199732579.013.0017

Lievens, F., Harris, M. M., Van Keer, E., & Bisqueret, C. (2003). Predicting cross-cultural

training performance: The validity of personality, cognitive ability, and dimensions

measured by an assessment center and a behavior description interview. *Journal of*

Applied Psychology, *88*(3), 476–489. https://doi.org/10.1037/0021-9010.88.3.476

Lievens, F., & Motowidlo, S. J. (2016). Situational judgment tests: From measures of situational

judgment to measures of general domain knowledge. *Industrial and Organizational*

Psychology, 9, 3-22. https://doi.org/10.1017/iop.2015.71

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests,

    low-fidelity simulations, and high-fidelity simulations for predicting job performance in

    advanced-level high-stakes selection. *Journal of Applied Psychology*, *96*(5), 927-940.

    https://doi.org/10.1037/a0023496

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.

    https://doi.org/10.1037/0021-9010.91.5.1181

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. https://doi.org/10.1037/a0025741

Lievens, F., Schäpers, P., & Herde, C. N. (2020). Situational judgment tests: From low-fidelity

    simulations to alternative measures of personality and the person-situation interplay.

    Measuring and Modeling Persons and Situations, 285–311. https://doi.org/10.1016/b978-

    0-12-819200-9.00017-x

Lucas, R. E., & Donnellan, M. B. (2009). Age differences in personality: evidence from a

    nationally representative Australian sample. *Developmental Psychology*, *45*(5), 1353.

    https://doi.org/10.1037/a0013914

Malouff, J. M., Thorsteinsson, E. B., & Schutte, N. S. (2005). The relationship between the five-

    factor model of personality and symptoms of clinical disorders: A meta-analysis. *Journal

    of Psychopathology and Behavioral Assessment*, *27*, 101-114.

    https://doi.org/10.1007/s10862-005-5384-y

Matthews, G., Reinerman-Jones, L. E., Burke, C. S., Teo, G. W., & Scribner, D. R. (2018).

    Nationalism, personality, and decision-making: Evidence from an SJT for military multi-

national teaming scenarios. *Personality and Individual Differences*, *127*, 89-100.

https://doi.org/10.1016/j.paid.2018.01.045

Matthews, G., Reinerman-Jones, L. E., Burke, C. S., Teo, G. W., & Scribner, D. R. (2018).

Nationalism, personality, and decision-making: Evidence from an SJT for military multi-

national teaming scenarios. *Personality and Individual Differences*, *127*, 89-100.

https://doi.org/10.1016/j.paid.2018.01.045

McAdams, D. P., & Olson, B. D. (2010). Personality development: Continuity and change over

the life course. *Annual Review of Psychology*, *61*, 517-542.

https://doi.org/10.1146/annurev.psych.093008.100507

McCrae. R. R., & Costa. P. T., (2008). Empirical and theoretical status of the Five-Factor Model

of personality traits, in G. J. Boyle, Matthews, G., & D. H. Saklofske (eds.), *Personality*

Theory and Assessment: Volume 1 Personality Theories and Models. Thousand Oaks,

CA: SAGE Publications, Inc.

McDaniel, M. A., Bruhn Finnegan, E., Morgeson, F. P., & Campion, M. A. (2001). Use of

situational judgment tests to predict job performance: A clarification of the literature.

Journal of Applied Psychology, 86, 730-740. https//doi.org/10.1037//0021-9010.86.4.730

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. L. (2007). Situational judgment

tests, response instructions and validity: A meta-analysis. *Personnel Psychology, 60*, 63-

91. https://doi.org/10.1111/j.1744-6570.2007.00065.x

McDaniel, M. A., List, S. K., & Kepes, S. (2016). The "hot mess" of situational judgment test

construct validity and other issues. *Industrial and Organizational Psychology*, *9*(1), 47-

51. https://doi.org/10.1017/iop.2015.115

McDaniel, M. & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and

    constructs assessed. *International Journal of Selection and Assessment*, *9*(1-2), 103-113.

    http://dx.doi.org/10.1111/1468-2389.00167

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of

    employment interviews: A comprehensive review and meta-analysis. *Journal of Applied*

    Psychology, *79*(4), 599. https://doi.org/10.1037/0021-9010.79.4.599

Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure

    really better? A comparison of frame-of-reference training and descriptively anchored

    rating scales to improve interviewers'rating quality. *Personnel Psychology*, *64*(1), 53-87.

    https://doi.org/10.1111/j.1744-6570.2010.01202.x

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N.

    (2007). Are we getting fooled again? Coming to terms with limitations in the use of

    personality tests for personnel selection. *Personnel Psychology*, *60*(4), 1029-1049.

    https://doi.org/10.1111/j.1744-6570.2007.00100.x

Motowidlo, S. J. (2003). Job performance. *Handbook of Psychology: Industrial and*

    Organizational Psychology, *12*(4), 39-53. https://doi.org/10.4135/9781848608320.n7

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.

    https://doi.org/10.1037/0021-9010.75.6.640

Motowidlo, S. J., Ghosh, K., Mendoza, A. M., Buchanan, A. E., & Lerma, M. N. (2016). A

    context-independent situational judgment test to measure prosocial implicit trait

    policy. *Human Performance*, *29*(4), 331-346.

    https://doi.org/10.1080/08959285.2016.1165227

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006a). A theoretical basis for situational

    judgment tests. In J. A. Weekley & R. E. Ployhart (Eds.), *Situational judgment tests* (1st

    ed., pp. 72-94). Erlbaum. https://doi.org/10.4324/9780203774878-12

Motowidlo, S. J., Hooper, A. C., & Jackson, H. L. (2006b). Implicit policies about relations

    between personality traits and behavioral effectiveness in situational judgment

    items. *Journal of Applied Psychology*, *91*(4), 749 -761. https://doi.org/10.1037/0021-

    9010.91.4.749

Motowidlo, S. J., Lievens, F., & Ghosh, K. (2018). Prosocial implicit trait policies underlie

    performance on different situational judgment tests with interpersonal content. *Human*

    Performance, *31*(4), 238-254. https://doi.org/10.1080/08959285.2018.1523909

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,*

    337-344. https://doi.org/10.1111/j.2044-8325.1993.tb00543.x

Neal, A., Yeo, G., Koy, A., & Xiao, T. (2012). Predicting the form and direction of work role

    performance from the Big 5 model of personality traits. *Journal of Organizational*

    Behavior, *33*(2), 175-192. https://doi.org/10.1002/job.742

Ng, T. W., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). Predictors of objective and

    subjective career success: A meta-analysis. *Personnel Psychology*, *58*(2), 367-408.

    https://doi.org/10.1111/j.1744-6570.2005.00515.x

Olaru, G., Burrus, J., MacCann, C., Zaromb, F. M., Wilhelm, O., & Roberts, R. D. (2019).

    Situational judgment tests as a method for measuring personality: Development and

    validity evidence for a test of dependability. *PLOS ONE*, *14*(2).

    https://doi.org/10.1371/journal.pone.0211884

Oostrom, J. K., de Vries, R. E., & de Wit, M. (2018). Development and validation of a HEXACO situational judgment test. *Human Performance*, *32*(1), 1–29. https://doi.org/10.1080/08959285.2018.1539856

Patterson, F., & Driver, R. (2018). Situational judgment tests (sjts). *Selection and Recruitment in the Healthcare Professions*, 79–112. https://doi.org/10.1007/978-3-319-94971-0_4

Patterson, F., Ashworth, V., Kerrin, M., & O'Neill, P. (2013). Situational judgment tests represent a measurement method and can be designed to minimize coaching effects. *Medical Education*, *47*(2), 220-221. https://doi.org/ 10.1111/medu.12100

Ponterotto, J. G. (2005). Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of Counseling Psychology*, *52*(2), 126. https://doi.org/10.1037/0022-0167.52.2.126

Ployhart, R. E. (2006). The predictor response process model. In J. A. Weekley, & R. In E. Ployhart & Weekley (Eds.), *Situational judgment tests: Theory, measurement and application* (1st ed., pp. 157-182). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203774878-12

Reiser, S., Schacht, L., Thomm, E., Figalist, C., Janssen, L., Schick, K., ... & Bauer, J. (2022). A video-based situational judgment test of medical students' communication competence in patient encounters: Development and first evaluation. *Patient Education and Counseling*, *105*(5), 1283-1289. https://doi.org/10.1016/j.pec.2021.08.020

Sackett, P. R., Zhang, C., Berry, C. M., & Lievens, F. (2021). Revisiting meta-analytic estimates of validity in personnel selection: Addressing systematic overcorrection for restriction of range. *Journal of Applied Psychology, 107*(11), 2040–2068. https://doi.org/10.1037/apl0000994

Salgado, J. F. (2003). Predicting job performance using FFM and non-FFM personality measures. *Journal of Occupational and Organizational Psychology*, *76*(3), 323-346. https://doi.org/10.1348/096317903769647201

Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality*, *69*(6), 847-879. https://doi.org/10.1111/1467-6494.696167

Saucier, G., & Srivastava, S. (2015). What makes a good structural model of personality? Evaluating the big five and alternatives. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *APA handbook of personality and social psychology: Vol. 4. Personality processes and individual differences* (pp. 177-204). American Psychological Association. https://doi.org/10.1037/14343-001

Serfass, D. G., & Sherman, R. A. (2013). Personality and perceptions of situations from the Thematic Apperception Test. *Journal of Research in Personality*, *47*(6), 708-718. https://doi.org/10.1016/j.jrp.2013.06.007

Schäpers, P., Freudenstein, J.-P., Mussel, P., Lievens, F., & Krumm, S. (2020). Effects of situation descriptions on the construct-related validity of construct-driven situational judgment tests. *Journal of Research in Personality*, *87*, 103963. https://doi.org/10.1016/j.jrp.2020.103963

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, *71*(3), 432-457. https://doi.org/10.1037/0021-9010.71.3.432

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420. https://doi.org/10.1037/0033-2909.86.2.420

Singer, M., & Bruhns, C. (1991). Relative effect of applicant work experience and academic qualification on selection interview decisions : a study of between-sample generalizability. *Journal of Applied Psychology*, 76, 550-559. https://doi.org/10.1037/0021-9010.76.4.550.

Smithikrai, C. (2008). Moderating effect of situational strength on the relationship between personality traits and counterproductive work behavior. *Asian Journal of Social Psychology*, *11*(4), 253-263. https://doi.org/10.1111/j.1467-839X.2008.00265.x

Takacs, E. B., & Tracy, C. R. (2022). Evaluating the whole applicant: use of situational judgment testing and personality testing to address disparities in resident selection. *Current Urology Reports*, *23*(11), 309-318. https://doi.org/10.1016/j.paid.2020.110178

Takacs, E. B., & Tracy, C. R. (2022). Evaluating the whole applicant: Use of situational judgment testing and personality testing to address disparities in resident selection. *Current Urology Reports*, 1-10. https://doi.org/10.1007/s11934-022-01115-8

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, *88*(3), 500. https://doi.org/10.1037/0021-9010.88.3.500

Tett, R. P., & Guterman, H. A. (2000). Situation trait relevance, trait expression, and cross-situational consistency: Testing a principle of trait activation. *Journal of Research in Personality*, *34*(4), 397-423. https://doi.org/10.1006/jrpe.2000.2292

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

    Christiansen & R. Tett (Eds.), *Handbook of personality at work* (1st ed., pp.71-100).

    Routledge. https://doi.org/10.4324/9780203526910

Tett, R., & Simonet, D. (2021). Applicant faking on personality tests: Good or bad and why

    should we care?. *Personnel Assessment and Decisions*, *7*(1), 2, 6-19.

    https://doi.org/10.25035/pad.2021.01.002

Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait activation theory: A review of the literature

    and applications to five lines of personality dynamics research. *Annual Review of*

    Organizational Psychology and Organizational Behavior, *8*, 199-233.

    https://doi.org/10.1146/annurev-orgpsych-012420-062228

Trull, T. J., & Widiger, T. A. (2013). Dimensional models of personality: The five-factor model

    and the DSM-5. *Dialogues in Clinical Neuroscience*, 15(2), 135–146.

    https://doi.org/10.31887/dcns.2013.15.2/ttrull

Walton, K. E., Burrus, J., Murano, D., Anguiano-Carrasco, C., Way, J., & Roberts, R. D. (2022).

    A Big Five-based multimethod social and emotional skills assessment: The Mosaic™ by

    ACT® Social Emotional Learning Assessment. *Journal of Intelligence*, *10*(4), 72-73.

    https://doi.org/10.3390/jintelligence10040072

Weekley. J. A., & Jones, C. (1999). Further studies of situational tests. *Personnel Psychology,*

    50, 679-700. https://doi.org/10.1111/j.1744-6570.1999.tb00176.x

Weekley, J. A., Ployhart, R. E., & Holtz, B. C. (2006). On the development of situational

    judgment tests: Issues in item development, scaling, and scoring. In J. A. Weekley, & R.

E. Ployhart & Weekley (Eds.), *Situational judgment tests: Theory, measurement and application* (1ˢᵗ ed., pp. 117-149). Lawrence Erlbaum Associates. https://doi.org/10.4324/9780203774878-12

Weng, Q. D., Yang, H., Lievens, F., & McDaniel, M. A. (2018). Optimizing the validity of situational judgment tests: The importance of scoring methods. *Journal of Vocational Behavior*, *104*, 199-209. https://doi.org/10.1016/j.jvb.2017.11.005

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. *Journal of Applied Psychology*, *52*(5), 372. http://dx.doi.org/10.1037/h0026244

Wolcott, M. D., Lobczowski, N. G., Zeeman, J. M., & McLaughlin, J. E. (2020). Situational judgment test validity: An exploratory model of the participant response process using cognitive and think-aloud interviews. *BMC Medical Education*, *20*(1), 1-12. https://doi.org/10.1186/s12909-020-02410-z

Appendix A

Sample Video Situational Judgment Test Item – Video Script

Scene: Two firefighters standing around. Supervisor walks up.

Supervisor:	Supervisor: Hey guys. I need one of you to head over to District 2 for the rest of your shift. They are dealing with staffing issues right now and I need you to help them by filling in.
Firefighter 1:	*(uncertain)* District 2? I've been there once and didn't get along with the lieutenant at all. I bet he's the reason for the staffing issues. No one wants to work with him.
Supervisor:	Listen, I don't care about the lieutenant. They need help and given your knowledge of the crew, why don't you go grab your gear and go.
Firefighter 2:	*(a little sarcastic)* Good luck….haha.

Question: If you had a bad experience with the lieutenant at that station, and you were the one being told to go there, how would you respond?

    a. Say you can't accept the assignment because you cannot work for that lieutenant.
    b. Suggest they send the other firefighter.
    c. Tell the supervisor you need more detailed information about what you will be doing and then decide whether to accept the assignment.
    d. Accept the assignment and do your best to maintain a positive attitude.

Appendix B

Sample SJT Items with and without Situation Descriptions

	SJT item with situation description	SJT item without situation description	SJT item without situation description and situation-neutral responses
Situation description	A new computer program was installed in your department. No detailed training was provided to save time and money. Some of your colleagues and you feel insecure in dealing with this new program. Errors frequently happen which leads to a loss of time.	*(omitted)*	*(omitted)*
Response instructions	What would you do?	What would you do?	What would you do?
Response options	a) I organize an internal training in which more experienced colleagues share their knowledge. (*correct answer*)	a) I organize an internal training in which more experienced colleagues share their knowledge.	a) I organize internal trainings in which more experienced colleagues share their knowledge.
	b) I accept working overtime if I have to correct some of the errors	b) I accept working overtime if I have to correct some of the errors	b) I accept working overtime if I have to correct errors.
	c) I read books to understand the computer program in my free time to avoid time-consuming errors.	c) I read books to understand the computer program in my free time to avoid time-consuming errors.	c) I read professional books in my free time to avoid time-consuming errors at work.
	d) I don't get upset about it because with more practice I will stop making errors.	d) I don't get upset about it because with more practice I will stop making errors.	d) I am not overly worried about errors at work because with more practice I will stop making errors.

Note. Example item taken from an SJT on personal initiative (Bledow & Frese, 2009).

Appendix C

Firefighter Assessment Subscales

---

1. Considers Coworkers Before Self
2. Congenial
3. Leadership
4. Industrious/Hardworking
5. Adaptability
6. Customer Relations
7. Communication
8. Compassion
9. Dealing with Difficult Situations
10. Integrity/Ethics
11. Supervisory/Management Support
12. Persistence in Ongoing Learning
13. Safety Orientation
14. Reading Comprehension
15. Mechanical Reasoning
16. Math
17. Written Communication
18. Emergency Medical Knowledge
19. Physical Skill and Ability
20. Overall Performance

Appendix D

IPIP Items by Dimension

Conscientiousness

1 I don't quit a work task before it is finished.
2 I am a goal-oriented employee.
3 I finish things at work, despite obstacles in the way.
4 I am a hard worker.
5 I don't finish what I start.
6 I do not tend to stick with what I decide to do at work.
7 I accomplish my work on time.
8 I am careful to avoid making mistakes at work.
9 I forget to put things back in their proper place at work.
10 I neglect my duties at work.
11 I leave my work undone.
12 I put off unpleasant work tasks.
13 I am often late to work.
14 I like to organize things at work.
15 I am exacting in my work.
16 When my work shift begins, I start working right away.
17 I waste time at work.
18 I disregard work rules.
19 I avoid responsibilities at work.
20 I pay attention to details at work.
21 I make plans at work and stick to them.
22 I follow through with my work plans.
23 I finish what I start at work.
24 I have difficulty starting work tasks.

Neuroticism (or Emotional Stability)

1 I seldom get mad at work.
2 I am not easily bothered by things at work.
3 I am not easily frustrated at work.
4 I get stressed out easily at work.
5 I get upset easily at work.
6 I am easily disturbed at work.
7 I rarely get irritated at work.
8 I am not easily annoyed at work.
9 I remain calm under pressure at work.
10 At work, I do things I later regret.
11 I am calm at work, even in tense situations.

12   I panic easily at work.

13   At work, my feelings are easily hurt.

14   I feel crushed by setbacks at work.

15   I become overwhelmed by events at work.

16   I keep my emotions under control at work.

17   I react intensely to things at work.

18   I say whatever comes into my mind at work.

19   I can't stand being contradicted at work.

20   I adjust easily to what is happening at work.

21   I get upset if others change the way that I have arranged things at work.

22   I am annoyed by others' mistakes at work.

23   I am hard to reason with at work.

Agreeableness

1   I respect authority at work.

2   I do not believe that I am better than others I work with.

3   At work, I find it best to focus on my own personal gain.

4   When I'm at work, I want people to finish what they have to say before I reply.

5   I try to refrain from behavior that might disturb my coworkers' concentration.

6   I try to refrain from behavior that is annoying to coworkers.

7   I try to refrain from behavior that is offensive to coworkers.

8   I am concerned about how my behavior impacts coworkers.

9   I am concerned about my coworkers' welfare.

10   When a coworker is offended by my behavior, I apologize and try to change.

11   I behave how I want at work, whether people like it or not.

12   I appreciate good manners at work.

13   I keep my promises at work.

14   I am good at taking advice on how to do things at work.

15   When interacting with a group of coworkers, I am often annoyed by at least one of them.

16   I am hard to convince at work.

17   I like to participate in group activities at work.

18   I seldom take offense at work.

19   I cannot stand criticism at work.