A Conditional Reasoning Test for Risk and Incident Propensity

by

Alexa J. Doerr

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of Industrial/Organizational Psychology
Doctor of Philosophy

Auburn, Alabama
August 6, 2016

Keywords: Conditional Reasoning Test, Personality Testing, Test Development, Test Validation

Approved by

Daniel Svyantek, Chair, Professor of Psychology
Jinyan Fan, Associate Professor of Psychology
Jesse Michel, Assistant Professor of Psychology
Alejandro Lazarte, Associate Professor of Psychology

Abstract

Occupational injuries and fatalities persist in large numbers in the United States. Injuries and fatalities that occur at work have a high social and financial cost to employers. Moreover, human error is often to blame for these outcomes. Countless organizations could stand to benefit from the capability to predict safety performance, or alternatively risky behavior, of individuals. Traditional measures of risk propensity generally take the form of self-report test batteries, which are prone to faking and socially desirable responding. Additionally, evidence suggests that individuals are not fully aware of certain aspects of their personality and motivational tendencies. A conditional reasoning test for risk and incident propensity (CRT-RIP) was developed with hopes of tapping into implicit patterns of thought involved in risk propensity. Test items sought to identify individuals who were inclined to endorse biased answer choices in what was presented as a logic test with inductive reasoning problems. Of the 27 CRT-RIP items, 12 items were retained. With initial versions of CRTs, item retention is generally fairly low. As was expected, the CRT-RIP showed small nonsignificant correlations with explicit measures of risk propensity demonstrating that implicit personality is distinct from explicit personality. The CRT-RIP had positive and significant correlations with a risk-taking report and a report of substance use and/or abuse. In hierarchical multiple regressions, the CRT-RIP was shown to have incremental validity over the explicit measures of risk propensity when the dependent measures were risk-taking and substance use and/or abuse. This was true for both time points collected for each scale (i.e., in the past 12 months and in the lifetime). Exploratory factor analysis (EFA) and item response theory (IRT) were used to further examine the scale properties and individual item characteristics, respectively. The EFA suggested that a four-factor model is reasonable, but that a more parsimonious model is also feasible based on multiple fit statistics. A two-parameter logistic model fit both the 27-item and 12-item CRT-RIP well. A few items stood out as high-performing items.

Acknowledgements

I would like to thank my advisor and chair, Daniel Svyantek, for his ongoing guidance and unyielding support during my years at Auburn University. I would also like to express gratitude to the members of my dissertation committee, Jinyan Fan, Jesse Michel, and Alejandro Lazarte, as well as my outside reader, Bonnie Sanderson, for their practical and constructive feedback. In addition to members of my academic committee, I would like to thank my mentor at the Society for Human Resource Management, Alexander Alonso, for his continued investment in my professional efforts.

My personal support system has served as a source of strength without which I would not have been able to accomplish what I have. I would like to thank my parents, Gail and Bruce Doerr, for encouraging me throughout my education and inspiring me to reach my goals. Finally, my friends and other family members deserve thanks for always listening, advising, and sharing their perspectives.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INRODUCTION

Workplace injuries and fatalities persist in high numbers. Occupational accidents account for thousands of deaths and disabilities each year in the United States. In 2014 alone, this amounted to 4,821 work-related fatalities and over 3.7 million nonfatal occupational injuries and illnesses (Bureau of Labor Statistics, U.S. Department of Labor, 2016). Moreover, in 2007 the national cost of workplace injuries and fatalities were estimated to be $6 billion and $186 billion, respectively (Leigh, 2011). These statistics become even more overwhelming when acknowledging that reports of workplace injuries may well be underestimated. Such data typically exclude individuals in the military, and epidemiologic studies have indicated that numerous injuries are either unreported or undocumented (Conway & Svenson, 1998; Parker, Carl, French, & Martin, 1994; Veazie, Landen, Bender, & Amandus, 1994).

Many professions require the recruitment, employment, and retention of workers who are safe and who are motivated to adhere to safety regulations. From the pilots flying our commercial aircrafts, to the medical professionals who care for us, to the chefs preparing our food, safety compliance at work can mean the difference between life and death. Other such workplaces that require safe employees include construction sites, manufacturing plants, and engineering industries. On these job sites, individuals are commonly confronted with precarious conditions that require the use of good judgment, the ability to exercise extreme caution, and the aptitude to accurately appraise potentially hazardous situations.

Data on work-related injuries and fatalities draw the attention of researchers across various fields of study, and stimulate questions regarding potential causes of workplace accidents. Oliver, Cheyne, Tomás, and Cox (2002) propose that numerous factors should be analyzed in an effort to explain the occurrence of occupational accidents, including the characteristics of the individual as well as those of the organizational environments (Iverson & Erwin, 1997; Sheehy & Chapman, 1987). Often times safety articles will discuss safety as it pertains to the organization as a whole, using terms such as "safety culture" and "safety climate." In comparison, there is substantially less research and theoretical groundwork to describe the individual's role in ensuring their own safety and the safety of others in the workplace. There are many ways to assess whether someone is safe at work. Importantly, safety performance must be defined in order to understand how risk can be measured and evaluated.

Conceptualizing Safety

Safety performance is a term that has been used to define both safety outcomes, such as number of injuries per year, and the safety-related behaviors conducted by individuals (Christian, Bradley, Wallace, & Burke, 2009). Safety outcomes are distinct from safety-related behaviors in that they are tangible events or results including accidents, injuries, or fatalities. More proximally related to psychological factors, individual safety behaviors provide researchers with a measurable criterion. Safety-related behaviors, or safety performance, can be predicted with greater precision than safety outcomes. Comparable to general job performance, safety performance behaviors can be measured by the frequency with which individuals participate in the behaviors (Burke, Sarpy, Tesluk, & Smith-Crowe, 2002). Though safety performance is theoretically comparable to general job performance, it does not adequately fit into task, contextual, or adaptive performance and thus, should be considered as distinct from job performance (Burke, Sarpy, Tesluk, & Smith-Crowe, 2002; Parker & Turner, 2002). Burke et al. (2002) have defined safety performance as the behaviors that people engage in to promote the health and safety of colleagues, clients, the public, and the environment. A model of safety performance presented by Burke and colleagues (2002) consists of four components including a) using personal protective equipment, b) engaging in work practice to reduce risk, c) communicating threats and accidents, and d) exercising employee rights and responsibilities. Alternative conceptualizations of safety performance differentiate between generally mandated safety behaviors and safety behaviors that are often voluntary, which have been termed safety compliance and safety participation, respectively (Neal, Griffin, & Hart, 2000).

Neal and Griffin (2004) also posited a model based upon Campbell, McCloy, Oppler, & Sager's (1993) theory of performance, which identifies three proximal contributing factors of an individual's performance—knowledge, skills, and motivation to perform. This model posits that distal antecedents of performance (e.g., training, organizational climate, personality) seemingly influence performance via increases in these proximal determinants. Accordingly, Neal and Griffin (2004) suggested that antecedents such as safety climate or personality directly influence safety motivation and knowledge, which then directly influence safety performance behaviors. Those safety performance behaviors ultimately contribute to safety outcomes, such as accidents and injuries.

Utility of an Implicit Test of Risk and Incident Propensity

Individuals carry with them a wide array of experiences, attitudes, and dispositions that may influence their proneness for risk-taking or accident involvement in their work roles. Unfortunately, self-reported measures of risk propensity are often excessively transparent. Thus, these tests are prone to faking and socially desirable responding, particularly in employment situations, rendering them utterly useless in selection contexts. Researchers (Greenwald & Banaji, 1995; Nisbett & Ross, 1980) have also claimed that individuals may not be entirely aware of the preconceptions and biases they have that motivate their behavior. In the present study, a conditional reasoning test (CRT) was developed with aims of tapping into implicit patterns of thought to identify how readily individuals are able to justify risky or biased interpretations via inductive reasoning. CRTs are further described in chapter 2 and chapter 3.

CHAPTER 2

MEASURING RISK PROPENSITY

In 1919, Greenwald and Woods discovered that the distribution of accidents was not equal across individuals. This suggested that some individuals were more apt to experience accidents than others. In following years, attempts were made to identify the factors that might contribute to individual differences in safety. Various studies have found that accident involvement or risk-taking showed a positive relationship with neuroticism (Frone, 1998; Hansen, 1989; Iverson & Irwin, 1997; Klen & Ojanen, 1998; Sutherland & Cooper, 1991). One proposed explanation for this finding is that people with high levels of neuroticism are more distractible than individuals with low levels and these individuals may suffer from lapses of attention causing them to make more mistakes. Other evidence links risk-taking or accident involvement with Type A behavior, sensation seeking, and extremely high levels of extraversion. Individuals with Type A behavior may be more inclined to be involved in accidents because they experience an elevated sense of time urgency and haste (Frone, 1998). Likewise, high levels of extraversion may lead people to take more risks because they are overconfident. Numerous studies have also looked into the relationship between locus of control and accident involvement with several studies indicating that individuals with an internal locus of control are less likely to have accidents than those with an external locus of control (Jones & Wuebker, 1985).

Particularly under circumstances of uncertainty, people are said to rely on a limited number of heuristic principles, which are used to confine complex tasks of evaluating probabilities and predicting values into simpler judgment operations (Tversky & Kahneman, 1974). Generally, these heuristics are advantageous, but at times, use of heuristics can lead to extreme and systematic errors. For example, when visibility is poor, distances are often overestimated whereas when visibility is good, distances are often underestimated. Therefore, confidence in clarity as in the assessment of distances leads to common biases. Likewise, biases are pervasive in the intuitive judgment of likelihood and can lead to systematic errors. For example, research within the safety literature proposes that as much as 80% to 90% of industrial accidents are due to errors and violations made by individuals (Reason, 1990). Specific biases that may impact safety performance will be outlined in the discussion of the present study. Some individual differences have been posited to increase individuals' vulnerability to the slips, mental lapses, and mistakes that lead to risk-taking and involvement in accidents (Neal & Griffin, 2004). To date, relatively little research has been dedicated to understanding the mechanisms attributable for the individual differences related to safety.

Individual differences that predispose someone to think in systematic and erroneous ways can adversely affect the decisions that he or she makes. When an employee is inclined to make risky decisions or to act without caution, this can become a concern not only for their health and safety, but also for the health and safety of colleagues and the general public. This should be a consideration in the recruitment, selection, and maintenance of a safe workforce. As will be discussed further, individual differences have been shown to influence decision-making with regard to risk-taking. These differences may be assessed as means to predict employee safety performance. Therefore, individual differences with regard to risk propensity may have applications in personnel selection and hiring contexts.

Personality Assessment

The aim of personnel selection is to identify and select the right people for the right jobs by matching individual abilities and needs to organizational rewards and demands (Salgado, Viswesvaran, & Ones, 2001). There are countless ways organizations attempt to recruit and hire the best individuals for vacant positions. Common selection batteries include aptitude and ability tests, job knowledge tests, work sample tests, and personality tests (Salgado, Viswesvaran, & Ones, 2001). Among these, the use of personality tests in staffing decisions has received strong endorsement in hundreds of primary studies and dozens of meta-analyses conducted and published since the mid-1980s (Ones, Dilchert, & Viswesvaran, 2007). Personality constructs have proven useful in the explanation and prediction of various attitudes, behaviors, performance, and outcomes in organizational domains. Moreover, several professionally developed measures of personality traits demonstrate useful degrees of criterion-related validity for job performance and its facets.

Personality predictors can be parsed into two categories (Salgado, Viswesvaran, & Ones, 2001). First, general measures of adult personality aim to accurately describe individual differences in personality and were created to give a broad description of personality that could be applied to a wide range of contexts. The second category of personality measures used in personnel selection can be referred to as measures of personality at work. These tests are constructed to provide accurate prediction of individual differences in work behaviors of interest. These individual difference measures may include integrity tests, violence measures, sales potential scales, and managerial potential assessments (Salgado, Viswesvaran, & Ones, 2001). Some of these measures are intended for the prediction of specific criteria such as theft or violence at work whereas others have been created for specific occupational groups such as sales employees or managers.

Social-Cognitive Theory of Personality

Personality is multifaceted and that makes accurate and comprehensive measurement challenging. Research in the domains of both self-perception (Bern, 1972) and social inference (Nisbett & Ross, 1980) suggests that individuals do not necessarily know why they do things and that people are not necessarily cognizant of the causes of their behavior. The high level of dependence on direct measures of attitudes observed by Greenwald and Banaji (1995) designates a pervasive assumption that attitudes operate predominantly in a conscious mode. Conversely, Campbell (1963) posited that acquired dispositions such as attitudes retain residue of experience as to guide or bias, or otherwise impact future behavior.

As such, social-cognitive theory aims to explain how people perceive and interpret their social world, and how that manifests in patterns of human behavior (Dweck & Leggett, 1988; Fazio, Jackson, Dunton, & Williams, 1995; Greenwald & Banaji, 1995; Mischel & Shoda, 1995). The social-cognitive perspective and corresponding research have found support for the notion that both implicit and explicit social cognitions are instrumental in understanding how personality shapes behavior (Bing, LeBreton, Davison, Migetz, & James, 2007; Frost, Ko, & James, 2007; Gawronski & Bodenhausen, 2006). Implicit social cognitions pertain to the unconscious and automatic thoughts related to a person's beliefs, attitudes, and behaviors whereas explicit social cognitions pertain to the easily accessible and controlled thoughts related to a person's beliefs, attitudes and behaviors. Theoretically, operationally, and empirically distinct, implicit and explicit social cognitions jointly serve as a precursor to stable patterns of behavior (Michel, Pace, Edun, Sawhney, & Thomas, 2015; Nieminen, 2012). Each of these constructs has been evidenced to contribute unique yet often intersecting explanatory ability (see Bing et al., 2007; Frost et al., 2007; Gawronski & Bodenhausen, 2006).

Implicit cognition is largely ignored in most selection assessments. Nisbett and Wilson (1977) reasoned convincingly that psychology's dependence on self-report measures was unwarranted due to evidence exhibiting the lack of introspective availability regarding the causes of behavior. Wegner and Vallacher (1977, 1981) made similar remarks concerning influences on social behavior that are not always accessible. Greenwald and Banaji (1995) also called for methods to measure the individual differences attributable to implicit social cognition and claimed that, at the time of their publication, measurement of those individual differences was beyond the "means of present assessment technology" (p. 6). Due to this limitation, a vast subset of the empirical implications of implicit social cognition remained largely untestable. However, since their call for measures of implicit attitudes, more recent developments have made the so called, "untestable," testable. Among these, conditional reasoning tests (CRTs) have provided an assessment technology that has been utilized to test unconscious mental processes of interest.

Conditional Reasoning: A New Format for Personality Testing

In response to popular concerns with traditional personality tests such as job candidates distorting their responses coupled with the desire for a method of indirect personality measurement, James (1998; James & Mazzerolle, 2002; James & McIntyre, 2000; James, McIntyre, Glison, Bowler, & Mitchell, 2004; James et al., 2005) sought to design a motive-based personality test that would tap into implicit thought patterns and that would be resistant to such issues as socially desirable responding and faking. Conditional reasoning tests were constructed to measure implicit, or unconscious, cognitive biases which individuals with certain latent motives use to substantiate their actions (LeBreton, Barksdale, Robin, & James, 2007). CRTs allow assessors to fool the test-takers before they can fool the assessors. By concealing the true nature of the test, these batteries intend to measure biases that function beneath the surface of consciousness. These tests are presented as a method for evaluating how people solve seemingly tradition inductive reasoning problems whereas the real intent of the test is to determine if response choices based on relevant implicit biases are logically appealing to a test taker (LeBreton et al., 2007).

Very few conditional reasoning tests are currently in existence. This is to be expected, as the technology is still rather novel. CRTs have been constructed for achievement motivation (James, 1998), aggression (James & McIntyre, 2000), addiction proneness (Bowler, Bowler, & James, 2011), counterproductive work behaviors (Fine & Gottlieb, 2013), team orientation (Driskell, Goodwin, Salas, & O'Shea, 2006), and creative personality (Schoen, Bowler, & Schilpzand, 2016). Through various searches, these are the CRTs that emerged as presented in published scholarly articles. A few other CRTs have also been presented at conferences or were developed for dissertations, but have yet to be published. Based upon the abundance of literature on the CRT-A aggression scale, it is perhaps the most popular and successful CRT yet.

Deception

As previously noted, a chief component of these indirect assessments is that the subject must not be alerted to the construct or variable that is actually being measured. LeBreton, Barksdale, Robin, and James (2007) empirically tested the importance of concealing the actual intent of the scale. Results indicated that scores on the CRT-A were influenced by a respondent's knowledge of what was being tested (i.e., aggression). Specifically, results revealed that disclosing the intent of the test affected individuals' aptitudes to see through the test and identify the aggressive item responses. Compared to participants in the control group who were not informed of the purpose of the test, individuals who were made cognizant of the true nature of the scale and then asked to present themselves in the worst conceivable manner (fake bad) were able to identify and select the aggressive responses. This evidence suggests that concealing the true purpose of a conditional reasoning test is a critical feature of this type of assessment and should be upheld.

This deception may also be advantageous in getting a more accurate portrayal of personality. Abundant research indicates that, when instructed to do so, people are capable of distorting their responses in the desired direction (Dunnett, Koun, & Barber, 1981; Furnham & Craig, 1987; Hinrichsen, Gryll, Bradley, & Katahn, 1975; Schwab, 1971; Thornton & Gierasch, 1980; Walker, 1985). This apparent ability to distort responses in a desired direction coupled with the motivation of acquiring a sought after job position may make for a dangerous combination. Often self-report measures such as personality and trait-based tests fall victim to response distortion or faking (Rosse, Stecher, Miller, & Levin, 1998). Many selection researchers have retained the position that response distortion is more predominant in employment circumstances (Schmit & Ryan, 1993) and that it does, in fact, influence the rank order of job candidates, ultimately affecting the selection decision made by the organization (Rosse et al., 1998). The use of impression management tactics has proven incredibly pervasive in employment interviews with empirical results implying that all, or almost all, interviewees use one or more impression management tactics during interviews (Ellis, West, Ryan, & DeShon, 2002; Levashina & Campion, 2007; Stevens & Kristof, 1995; Weiss & Feldman, 2006).

Traditional personality inventories are often particularly vulnerable because of the high level of transparency of what is being tested (Fan et al., 2012). Moreover, personality measures have been found to correlate considerably with cognitive ability or general intelligence (*g*) (Pauls & Crost, 2005). Mersman and Shultz (1998) demonstrated that mean scores of faking and mean difference scores between faked and honest scores of emotional stability and conscientiousness were associated with general intelligence (*g*). The reason people, especially those with higher cognitive ability, can "fake good" on personality tests is that they can use the high level of transparency that is common among traditional personality tests to guess what responses are deemed desirable and then answer accordingly. If individuals know what is being measured, they can often identify the "right" answer or the answer that would be regarded most favorably by a potential employer. Issues that may arise in the event of faking include reduced scale variance, augmented or exaggerated scale mean scores, and pronounced collinearity among scales (see Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; Viswesvaran & Ones, 1999; Zickar & Robie, 1999). Because conditional reasoning both removes transparency that is typical of traditional personality tests and seeks to measure implicit cognitive processes, the conditional reasoning technique should be more resistant to faking.

CHAPTER 3

THE PRESENT STUDY

Individual risk propensity and risk involvement have been shown to be highly related to more general personality components (i.e., the Big Five), which are considered fairly stable across time and circumstances (Beus, Dhanani, & McCord, 2015). If predisposition to act safely is highly personality-dependent, it brings about the question of how readily safety decision-making and safety performance may be trained or learned. As such, selecting individuals who are more likely to conduct themselves in a less risky manner may be vital to achieving a safe workforce. Personnel selection experts have called for new research devoted to investigating the criterion validity of personnel selection methods for predicting work accidents and injuries (Salgado, Viswesvaran, & Ones, 2001).

In the present study, a conditional reasoning test was developed for the purpose of measuring an individual's risk and incident propensity. Potential applications include selecting valuable employees, evaluating safety of current employees, and assessing the success of safety training programs. Specifically, this could be used to assess the success of safety training programs by having individuals take the assessment before and after the training and comparing pre-post scores. Importantly, if conditional reasoning were used in this fashion, test-retest reliability would have to first be established. There may be issues with the test becoming more transparent with greater exposure. Alternatively, if the work group is large enough, one randomly assigned group of individuals could take the test prior to training and the other group could take the test after training and scores of the two groups could be compared in order to assess the success of training. The hope for this test is to predict one's propensity to make risky decisions and their potential to be involved in incidents. Much of safety performance involves sound decision-making under pressure and in conditions of uncertainty. If this test is shown to have high predictive validity and low transparency, it may be particularly useful for selection purposes for jobs that require the employment of exceptionally safe workers. As previous CRTs, the test was developed on the basis of justification mechanisms by which individuals tend to explain their unsafe and less than optimal choices.

Justification Mechanisms

The premise underlying conditional reasoning is that the process of rationalizing typically involves inductive forms of reasoning (LeBreton, Barksdale, Robin, & James, 2007). Motives for behavior are said to possess explicit (conscious) as well as implicit (unconscious) components (McClelland, Koestner, & Weinberger, 1989). Generally, people must see themselves as possessing self-worth. This means that they aspire to think of themselves as being moral, stable, and capable of self-control (Bersoff, 1999; Loewenstein, Weber, Hsee, & Welch, 2001). Thus, motives that are typically viewed as negative remain largely implicit because they conflict with the motive to maintain a favorable self-image (Baumeister, Campbell, Krueger, & Vohs, 2003; Cramer, 1998, 2000). Unconscious defensive practices make potentially negative behavior permissible while concurrently guarding one's sense of self-worth. These unconscious self-guarding processes have been termed "defense mechanisms." Defense mechanisms provide mental operations intended to keep painful thoughts and emotions out of consciousness (Cramer, 2000; Kilstrom, 1999). Painful thoughts may include that one possesses some negative disposition or inclination.

Specifically, conditional reasoning relies on the defense mechanism of rationalization by which individuals use seemingly reasonable explanations to justify behaviors that are unknowingly attributable to unconscious, unacceptable, and/or undesirable motives (Baumeister, Smart, & Boden, 1996; Westen & Gabbard, 1999). In other words, rationalization involves making inferences to interpret their behavior, explain their behavior, and project consequences of their behavior. The unconscious, yet, true intent of reasoning is to augment the rational appeal of adverse or negative behavior. To accomplish this, individuals may use, justification mechanisms (JMs), which are self-preserving biases that implicitly form validation so as to boost the rational appeal of conducting oneself in an unfavorable manner (James, 1998). Essentially, JMs are mental processes operating below the surface of consciousness to yield rationalizations. The author searched both scholarly and lay text to identify biases thought to contribute to individual safety. For risk and incident propensity, brief descriptions of each JMs are available as Table 1. More comprehensive explanations are presented following Table 1.

Table 1
Justification Mechanisms for Risk and Incident Propensity

Bias	**Explanation**
Optimism Bias	A belief that one is invulnerable to harm. Individuals who exhibit optimism bias tend to overestimate the likelihood of favorable events happening to them and to underestimate the likelihood of unfavorable events occurring to them.
Social Proof	A propensity to follow the example of others, particularly in the case of groups. Individuals tend to take on the action(s) and behaviors of others with the intent of acting appropriately or correctly in a given situation. Also see informational social influence.
Externality Bias	A tendency to believe that an individual’s outcomes are largely due to chance, powerful others, or external sources rather than due to one's own influence and decisions.
Immediate Gratification Bias	The desire or need for pleasure and satisfaction on a momentary basis. The inclination to overvalue instant happiness and to place little to no emphasis on the long-term consequences of decisions or behavior.

Optimism Bias

Optimism bias refers to an individual's systematic tendency to believe, often in error, "that they are less likely than the average person to experience negative life events and more likely than the average person to experience positive life events" (Borschmann, Lines, & Cottrell, 2012, p. 182). Other researchers have made mention of similar constructs which they have referred to as unrealistic optimism (Sharot, Korn, & Dolan, 2011), bulletproof mentality (Anderson & Lorber, 2006), comparative bias (Harris, Griffin, & Murray, 2008) and illusions of invulnerability (Roe-Burning & Straker, 1997). Optimism bias may occur when individuals form perceptions of risks using incomplete or unclear evidence regarding the true nature of those risks (Costa-Font, Mossialos, & Rudisill, 2009). Consequently, individuals perceive risks to be smaller for themselves than others.

Research on perceptions of fatigue in the trucking industry have shown that drivers and company representatives think that, compared to other companies' employees, fatigue is less of a problem for themselves and their employees (Arnold et al., 1997). This disproportionate optimism may encourage risk-taking (Stone, Choi, Bruine de Bruin, & Mandel, 2013). Data obtained using a focus-group methodology has contributed anecdotal evidence of optimism bias with regard to organizational health and safety (Workcover New South Wales, 2002). In nine focus groups with electricians, a concern was consistently raised that some electricians take shortcuts that may be unsafe guided by the mentality that "it won't happen to me." Bias concerning one's supposed invulnerability has the potential to hinder the adoption and maintenance of preventive behaviors because if one believes that particular risks usually apply more to others than to oneself, there may be insignificant motivation to employ preventive measures (van der Pligt, 1996). Although Caponecchia (2013) observed no relationship between optimism bias and reported use of precautionary behaviors, the author suggested that this may be a function of the self-report methodology.

Social Proof

From the time we are born, humans learn that the judgments and perceptions of others are often dependable sources of evidence about reality (Deutsch & Gerard, 1955). The principle of social proof implies that in a given situation individuals perceive a behavior to be correct to the extent that they see others performing it (Cialdini, 1993). Cialdini (1993) has suggested that situations in which there is a high degree of uncertainty contribute to the effect of social proof. Many patterns of behavior that increase the risk of illness and injury can arise when people are apprehensive with how they are perceived by others (Leary, Tchividjian, & Kraxberger, 1994). The aim to convey certain social images can encourage behaviors that are hazardous for the individual, if not for others as well. Self-presentation concerns may also lead an individual to practice less caution which could contribute to the occurrence of accidents or injuries. This can be seen in people's reluctance to wear protective equipment or gear (e.g., safety goggles, gloves, and helmets) when operating power tools or hazardous machinery for fear that they will be viewed as neurotic or exceedingly careful.

Externality Bias

The term locus of control (LOC) has been used to describe how inclined someone is to believe they have control over their life circumstances (Eatough & Spector, 2014). LOC is typically conceptualized such that an individual is predisposed to believe that he or she has considerable personal control over the events in his or her life (internal locus of control) or that luck, fate, or powerful others are more frequently the source of control in his or her life (external locus of control; Rotter, 1966). Internal locus of control and external locus of control have been referred to as internality and externality, respectively (Hunter & Stewart, 2012; Samreen & Zubair, 2013).

In a sample of military aviators, Hunter and Stewart (2012) found a negative relationship between internality scores and recent accident status. Lack of perceived control can be hazardous to an individual's health and wellbeing due to the diminished feelings of responsibility for certain actions. Demonstrating this notion, a positive correlation was observed between externality and involvement in hazardous events among a sample of 205 Indian Army aviators (Reddy & Sharma, 2013). Additionally, this study found that external locus of control was positively related to higher impulsivity, anxiety, and weather anxiety. Lower self-confidence and lower safety orientation were also found to be positively associated with external locus of control.

In a cross-sectional telephone survey administered to randomly selected individuals who resided and worked on family farms, individuals who reported having high internal safety locus of control were more likely to openly communicate their mistakes at work (Cigularov, Chen, & Stallones, 2009). This is a safe and preventative practice because openly communicating about mistakes or errors in their work allows greater opportunity to effectively manage those errors while producing the chance to learn from mistakes (Frese, 1995). Similarly, frequent error communication among nurses was evidenced to be associated with fewer back injuries and medication errors (Hofmann & Mark, 2006; Mark et al., 2007). The implications of internality-externality on health and safety have long been apparent.

Immediate Gratification Bias

An individual's ability to postpone receiving an immediate reward in order to reap additional benefits in the future (i.e., shallow delay discounting) has positive implications for one's health, wealth, and happiness (Cheng, Shein, & Chiou, 2012; Cherukupalli, 2010; Daugherty & Brase, 2010; Dittmar & Bond, 2010). Those who prefer a smaller, more immediate reward to a larger yet delayed reward demonstrate a steep discounting rate. Choosing a smaller, more immediate reward instead of a larger, but delayed reward has been considered as, 'impulsive' and signifies lack of 'self-control' (Cheng, Shein, & Chiou, 2012). Zimbardo, Keough, and Boyd (1997) established that individuals who hold a present time perspective tend to take more risks and to act more impulsively. Furthermore, time perspective has been shown to be a significant predictor of many risk behaviors such as smoking, drug use, and drinking (Daugherty & Brase, 2010).

Item Development and Scoring

Conditional reasoning tests (CRTs) satisfied the need for a method to measure the extent to which JMs are instrumental in establishing a person's reasoning. The chief struggle in developing a method to measure this was the implicit or unconscious nature of biases. JMs cannot be evaluated by the popular and simple method of self-report (Nisbett & Wilson, 1977; Winter, John, Stewart, Klohnen, & Duncan, 1998). Instead, assessment of JMs requires indirect measurements (Greenwald & Banaji, 1995). The key feature of indirect measures is that they do not alert the subject to the identity of the construct being measured (Greenwald & Banaji, 1995).

Conditional reasoning items are intended to tap into the implicit reasoning tactics that individuals employ to enhance the logical appeal of their behavior (James, 1998). Conditional reasoning items consist of the same basic structure and test administration conditions as objective reasoning items. For instance, they usually include an item stem and four response alternatives, and respondents are instructed to select the most reasonable response choice. Inductive reasoning problems, by definition, need to have logically valid and logically invalid response alternatives (LeBreton et al., 2007). Inductive reasoning involves a logical and probabilistic extrapolation from the information presented in the problem. Probabilistic inferences are judgments regarding what is most likely to be true on the basis of available pieces of information (Nieminen, 2012). This implies that alternative solutions could be possible if additional information is provided. Differing from inductive reasoning, deductive reasoning implies that there is only one conclusion that must be correct following from true premises (Moore, 1998).

Traditional inductive reasoning questions aim to measure cognitive ability and as such, have one logical response and three or four illogical responses. These problems along with the answer choices available vary in degree of difficulty. This denotes that for the more challenging items, a greater amount of the respondents will choose an illogical and therefore incorrect response. Whereas objective reasoning items contain only one correct answer, conditional reasoning items essentially have two "correct" answers (James, 1998). One response alternative echoes the implicit assumptions, JMs, and information processing approaches that are characteristic of the focal personality trait or motive and the other reflects opposing implicit justification strategies.

Conditional reasoning problems are designed to take on the appearance of traditional inductive reasoning problems (LeBreton et al., 2007). In keeping with traditional inductive reasoning tasks, the respondents are asked to determine which general conclusion follows most reasonably from the evidence provided. In order to do so, respondents must make judgments about which evidence is plausible, which points are valid, which assumptions are reasonable, and, decisively, which one of the provided conclusions is most probable to be correct (LeBreton et al., 2007). This ensures respondents are entirely focused on finding logical answers to reasoning problems and thus, allows the implicit processes to operate free from the self-protective mechanisms that typically pose threat to traditional self-report personality instruments (e.g., socially desirable responding, self-deception).

LeBreton and colleagues (2007) have found that respondents largely trust that their critical intellectual abilities guide their attempts to identify a logically correct conclusion for each problem. Yet, the test is deceptive in that the true demands of the inductive reasoning tasks are produced by an underlying requirement to assess whether reasoning based on a JM is more or less sensible than reasoning based on more moderate ideologies and rationales.

Illogical Response Choices

Assuming the appearance of traditional inductive reasoning problems, in this case, necessitates inclusion of illogical responses, or "distractor" response choices to the problems. Each question contains two illogical answer choices. One might inquire as to whether it is truly essential to include these illogical responses. Another consideration is how illogical responses are to be interpreted and scored (LeBreton et al., 2007). To ease these concerns, conditional reasoning problems intentionally use distractor answers that are obviously illogical to individuals who are operating with cognitive skills at least at the seventh grade reading level (see James & McIntyre, 2000). As a result, practically no one attempts to solve conditional reasoning problems by picking either of the two logically incorrect response choices. James et al. (2005) posit that, in their experience with thousands of administrations, one of the logical answers is nearly always selected to solve the problem.

Departing from traditional inductive reasoning items, the conditional reasoning problems involve two logical responses and two illogical responses. The problems all have high item difficulty values. This means that all items have high *p* values, signifying that they are very easy and that almost everyone selects a logically correct answer (LeBreton, Barksdale, Robin, & James, 2007). Thus, the technique used to build conditional reasoning items causes them to be poor measures of cognitive ability because all items are simple and almost everyone would get a perfect score if the items were rescored based on logically valid answers.

Hypotheses

The scale was developed on the basis of the four implicit personality facets (i.e., JMs; optimism bias, externality bias, social proof, and immediate gratification bias) outlined above that should theoretically be linked to risk propensity. The overall scale is expected to predict propensity for risky behavior that may contribute to the occurrence of incidents. Because implicit personality is largely unconscious and not readily available (Greenwald & Banaji, 1995; Haidt, 2001), implicit and explicit personality are considered distinct constructs. Thus, implicit and explicit personality attributes are not expected to correlate highly (Frost et al., 2007; James & LeBreton, 2012). Individuals typically have sparse insight into their own reasoning processes, and there is little theoretical ground to substantiate these entities of personality being interrelated. As such, implicit and explicit risk and incident propensity are not expected to correlate highly.

Hypothesis 1: Scores on the CRT-RIP and explicit measures of risk propensity will not be highly correlated.

The relationship between personality and behavior has been long established. In particular, an association between some attributes of personality and risk behavior has been evidenced (Frone, 1998; Hansen, 1989; Iverson & Irwin, 1997; Klen & Ojanen, 1998; Sutherland & Cooper, 1991). The aforementioned JMs are expected to influence risk behavior without conscious awareness (i.e., implicitly). If individuals are prepared to justify biased thinking related to risk behavior, they should be more likely to have partaken in risky behavior or to have been involved in incidents.

Hypothesis 2: Scores on the CRT-RIP will be positively correlated with scores on a) a risk-taking report, and b) substance use and/or abuse.

Current measures of risk and incident propensity often utilize self-report, which makes them susceptible to faking and response distortion (Rosse et al., 1998). These self-report measures are particularly vulnerable to faking and other issues when examinees are in a high-stakes situation such as under employment circumstances (Schmit & Ryan, 1993). This renders the explicit measures of risk and incident propensity ineffectual in selection contexts. If someone is applying for a safety-related job and they know that the test they are asking is attempting to gauge how safe they are, the test essentially acts as a cognitive ability test rather than a test of risk propensity. In addition to the issues outlined above, explicit measures of risk propensity may not be appropriate for theoretical reasons. Individuals prefer to maintain a positive self-image (Bersoff, 1999; Loewenstein, Weber, Hsee, & Welch, 2001) and thus, individuals may truly believe that they are less risky than is accurate. For these reasons, the CRT-RIP is expected to show incremental validity over previous explicit measures of risk propensity.

Hypothesis 3: The CRT-RIP will have incremental validity above and beyond previous scales measuring similar constructs for predicting risk propensity.

Factor analysis and item response theory (IRT) will be used to examine the underlying structure of the CRT-RIP and to assess individual items within the CRT-RIP. Because the test consists of items measuring four JMs, it is predicted that four factors will emerge in an exploratory factor analysis. Additionally, the 2PL IRT model will be fitted to the data and item quality will be evaluated. The IRT model should show adequate fit and items should demonstrate high performance in terms of measuring the latent trait (risk and incident propensity). Based on these pretenses, the following are hypothesized:

Hypothesis 4: A four-factor model will exhibit superior fit over other solutions in a confirmatory factor analysis of the observed data.

Hypothesis 5: Item response theory will demonstrate that items are measuring the intended latent trait.

CHAPTER 4

METHOD

The present study was designed with the goal of developing and validating a new conditional reasoning test for identifying risk and incident propensity. This study was conducted with a sample of college students during a one-hour lab session during which the CRT-RIP and additional measures were administered. Students' risk-taking history was collected using a risk-taking incident report that served as the primary criterion for testing validity. Information regarding substance use and/or abuse was also collected.

Participants

Participants were 331 undergraduate students at a large public Southeastern university in the U.S. Participants were recruited through SONA, an online research participant registration program. Individuals who participated were granted extra credit for participating in the study. A majority of participants (74.6%) were female. The mean age of the participants was 20 (range 18-41). The sample consisted of 85.8% (N = 284) Caucasian, 9.4% (N = 31) African American, 1.8% (N = 6) Asian American/Pacific Islander, 1.2% (N = 3) Hispanic, .9% (N = 3) Arabic, and .9% (N = 3) who identified as other. Three hundred twenty-five (98.2%) participants were native English speakers, five (1.5%) reported that they were not native English speakers, but were fluent in English, and only one reported not being a native English speaker and not being fluent in English. The entire sample consisted of 340 participants to which exclusion criteria was applied. In following recommendations for exclusion outlined below, the resulting sample was brought down to 331 individuals.

Exclusion Criteria

Prior to running any analyses, individuals were removed using specified exclusion criteria. As previously noted, distractor items are designed to be noticeably wrong, and are often ruled out quickly, leaving the respondent with essentially two answers to consider. A high frequency of responding with distractor items indicates that perhaps an individual is not paying full attention. Therefore, a high rate of distractor selection suggests careless responding. James (1998) recommended excluding individuals who endorsed five or more distractor answer choices. However, the earlier CRTs were much shorter in length. To compensate for having a longer test, Schoen et al. (2016) made the decision to exclude individuals who responded to approximately one quarter of the test selecting distractors. This aligns with actual practice of many researchers including James (1998).

In following with previous CRT research, individuals were excluded from the CRT-RIP analyses if the participant responded with six or more distractors. On this basis, eight individuals were excluded from the analyses. One additional participant who endorsed five distractors and had one missing response was also excluded from the analyses.

Measures

Conditional Reasoning Test for Risk and Incident Propensity

A CRT for individual risk and incident propensity was developed, which drew on biases that were viewed as relevant to riskiness. Thus, each item was established with one of the previously mentioned biases in mind. In each item of the test, one answer choice was designed to logically appeal to an individual who possessed the bias. Another answer choice was designed to logically appeal to someone who does not possess the bias. Furthermore, to enhance the façade that test-takers were responding to a logic test, two additional answer choices were included as distractor items. The illogical choices are expected to be seldom selected. Five actual logic questions were also included to increase fidelity of the test, but these will not be included in the scoring.

A sample item for the social proof JM is:

Many bosses do not like to punish high-performing employees who break company rules. Instead, the bosses will turn a blind eye. They do not want to discourage hard working employees.

Which of the following does this imply?

a) Safety policies are particularly useless.
b) Many rules are not that important to getting the job done adequately and safely. (Social Proof)
c) Bosses wear more expensive clothing to show off their status.
d) Bosses are not enforcing rules that are put in place for a reason. (Non-biased)

This question is intended to measure one's use of social proof (SP) as a means to justify or rationalize risky behavior. In this example, answer choice B represents the biased (SP) response, whereas answer choice D represents a non-biased response. Answer choice B rationalizes that if bosses are not enforcing certain rules, that those rules do not need to be followed. This hints at an individual's inclination to use social cues and social information in order to justify acting in a potentially unsafe manner. The non-biased response, D, acknowledges the potential for negative consequences of rules not being enforced at work. Answer choices A and C serve as distractor items in this example. The only purpose of the distractor terms is to maintain the appearance of a true logic test. The distractor response choices are designed to be ruled out quickly, and they are typically not endorsed. If a test-taker endorses a high number of distractor responses, this may serve as a manipulation check because the participant is likely not taking the test seriously and their responses could skew the data. Likewise, if particular items experience a high number of participants endorsing distractor terms, this may designate that an item may need to be dropped from the scale. For this scale, each item was recoded as such: 1 = biased choice endorsed, 0 = distracter or non-biased choice endorsed.

A sample item for the optimism bias JM is:

When choosing a horse to bet on at a race, selecting a horse that has a good record and seed will increase the chances of you winning. However, the reward for betting on a favored horse that wins is often minimal compared to the reward for betting on an underdog if they were to win.

Based on the above, which statement below is the most logical?

a) The least attractive horses usually win the race.
b) Horse racing is not considered a sport.
c) Betting on underdogs can yield larger sums of prize money. (Optimism Bias)
d) Conservative bets will more consistently pay off, but in smaller quantities. (Non-biased)

In this item, answer choice C represents the biased response. Specifically, endorsement of answer choice C signifies that the respondent may be inclined to rely upon optimism bias to justify risky behavior. Answer choice C sanctions betting on horses that are less likely to win races because on the off chance that the underdog does win, the payoff will be larger. An individual's tendency to overestimate the likelihood of favorable events happening to them compared to others may cause him or her to perceive more risky bets as more appealing. Answer choice D represents the non-biased response. There is generally less risk involved in betting on horses that are more likely to win given their track record and other qualifications. This item, and all other CRT items were also scored dichotomously with a score of one signifying endorsement of the biased response.

Domain-Specific Risk-Taking

The DOSPERT (Domain-Specific Risk Taking) scale was used to measure explicit risk-taking inclination. This scale consists of 30 items. The DOSPERT model proposed by Weber, Blais, and Betz (2002) predicts five domains for which the authors developed scale items including social, recreational, health-and-safety, ethical, and financial risks. Participants were asked how likely they would be to engage in specific behaviors. Participants responded to six questions from each domain using a 7-point scale ranging from 1 (extremely unlikely) to 7 (extremely likely). An example item is, "Driving a car without wearing a seat belt." This scale was expected to have little to no correlation with the CRT-RIP. The DOSPERT had an acceptable Cronbach alpha of .79. The full DOSPERT can be found in Appendix A.

Risk Propensity

The risk propensity scale (RPS; Meertens & Lion, 2008) was used for the purpose of measuring general tendency to take risks. The scale consists of seven items that were originally rated on a 9-point scale ranging from 1 (totally disagree) to 9 (totally agree). Higher scores signify greater risk-seeking tendencies. For consistency with other measures in this study, the RPS will be rated on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). This scale is expected to have little to no correlation with the CRT-RIP. The RPS had a Cronbach alpha of .82. The RPS appears in Appendix B.

Risk-Taking Incident Report

A risk-taking measure (Brache & Stockwell, 2011) was utilized to obtain information about the risk-taking behaviors of participants in the past 12 months and in their lifetime. The risk-taking report consisted of six items and participants were asked to respond with 'Yes' or 'No' to each item with regard to the past 12 months and in their lifetime. An example item is, "Have you been ticketed for speeding." Scores on this scale will be used to identify whether the CRT-RIP is associated with risky behavior and likelihood of experiencing incidents. Because this is asking about events that have already happened, it will not be a truly predictive measure, but it will provide correlational data between performance on the CRT-RIP and involvement in incidents. The Cronbach alphas of this scale were .44 for in the past 12 months and .47 for in the lifetime, which are fairly low reliabilities. The low reliability may have influenced relationships with other variables. The risk-taking incident report can be found in Appendix C.

Substance Use and/or Abuse

Participants also responded to seven items that asked whether they had used and/or abused substances, both legally and illegally. Again, participants were asked to respond with 'Yes' or 'No' to each item with regard to the past 12 months and in their lifetime. An example item is, "Have you taken a higher dosage of medicine than recommended by your doctor or the package insert?" Scores on this scale were used as a secondary criterion. The Cronbach alphas for this were .73 for in the past 12 months and .76 for in the lifetime. The substance use and/or abuse scale is available in Appendix D.

Procedure

The CRT for risk and incident propensity was developed and was administered in a proctored setting. Participants responded to the scale under the impression that they were taking a logic test. As a manipulation check, at the end of the test, respondents were asked what they thought the test measured. A vast majority of participants responded with, "critical reasoning skills," "logic skills," or "decision-making." Only one of all 331 participants responded to this question that they thought the test measured how cynical they were and how willing he or she was to take risks. The fact that less than 1% of participants were able to guess the true nature of the test demonstrates lack of transparency. Following the proctored test, participants completed a survey on the computers in the lab. The participants responded to the DOSPERT, risk propensity scale, a risk-taking incident report described above, and items that measured substance use and/or abuse. Lastly, participants responded to a few demographic questions. Upon completion, participants were debriefed and this completed their participation.

CHAPTER 5

RESULTS

After applying the exclusion criteria described in the method section, the subsequent analyses were conducted with a sample of 331 individuals. The results are presented in two components. First, individual item characteristics were used to refine the scale. Items with the weakest statistical properties were removed. Second, hypothesis testing was reported. With the revised CRT-RIP, hypotheses with regard to concurrent and incremental validity were assessed. Scale properties were also examined in an exploratory factor analysis (EFA) and item quality was further assessed using item response theory (IRT).

Scale Refinement and Item Deletion

The scale refinement process will follow the recommendations made by James and LeBreton (2011). This involved examining the (a) item response characteristics, (b) item-total correlations, and (c) item-criterion correlations. Items that were identified as needing further consideration were flagged and these items were later assessed for the purpose of removing poor performing items.

Item-total correlations

Items were deleted if they exhibited a negative item-total correlation. In following with James and LeBreton's (2011) recommendations, biserial item-total correlations were examined. A biserial correlation corrects for an artificial dichotomy in one of the two correlated variables and instead estimates what the correlation would be had the measure reproduced the variable's underlying continuous nature. Biserial correlations were appropriate because even though items were scored dichotomously, they are intended to measure a continuous construct. All item-total correlations were positive and larger than .07, a cut-off advised by Schoen and colleagues (2016). Thus, no items were deleted on this basis.

Item response characteristics

Next, items were deleted if more than 95% or less than 5% of participants endorsed the biased response. None of the items demonstrated the biased response being endorsed by more than 95% of participants. In two items, less than 5% of respondents endorsed the biased response. However, one of these items (item 12) was retained because its other item statistics were strong. One additional item (item 15) was deleted because a high proportion of respondents (19%) endorsed a distractor answer choice.

Item-criterion correlations

Lastly, the direction and magnitude of item-criterion correlations were also examined. Items were deleted if they were negatively correlated with the dependent measures. In this analysis, each item was inspected in terms of the direction and statistical significance with each dependent measure (i.e., risk-taking scale and substance use and/or abuse). In general, positive correlations were anticipated, suggesting that endorsement of the JM response choice is associated with a higher likelihood of experiencing incidents related to risky behavior. Because total scores of the dependent measures were continuous and the items are dichotomously scored, biserial correlations were used. In the case that an item had a small negative correlation with one or more dependent measures, but also exhibited a positive correlation with other dependent measures, the item was retained. Thirteen additional items were removed because they had negative correlations with one or more of the dependent measures. Not all of these negative correlations were significant. However, at best, this signified no relationship with the dependent measures. Ideally, only items that contribute to the predictive validity should be retained.

Resulting Scale

Correlations with the criterion variables were weighted more heavily in all verdicts about item retention or deletion. In isolation, a positive item-total correlation was not considered sufficient evidence to warrant an item's retention. After applying all recommendations regarding item deletion, the resulting scale consisted of 12 items. Item deletion decisions are depicted in Table 2. The 12-item scale was used to test hypothesis 1, 2, and 3.

Table 2
Item Deletion Decisions

			Proportion of Responses			
Item	Bias	Item-Total R_{bis}	Biased	Distractor	Negative with DVs	Decision
1	EB	0.37	36.6%	0.9%	Yes	Reject
2	OB	0.34	23.6%	1.5%	No	Retain
3	EB	0.26	71.0%	1.5%	Yes	Retain
4	SP	0.15	84.0%	6.6%	Yes	Reject
5	IG	0.21	11.5%	19.3%	Yes	Reject
6	OB	0.20	84.0%	0.3%	Yes	Reject
7	EB	0.26	59.5%	4.8%	Yes	Reject
8	OB	0.20	53.8%	1.5%	Yes	Retain
9	EB	0.38	26.9%	0.9%	No	Retain
10	SP	0.17	59.2%	3.6%	Yes	Reject
11	IG	0.37	31.1%	0.9%	Yes	Reject
12	OB	0.34	1.5%	1.5%	No	Retain
13	EB	0.26	6.0%	0.6%	No	Retain
14	OB	0.19	80.1%	1.2%	Yes	Reject
15	EB	0.42	45.0%	0.6%	Yes	Reject
16	SP	0.31	39.3%	3.3%	No	Retain
17	IG	0.35	34.4%	0.6%	Yes	Reject
18	IG	0.24	15.7%	0.0%	No	Retain
19	EB	0.41	54.4%	3.3%	Yes	Reject
20	OB	0.31	6.3%	0.9%	Yes	Reject
21	EB	0.22	35.6%	9.1%	Yes	Reject
22	SP	0.40	7.3%	1.8%	No	Retain
23	IG	0.39	21.5%	2.1%	Yes	Reject
24	IG	0.36	19.0%	0.9%	No	Retain
25	EB	0.23	3.3%	3.9%	Yes	Reject
26	EB	0.28	12.7%	0.6%	Yes	Retain
27	OB	0.28	53.8%	1.5%	No	Retain

Important to note, neither the full 27-iem scale nor the 12-item scale exhibited positive correlations with a sum of the five actual logic questions. This provides evidence that the CRT is not a test of intelligence and individuals do not tend to score higher or lower based on their ability to answer real logic questions.

Hypothesis Testing

Correlations with Explicit Measures

Evidence was found in support of Hypothesis 1, which suggested the CRT-RIP (implicit) and explicit measures of risk would have small to no correlation. The 12-item CRT-RIP had a small non-significant correlation with both explicit measures of risk, the risk propensity scale (r =.04) and the DOSPERT (.02).

Correlations with Criterion Measures

Hypothesis 2 posited that the CRT-RIP would be positively correlated with a) risk-taking incidents and b) substance use and/or abuse. Based on the correlations shown in Table 3, the magnitude, direction, and significance of these correlations lend full support for hypothesis 2a and 2b.

Table 3
Means, Standard Deviations, and Correlations Among Study Variables

Variable	M	SD	1	2	3	4	5	6	7
1. CRT-RIP	3.31	1.44							
2. Risk Propensity (Explicit)	3.28	.99	.04						
3. DOSPERT	3.22	.61	.02	.53**					
4. Risk-Taking - 12 Months	.91	1.07	.19**	.26**	.37**				
5. Risk-Taking – Lifetime	1.83	1.37	.24**	.24**	.36**	.72**			
6. Substance - 12 Months	1.88	1.83	.13*	.20**	.32**	.61**	.53**		
7. Substance – Lifetime	2.37	2.03	.16**	.19**	.31**	.55**	.60**	.90**	

Note. CRT-RIP = Conditional Reasoning Test for Risk and Incident Propensity; DOSPERT = Domain-Specific Risk-Taking Scale ; N = 331

*p ≤ .05.

**p ≤ .01.

The 12-item CRT-RIP had correlations of .19 and .24 with reported risk-taking in the past 12 months and in the lifetime, respectively. Additionally, the 12-item CRT-RIP had correlations of .13 and .16 with reported substance use and/or abuse in the past 12 months and in the lifetime, respectively.

Incremental Validity

Hypothesis 3 proposed that the CRT-RIP would better predict risk and incident propensity than the explicit measures of risk propensity alone. As presented in Table 4, the results of a hierarchical regression supported hypothesis 3.

Table 4
Hierarchical Multiple Regression Results

	Risk-Taking - 12 Months				**Risk-Taking - Lifetime**			
Step and IV	β	SE	Total R^2	Δ R^2	β	SE	Total R^2	Δ R^2
Step 1								
Risk Propensity	.248	.057	.065	.065**	.235	.072	.060	.060**
Step 2								
CRT-RIP	.175	.039	.096	.030**	.227	.050	.111	.052**
Step 1								
DOSPERT	.363	.088	.134	.134**	.357	.111	.130	.130**
Step 2								
CRT-RIP	.178	.038	.166	.032**	.230	.048	.172	.053**
	Substance - 12 Months				**Substance - Lifetime**			
Step and IV	β	SE	Total R^2	Δ R^2	β	SE	Total R^2	Δ R^2
Step 1								
Risk Propensity	.192	.100	.039	.039**	.184	.110	.036	.036**
Step 2								
CRT-RIP	.124	.068	.015	.054*	.148	.076	.058	.022**
Step 1								
DOSPERT	.321	.155	.104	.104**	.304	.172	.094	.094**
Step 2								
CRT-RIP	.126	.066	.120	.016*	.150	.073	.116	.022**

*p ≤ .05.
**p ≤ .01.

Addition of the CRT-RIP to the regression was significant over the risk propensity scale and the DOSPERT for both dependent measures, risk-taking incidents and substance use and/or abuse. This is true for both dependent measures when the timeframe was in the past 12 months as well as lifetime.

Factor Structure and Dimensionality

Hypothesis 4 suggested that in an exploratory factor analysis (EFA), the four-factor solution would offer the best fit. To test this hypothesis, the full 27-item scale was used to run the EFA.

Due to the scoring of the items in this test, values of either 1 (biased) or 0 (non-biased) were observed. However, risk propensity should not be considered as an all or nothing phenomena. Additionally, the common factor model, on which factor analysis procedures are based, assumes a linear pattern that is continuous in nature (Hoijtink, Rooks, & Wilmink, 1999). Experts suggest using Weighted Least Squares for Mean and Variance (WLSMV) as the estimator in order to analyze theoretically continuous variables which have been collected as categorical or dichotomous (Beauducel, & Herzberg, 2006). Thus, factor solutions were obtained for 1 through 6 factors using WLSMV estimation with oblique rotation. Oblique, specifically Geomin, rotation was selected because I suspected that the factors would be correlated. Regardless of whether the factors are correlated oblique rotation should not harm results or interpretations (Fabrigar, Wegener, MacCallum, & Strahan, 1999).

To evaluate the model, I first observed eigenvalues greater than one. The analysis produced 13 eigenvalues greater than one, suggesting at most a 13-factor solution. This did not help much in narrowing down the appropriate solution. Next, the scree plot was analyzed (see

Figure 1). The elbow immediately after the four indicates that four or less factors should be retained. After this point, the scree plot experiences a fairly steady decline.

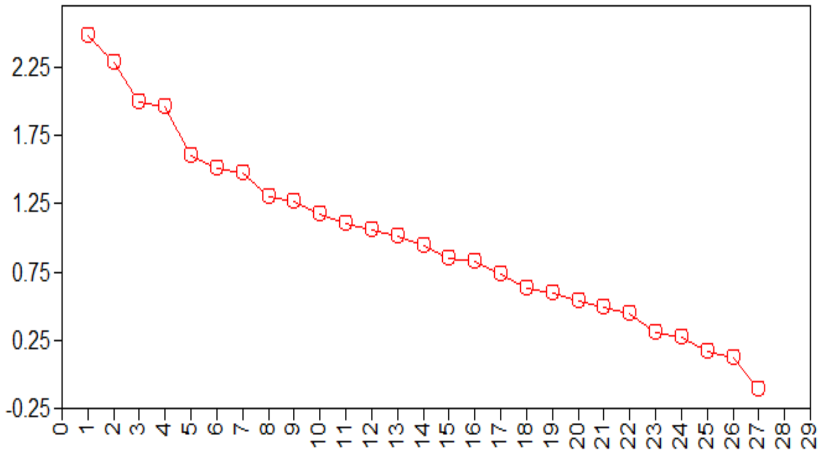

Figure 1. Scree Plot (EFA of 27-Item CRT-RIP)

Using Mplus software, parallel analysis is not an available option when conducting an EFA with categorical data. Unfortunately, this fit index was not obtained. Then, RMSEA values were examined for one to six factors in terms of model fit (see Table 5). The presence of Heywood cases, such as negative residual variances, was also examined. No solutions experienced negative residual variances, and therefore, no solutions could be ruled out on these grounds.

Table 5
Fit Indices for Different Factor Solutions for the 27-Item Scale

Factors	**Chi-square (χ^2, df, p)**	**RMSEA (90% CI)**	**CFI/TLI**	**SRMR**	**Negative Residual variance?**
1	324.16, 324, .49	.001 (0-.021)	1.00/.99	.11	No
2	279.79, 298, .77	0 (.00-.016)	1.00/1.00	.10	No
3	240.38, 273, .92	0 (.00-.01)	1.00/1.00	.09	No
4	205.92, 249, .98	0	1.00/1.00	.08	No
5	180.62, 226, .99	0	1.00/1.00	.08	No
6	157.48, 204, .99	0	1.00/1.00	.07	No

Note. RMSEA = Root Mean Squared Error of Approximation; CFI = Comparative Fit Index; and TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual

Based on prior theory and the fit statistics provided in the EFA, the four-factor solution was deemed most suitable. The four-factor solution had a nonsignificant chi-square [χ^2 (249) = 205.92, p = .98], which indicates the null hypothesis of perfect fit is retained. Additionally, the RMSEA estimate of 0.00 designates that the null hypothesis of close fit should be retained, and that I should reject the null hypothesis of poor fit. The CFI (1.00) and TLI (1.00) indicate very good fit, and the SRMR (.08) indicates fairly small discrepancy between the implied and observed models.

This suggests that a four-factor solution is plausible. However, the base model with one factor also had fairly good item statistics and the fit statistics only get marginally better for each solution following the one-factor solution. As such, hypothesis 4 was only partially supported because although the four-factor solution seems reasonable, other solutions fit nearly as well and offer a more parsimonious alternative. Additionally, some items experienced cross-loading and not all items load onto the appropriate factors (i.e., JMs). The factor loadings are presented in Table 6.

Table 6
Factor Loadings for Four-Factor Solution by JM

CRT Item #	JM	Factor 1	Factor 2	Factor 3	Factor 4
1	EB	-0.144	-0.059	0.114	**0.273***
3	EB	-0.217	-0.184	-0.061	**0.083**
7	EB	-0.025	-0.147	0.082	**0.199**
9	EB	0.168	**0.476***	0.019	0.275
13	EB	-0.064	0.13	**0.309***	-0.055
15	EB	-0.04	0.006	**0.458***	0.121
19	EB	0.134	0.119	**0.244**	0.121
21	EB	-0.053	0	-0.276	**0.406***
25	EB	**0.509***	-0.291	-0.066	0.241
26	EB	**0.384***	0.253	-0.072	-0.196
5	IG	**0.874***	0.012	0.113	-0.046
11	IG	0.169	0.032	0.114	**0.186**
17	IG	-0.025	-0.311*	0.038	**0.212**
18	IG	0.011	**0.687***	-0.042	-0.04
23	IG	0.29	-0.038	0.214	**0.582***
24	IG	-0.11	**0.469***	0.380*	0.009
2	OB	-0.001	**0.208**	0.166	-0.196
6	OB	0.021	-0.245*	0.067	-0.012
8	OB	0.019	-0.184	0.073	-0.350*
12	OB	0.168	-0.268	**0.292**	-0.374
14	OB	0.004	-0.076	-0.294*	**0.11**
20	OB	0	-0.029	**0.580***	-0.002
27	OB	-0.257*	-0.027	**0.123**	0.082
4	SP	-0.310*	0.05	-0.052	-0.059
10	SP	-0.001	-0.111	0.053	-0.393*
16	SP	-0.245*	-0.12	**0.117**	-0.02
22	SP	0.1	**0.294***	0.156	0.056

*$p \leq .05$.

A subsequent EFA was conducted with the 12-item scale. The same estimator and rotation selections were requested for this analysis. Due to the limited number of items, only one- through three-factor solutions could be obtained.

The analysis produced six eigenvalues greater than one, suggesting at most a six-factor solution. Next, the scree plot was analyzed (see Figure 2). The elbow after the two indicates that two or less factors should be retained. After this point, the scree plot experiences a fairly steady decline.

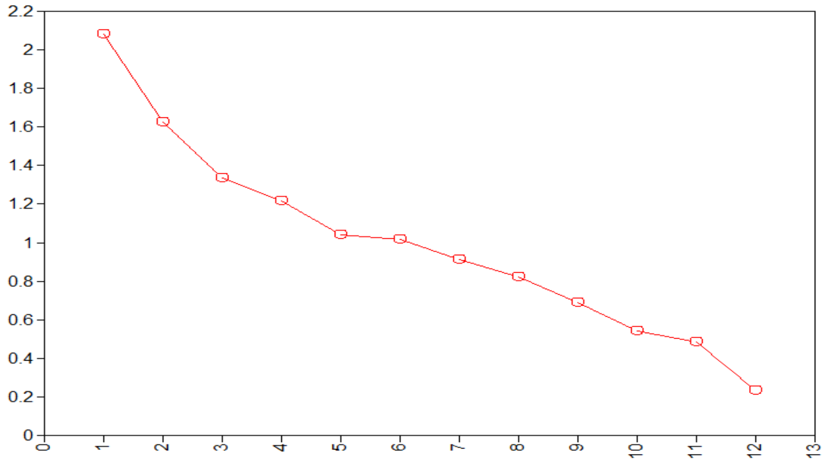

Figure 2. Scree Plot (EFA of 12-Item CRT-RIP)

Then, RMSEA values were examined for one to three factors in terms of model fit (see Table 7). The presence of Heywood cases, such as negative residual variances, was also examined. The three-factor solution experienced negative residual variances, and therefore, the three-factor solution was not found to provide an adequate fit.

Table 7
Fit Indices for Different Factor Solutions for the 12-Item Scale

Factors	**Chi-square (χ^2, df, p)**	**RMSEA (90% CI)**	**CFI/TLI**	**SRMR**	**Negative Residual variance?**
1	53.96, 54, .48	0 (0-.035)	1.00/1.00	.10	No
2	36.90, 43, .73	0 (.00-.028)	1.00/1.00	.08	No
3	22.75, 33, .91	0 (.00-.016)	1.00/1.00	.06	Yes

Note. RMSEA = Root Mean Squared Error of Approximation; CFI = Comparative Fit Index; and TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual

Theoretically, the hope was to find four factors, but with so few remaining items, a four-factor solution could not be produced. With the 12-item EFA, the one-factor model provides adequate fit. The one-factor solution had a nonsignificant chi-square [χ^2 (54) = 53.96, p = .48], which indicates the null hypothesis of perfect fit is retained. Additionally, the RMSEA estimate of 0.00 designates that the null hypothesis of close fit should be retained, and that I should reject the null hypothesis of poor fit. The CFI (1.00) and TLI (1.00) indicate very good fit, and the SRMR (.10) indicates fairly small discrepancy between the implied and observed models. As will be explained in the discussion section, it is likely that EFA is not the best method of evaluating a CRT.

Item Response Theory

In contrast to factor analysis, in which the intent is to explain correlations among indicators, IRT takes into consideration the relationship between a person's item responses with regard to both the level of latent trait and the item properties (Embretson & Reise, 2000). In an IRT model, latent trait level (denoted θ in the IRT literature) can be estimated with one, two, or three item parameters.

To test hypothesis 4, items in the CRT-RIP were fitted using a Two-Parameter Logistic (2PL) model. The 2PL model was deemed appropriate as this model is best-suited for measures in which the individual items may not be equivalent in measurement of the latent trait (Embretson & Reise, 2000). As demonstrated by the range of p-values across items, the items in the CRT-RIP do not each measure the latent trait equally. The 2PL model considers the CRT-RIP to be a dichotomously scored measure. This is suitable given that respondents rarely select the distractor answer choices, and that Conditional Reasoning Tests which seek to measure only one trait are typically scored dichotomously for analysis (James & LeBreton, 2011). Furthermore, evidence suggests that the 2PL model is the most parsimonious IRT model for the analysis of Conditional Reasoning Tests (DeSimone, 2012).

The 2PL model predicts the probability of endorsing a particular response (i.e., a correct response) on an item using the latent trait (θ) and two item parameters. The first parameter is *item difficulty,* which is typically denoted either as b or β in the IRT literature. The item difficulty parameter (*b*) characterizes the position of the curve along the horizontal axis of the ICC. This means that the "easier" items have lower *b* values and are characterized by curves nearer to the horizontal axis and has an inverse relationship with the proportion-correct score (*p)*, or the proportion of items endorsed if the symptom, trait, or attitude is present versus absent. Item difficulties have also been called *item threshold* or *item location* parameters. The item difficulty parameters correspond to item thresholds (τ) in CFA with categorical outcomes (Brown, 2006). The second parameter is *item discrimination* (α). In terms of the ICC, the discrimination parameters influence the steepness of the slope with steeper slopes representing high discrimination value and a stronger relationship to the latent variable (θ; Brown, 2006).

First, the 27-item CRT-RIP was fitted to the 2PL model. This 2PL model showed good overall fit with a Root Mean Square Error of Approximation (RMSEA) of 0.00. Embretson and Reise (2000) have suggested that a RMSEA of <.05 indicates good fit whereas a RMSEA of >.1 indicates poor fit in IRT. Though there are no specified cut-off values, lower Akaike's information criteria (AIC; Akaike, 1974) and Schwarz's Bayesian information criteria (BIC; Schwarz, 1978) values are indicative of better fit. These fit indices provide estimates of relative differences between models, but do not allow for significance testing (Kang, Cohen, & Sung, 2009). The 27-item CRT-RIP fitted to a 2PL model demonstrated an AIC of 9,014.59, and a BIC of 9,219.74.

Each item was individually reviewed for item quality. Upon this closer examination of the item characteristic curves (ICC) and validity coefficients for each item, many items showed poor fit and/or a lack of validity. These poor fitting items showed negative slopes or reflected standard error margins that contained zero.

A second 2PL model was applied to the 12-item final CRT-RIP measure. The 12-item IRT 2PL model exhibited an overall better fit with a RMSEA of 0.00, an AIC of 3,746.94, and a BIC of 3,838.19. However, the AIC and BIC values may be lower due to having less parameters to be estimated rather than being indicative of better fit because the items were superior. A comparison of the fit indices with the 12-item scale to those of the 27-item scale is presented as Table 8.

Table 8
2PL Model Fit Statistics

Model	M_2	df	p-value	RMSEA	-2lnL	AIC	BIC
2PL of 27-item CRT	315.62	324	0.62	0.00	8906.59	9014.59	9219.74
2PL of 12-item CRT	47.49	54	0.72	0.00	3698.94	3746.94	3838.19

Note. RMSEA = Root Mean Squared Error of Approximation; AIC = Akaike's information criteria; and BIC = Schwarz' Bayesian information criteria

Table 9 summarizes the 2PL difficulty, discrimination, and Pearson S-χ^2 index of item fit for both the 27-item and 12-item versions of the CRT-RIP. Because the previous correlation and regression analyses utilized the 12-item CRT-RIP and because the model with shortened scale seems to offer better fit, I will further explore item characteristics from the 2PL IRT of the 12-item scale.

Table 9
2PL Model Item Parameter Estimates

	27-Item CRT-RIP			12-Item CRT-RIP		
CRT Item	α (SE)	β (SE)	S-χ^2 (df)	α (SE)	β (SE)	S-χ^2 (df)
1	0.38 (0.29)	1.50 (1.16)	11.27 (9)	--	--	
2	-0.54 (0.23)	-2.34 (0.92)	8.82 (10)	0.47 (0.37)	2.64 (1.99)	5.44 (5)
3	0.54 (0.28)	-1.80 (0.88)	17.67 (10)	-0.29 (0.31)	3.12 (3.24)	3.04 (5)
4	0.15 (0.26)	-10.85 (18.31)	11.68 (9)	--	--	
5	-0.62 (0.36)	-3.50 (1.80)	8.45 (8)	--	--	
6	0.33 (0.24)	-5.10 (3.54)	9.19 (9)	--	--	
7	0.37 (0.26)	-1.06 (0.78)	4.59 (10)	--	--	
8	-0.10 (0.22)	1.64 (3.89)	12.98 (9)	-0.19 (0.20)	0.79 (0.98)	1.66 (5)
9	-0.50 (0.36)	-2.09 (1.43)	13.22 (10)	0.56 (0.25)	1.89 (0.81)	5.18 (5)
10	-0.21 (0.31)	1.74 (2.57)	11.81 (10)	--	--	
11	-0.03 (0.25)	-25.23 (204.54)	11.09 (10)	--	--	
12	0.11 (0.79)	39.12 (287.48)	5.62 (2)	-0.61 (1.31)	-7.11 (14.07)	1.20 (1)
13	-0.38 (0.38)	-7.37 (7.09)	13.04 (7)	0.49 (0.54)	5.87 (6.05)	4.45 (4)
14	0.33 (0.26)	-4.33 (3.39)	2.99 (9)	--	--	
15	-0.02 (0.24)	-12.59 (199.11)	16.04 (10)	--	--	
16	0.27 (0.20)	1.67 (1.25)	13.87 (10)	-0.25 (0.21)	-1.77 (1.52)	7.70 (5)
17	0.71 (0.28)	1.00 (0.39)	15.17 (10)	--	--	
18	-1.10 (0.45)	-1.86 (0.55)	8.74 (8)	2.21 (4.56)	1.29 (0.92)	0.17 (4)
19	-0.21 (0.24)	0.81 (1.06)	12.35 (10)	--	--	
20	-0.18 (0.42)	-14.83 (34.13)	3.48 (7)	--	--	
21	0.46 (0.24)	1.34 (0.73)	12.74 (10)	--	--	
22	-0.56 (0.38)	-4.81 (3.04)	8.31 (7)	0.88 (0.57)	3.25 (1.74)	1.47 (4)
23	0.26 (0.41)	5.06 (7.78)	18.89 (10)	--	--	
24	-0.69 (0.47)	-2.29 (1.36)	8.39 (9)	0.71 (0.56)	2.25 (1.48)	1.60 (4)
25	0.32 (0.54)	10.54 (16.92)	2.15 (5)	--	--	
26	-0.88 (0.65)	-2.48 (1.33)	5.24 (8)	0.72 (1.11)	2.94 (3.67)	8.00 (4)
27	0.24 (0.28)	-0.63 (0.81)	8.17 (11)	-0.23 (0.36)	0.66 (0.91)	2.46 (5)

Note. α = Item discrimination; β = Item difficulty; S-χ^2 = Pearson χ^2 index of item fit
*$p \leq .05$.

Item Characteristics of the CRT-RIP. S-χ^2 values reveal that none of the items in the 2PL with the 12-item CRT demonstrate misfit at the item level. In partial support of hypothesis 5, seven of the 12 discrimination parameters are positive, statistically different from zero, and monotonically increase as a function of their respective latent traits. Items 2, 9, 18, 22, 24, and

26 all showed positive slopes and varying difficulty levels. ICCs for these items are presented in Figures 3-9. Positive slopes indicate that these items of the CRT-RIP work as theoretically intended, whereas the other five items may not be as sufficient in measuring the latent trait. Discrimination parameters exhibited a range of -.61 (item 12) to 2.21 (item 18). Because the 2PL IRT model is a dichotomous model, the slope parameter signifies endorsement of the risky response option, and does not include analysis of the non-risky or illogical response options. Item difficulties ranged from -7.11 (item 5) to 5.87 (item 6). Ten of the 12 difficulty parameters are positive and eight of the 12 difficulty parameter estimates reach a maximum value above the $\theta = 1$ level.

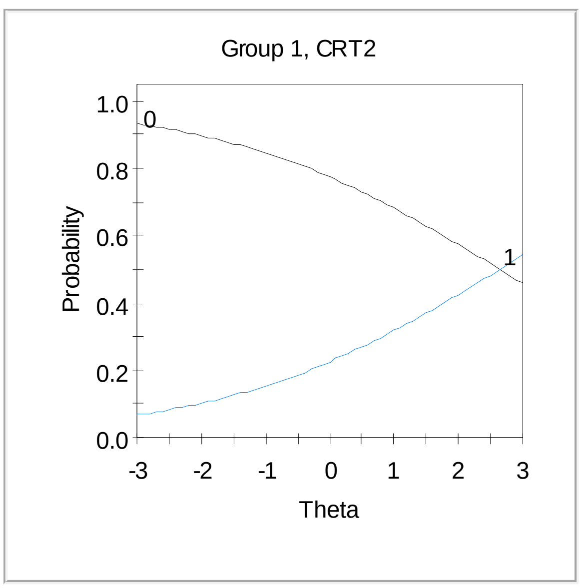

Figure 3. Item Characteristic Curve for CRT-RIP Item 2

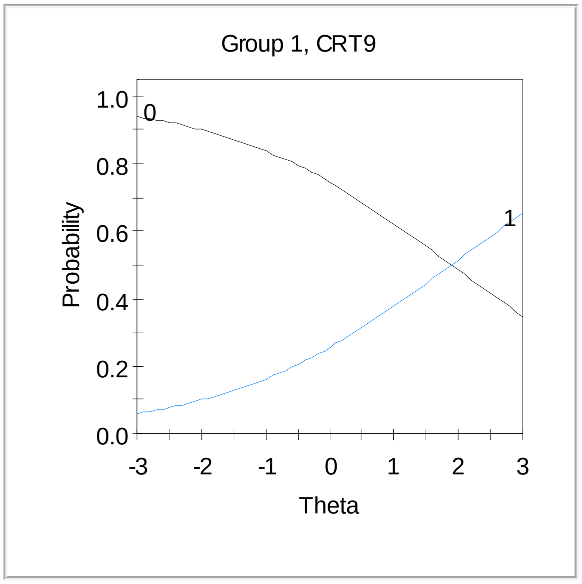

Figure 4. Item Characteristic Curve for CRT-RIP Item 9

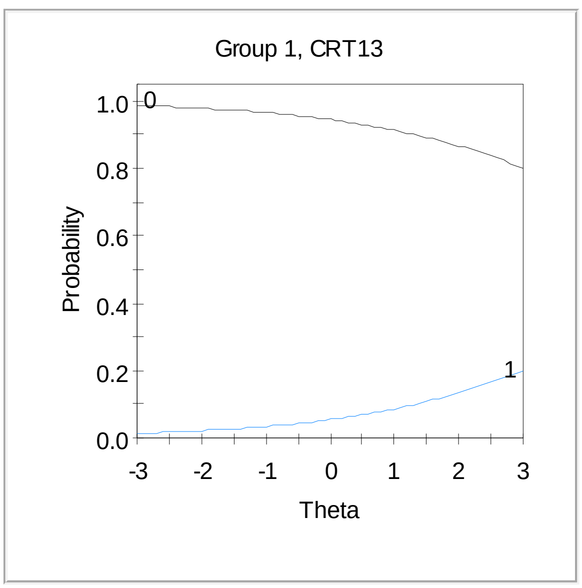

Figure 5. Item Characteristic Curve for CRT-RIP Item 13

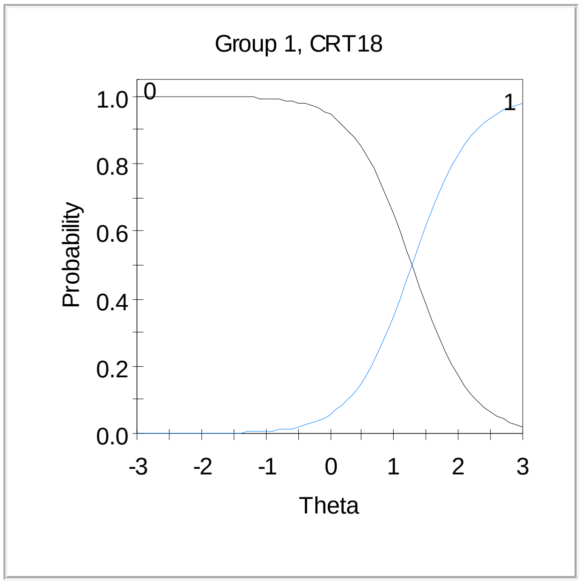

Figure 6. Item Characteristic Curve for CRT-RIP Item 18

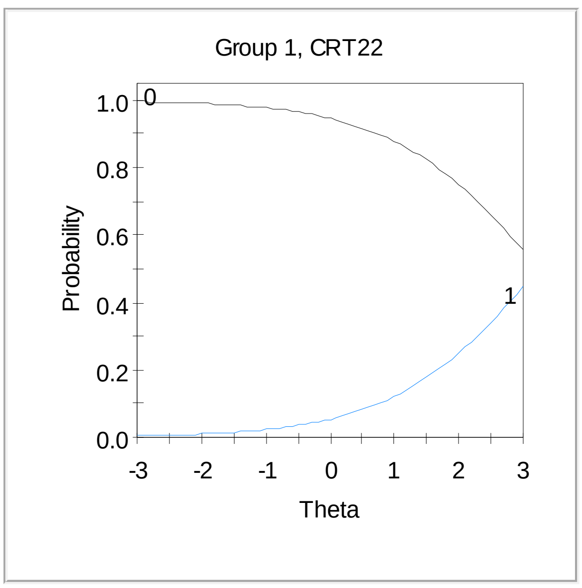

Figure 7. Item Characteristic Curve for CRT-RIP Item 22

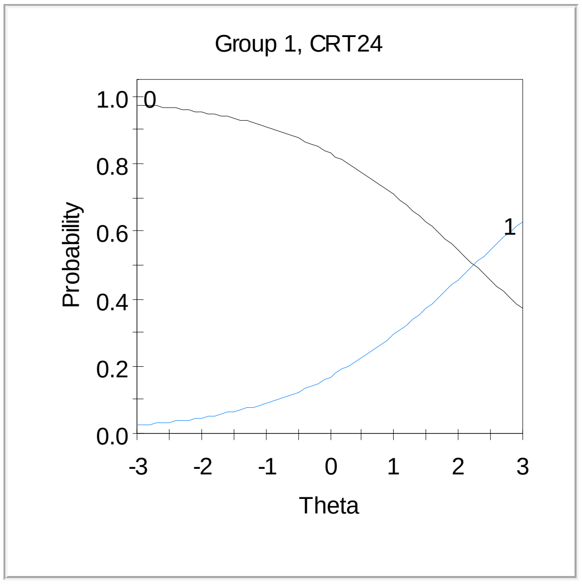

Figure 8. Item Characteristic Curve for CRT-RIP Item 24

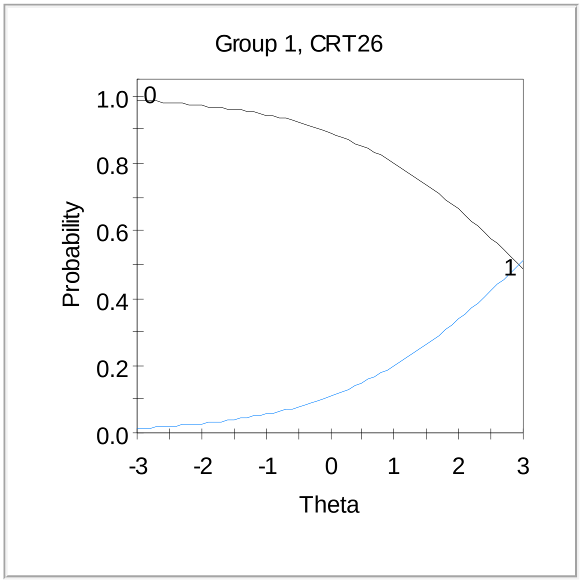

Figure 9. Item Characteristic Curve for CRT-RIP Item 26

CHAPTER 6

DISCUSSION

The small correlations between the CRT-RIP and the explicit measures of risk suggest that explicit risk and implicit risk are related yet distinct constructs. To a great extent, individuals like to believe that they have more favorable characteristics and few unfavorable characteristics. The ability to justify risk implicitly yet not claim to be risky on explicit measures is evident based on these very low correlations.

Moreover, the modest correlations found between the CRT-RIP and risk-taking incidents, as well as between the CRT-RIP and substance use and/or abuse suggest that there is a relationship between higher scores on the CRT-RIP and higher prevalence of risky behavior. Although the DOSPERT and risk propensity scale experienced slightly higher correlations with the risk-taking incidents and substance use and/or abuse, this was not what would be considered a 'high-stakes' situation such as an employment interview or performance appraisal context (Schmit & Ryan, 1993). Thus, the CRT-RIP may have application in selection contexts or contexts where faking or response distortion is expected. Assuming that the true nature of the test remains concealed, CRTs are typically not prone to faking or response distortion (LeBreton et al., 2007). The CRT-RIP could also be used to supplement explicit measures. The CRT-RIP offers additional and unique information. Bing and colleagues (2007) argue for integration of the measurement and use of implicit and explicit personality.

The notion that the CRT-RIP has the capacity to offer additional information is further substantiated by the multiple hierarchal regressions. In each instance, the CRT-RIP was found to offer additional information regarding risk propensity above the other (explicit) measures. These hierarchical regressions evidenced that the CRT-RIP has incremental validity and thus, potentially added value in predicting safety (or risk) behavior.

Some empirical support was found for a four-factor solution. However, this is not the only solution that exhibited decent properties. A more parsimonious model with fewer factors may be preferred. For example, the one-factor model demonstrated acceptable fit as well. Furthermore, there were many items which cross-loaded onto more than one factor. Some cross-loading was expected because the factors are not entirely conceptually disparate. For this reason and as mentioned above, oblique rotation was used to allow the factors to correlate with one another. If the factors had not been allowed to correlate with one another (i.e., if orthogonal rotation was used), cross-loading likely would have been much more prevalent. Another problematic finding was that some items did not load on the factor that they were theoretically expected to load most highly on. This may have occurred because of the similarity among the four theoretical factors (JMs).

Previously, researchers have suggested that CRTs may not be conducive to factor analysis (James & LeBreton, 2012). Each CRT item is intended to prompt the use of several different JMs, and consequently, the items often cross-load on several factors. The CRT-RIP may have been especially difficult to extract factors from due to the complexities of the risk construct. Risk propensity can pertain to a number of domains, such as those included in the DOSPERT (i.e., health and safety, social, ethical, financial, and recreational) and can differ based on contexts, such as when you are at work as opposed to in your car. Additionally, risk can refer to approach behavior, such as going skydiving, or avoidance behavior, such as not putting on sunscreen before going out in the sun. For the reasons outlined above, a clear and interpretable factor structure may not exist.

IRT was used to further examine item characteristics. The 2PL IRT offered greater insight into the functioning of individual items. The 2PL model offered information regarding item difficulty and the ability of individual items to discriminate between someone who in the latent trait of riskiness. In this analysis, a few items in particular emerged as high-performing items in terms of measuring risk propensity. The overall model fit was better for the 12-item CRT-RIP in comparison to the 27-item scale. However, it should be noted that, with less parameters to be estimated, lower AIC and BIC are expected. The fit of this model also may have suffered due to the varying contexts of risk and as a result of the avoid-approach behavioral element as to why individuals are not safe.

Limitations

One major limitation in this study was that performance measures were not collected, and predictive validity could not be verified. The present study only established what would better be described as concurrent or postdictive validity. A second limitation is that the sample was taken from a student population. There may be concerns regarding how well the results will translate to a working population, and ultimately, how generalizable these results are. The results are expected to generalize, but this notion should be corroborated with a different sample of participants.

Practical Implications

Injuries and fatalities in the workplace are costly both in terms of human capital and financial considerations. Increasing concern for and awareness of safety is a high priority of many organizations today. In numerous occupations, safety compliance and risk avoidance is a large aspect of the job. However, it has been demonstrated that incident involvement does not appear to be the same across all individuals. This suggests that some combination of traits may contribute to an individual's propensity to behave in a risky manner and be able to justify risky behavior. The CRT-RIP was designed under the assumption that individuals vary in the ways that they assess and think about risk in their everyday lives. If the CRT-RIP was revised and refined so as to increase predictive and incremental validity with regard to risk, this would be a huge win for organizations that regularly hire and fill safety-relevant jobs. Currently, there are scant methods by which organizations can attempt to measure individuals risk propensity aside from using either incredibly transparent and obvious test batteries or costly and involved simulations. The CRT-RIP aims to offer an alternative to these less than ideal approaches to selection and employment testing to screen out individuals who are likely to behave more risky. It may also have applications for measuring cognitions about risk before and after a safety training intervention. Much work is still to be done to perfect the technology of CR and to improve this specific application of the testing method. Nonetheless, the CRT-RIP shows promising potential to meet various employers' current and future needs.

Future Directions

Overall, the CRT-RIP demonstrated successful measurement of risk and incident propensity. However, many items were rejected due to poor statistical properties. In future iterations to this CRT, many items will be revised and new items will be developed in hopes of attaining a test with better predictive and incremental validity. In response to the limitation regarding not having performance measures, future data collection will involve information concerning self-reported safety behavior at work as well as supervisor performance ratings that will be used to assess predictive validity. The next iteration will also be compared with personality measures to verify incremental validity above the Big Five personality traits. To address the question of generalizability, data will be collected from working adults in a future sample.

Conclusion

This scale was an initial attempt to measure risk propensity using a conditional reasoning approach. The scale will likely need to go through a few more iterations before being finalized. In the next iteration to this test, a major objective will be to retain many items that have improved psychometric properties. Specifically, items will be developed or modified in order to address the concerns and poor item statistics that were demonstrated in this initial attempt to validate the scale. Moreover, this scale will ideally exhibit high predictive validity of individual risk propensity.

Establishing predictive validity of the CRT-RIP is the primary goal in moving forward. If the next version of the test exhibits high predictive validity, it may very well have implications in a number of organizational settings. Much work still needs to be done to revise and improve the functioning of individual items as well as the test as a whole. However, a test that measures risk propensity with low transparency could be exceptionally valuable in a number of contexts including the workplace.

References

Anderson, G. M., & Lorber, R. L. (2006). *Safety 24/7: Building an Incident-Free Culture*. Lafayette, LA: Results in Learning.

Arnold, P. K., Hartley, L. R., Corry, A., Hochstadt, D., Penna, F., & Feyer, A. M. (1997). Hours of work and perceptions of fatigue among truck drivers. *Accident Analysis and Prevention*, *29*, 471–477.

Babakus, E., Ferguson, C. E., & Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *Journal of Marketing Research, 37,* 72–141.

Bartholomew, D. J., Steele, F., Galbraith, J., & Moustaki, I. (2008). *Analysis of multivariate social science data*. London: CRC press.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Journal of the American Psychological Society*, *4*, 2-44.

Baumeister, R. F., Smart, L., & Boden, J. M. (1996). Relation of threatened egotism to violence and aggression: The dark side of high self-esteem. *Psychological Review*, *103*, 5-33.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of Maximum Likelihood versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling*, *13*, 186-203.

Bern, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1-62). New York, NY: Academic Press.

Bersoff, D. M. (1999).Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Society for Personality and Social Psychology*, *25*, 28-39.

Beus, J. M., Dhanani, L. Y., & McCord, M. A. (2015). A meta-analysis of personality and workplace safety: Addressing unanswered questions. *Journal of Applied Psychology*, *100*, 481-498.

Bing, M. N., LeBreton, J. M., Davison, H. K., Migetz, D. Z., & James, L. R. (2007). Integrating implicit and explicit social cognitions for enhanced personality assessment: A general framework for choosing measurement and statistical methods. *Organizational Research Methods*, *10*, 346–389.

Bock, R. D. (1997). A brief history of item response theory. *Educational Measurement: Issues and Practice*, *16*, 21–33.

Bock, R. D., Gibbons, R., & Muraki, E. J. (1988). Full information item factor analysis. *Applied Psychological Measurement*, *12*, 261–280.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Borschmann, R., Lines, K., & Cottrell, D. (2012). Sun protective behaviour, optimism bias, and the transtheoretical model of behaviour change. *Australian Journal of Psychology*, *64*, 181-188.

Bowler, J. L., Bowler, M. C., & James, L. R. (2011). The cognitive underpinnings of addiction. *Substance use & misuse*, *46*, 1060-1071.

Brache, K. & Stockwell, T. (2011). Drinking patterns and risk behaviors associated with alcohol and energy drink consumption in college drinkers. *Addictive Behaviors, 36*, 1133-1140.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of the covariance structures. *British Journal of Mathematical & Statistical Psychology*, *37*, 62–83.

Bureau of Labor Statistics, U.S. Department of Labor. (2016). *Industry injury and illness data.* Retrieved May 3, 2016, from http://www.bls.gov/iif/

Burke, M. J., Sarpy, S. A., Tesluk, P. E., & Smith-Crowe, K. (2002). General safety performance: A test of a grounded theoretical model. *Personnel Psychology, 55,* 429–457.

Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 6, pp. 94-172). New York: McGraw-Hill.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt &W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass.

Caponecchia, C. (2010). It won't happen to me: An investigation of optimism bias in occupational health and safety. *Journal of Applied Social Psychology*, *40*, 601-617.

Cheng, Y., Shein, P. P., & Chiou, W. (2012). Escaping the impulse to immediate gratification: The prospect concept promotes a future-oriented mindset, prompting an inclination towards delayed gratification. *British Journal of Psychology*, *103*, 129-141.

Cherukupalli, R. (2010). A behavioral economics perspective on tobacco taxation. *American Journal of Public Health, 100*, 609–615.

Christian, M. S., Bradley, J. C., Wallace, J. C., & Burke, M. J. (2009). Workplace safety: A meta-analysis of the roles of person and situation factors. *Journal of Applied Psychology*, *94*, 1103-1127.

Cialdini, R. B. (1993). *Influence: Science and practice (3rd Edition).* New York: HarperCollins College Publishers.

Cigularov, K. P., Chen, P. Y., & Stallones, L. (2009). Error communication in young farm workers: Its relationship to safety climate and safety locus of control. *Work & Stress*, *23*, 297-312.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression / correlation analysis for the behavioral sciences* (3rd edition). Mahwah, NJ: Lawrence Erlbaum Associates.

Conway, H., & Svenson, J. (1998). Occupational injury and illness rates, 1992-96: why they fell. *Monthly Lab. Rev.*, *121*, 36.

Costa-Font, J., Mossialos, E., & Rudisill, C. (2009). Optimism and the perceptions of new risks. *Journal of risk research*, *12*, 27-41.

Cramer, P. (1998) Defensiveness and defense mechanisms. *Journal of Personality*, *66*, 880-894.

Cramer, P. (2000). Defense mechanisms in psychology today. *American Psychologist*, *55*, 637-646.

Daugherty, J. R., & Brase, G. L. (2010). Taking time to be healthy: Predicting health behaviors with delay discounting and time perspective. *Personality and Individual Differences, 48*, 202–207.

DeSimone, J. A. (2012). *An analysis of the item characteristics of the Conditional Reasoning Test of Aggression*. (Dissertation), Georgia Institute of Technology.

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The journal of abnormal and social psychology*, *51*, 629.

Dittmar, H., & Bond, R. (2010). I want it and I want it now: Using a temporal discounting paradigm to examine predictors of consumer impulsivity. *British Journal of Psychology, 101*, 751–776.

Driskell, J. E., Goodwin, G. F., Salas, E., & O'Shea, P. (2006). What makes a good team player? Personality and team effectiveness. *Group Dynamics: Theory, Research, and Practice*, *10*, 249-271.

Dunnett, S., Koun, S., & Barber, P. J. (1981). Social desirability in the Eysenck Personality Inventory. *British Journal of Psychology, 72,* 19-26.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, *95*, 256–273.

Eatough, E. M. & Spector, P. E. (2014). *The role of workplace control in positive health and wellbeing.* In C. Cooper (Ed). *Wellbeing: A Complete Reference Guide*. Oxford, England: Wiley-Blackwell.

Ellis, A. P. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type. *Journal of Applied Psychology, 87,* 1200–1208.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.

Enders, C. K. (2010) *Applied missing data analysis*. New York: Guilford Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*, 272-299.

Fan, J., Gao, D., Carroll, S. A., Lopez, F. J., Tian, T. S., & Meng, H. (2012). Testing the efficacy of a new procedure for reducing faking on personality tests within selection contexts. *Journal of Applied Psychology*, *97*, 866-880.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Attitudes and social cognition. *Journal of Personality and Social Psychology*, *69*, 1013–1027.

Ferrando, P. J., & Lorenza-Sevo, U. (2005). IRT-related factor analytic procedures for testing the equivalence of paper-and-pencil and Internet-administered questionnaires. *Psychological Methods*, *10*, 193–205.

Fine, S., & Gottlieb-Litvin, Y. (2013). Justifying Counterproductive Work Behaviors and an Integrity-based Conditional Reasoning Test: Back to the drawing board?. *International Journal of Selection & Assessment*, *21*, 328-333.

Flora, D., & Curran, P. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*, 466–491.

Forcier, B. H., Walters, A. E., Brasher, E. E., & Jones, J. W. (2001). Creating a safer work environment through psychological assessment: a review of a measure of safety consciousness. In Stuhlmacher, A. & Cellar, D. (Eds.), *Workplace Safety: Individual Differences in Behavior* (pp. 53–65) Chicago, IL: Hayworth Press.

Frese, M. (1995). Error management in training: Conceptual and empirical results. In C. Zucchermaglio, S. Bagnara & S.U. Stuchy (Eds.), *Organizational learning and technological change* (pp. 112-124). New York, NY: Springer.

Frone, M. R. (1998). Predictors of work injuries among employed adolescents. *Journal of Applied Psychology*, *83*, 565.

Frost, B. C., Ko, C. H. E., & James, L. R. (2007). Implicit and explicit personality: A test of a channeling hypothesis for aggressive behavior. *Journal of Applied Psychology*, *92*, 1299–1319.

Furnham, A, & Craig, S. (1987). Fakability and correlates of the perception and preference inventory. *Personality and Individual Differences, 8,* 459-470.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Glöckner-Rist, A., & Hoijtink, H. (2003). The best of both worlds: Factor analysis of dichotomous data using item response theory and structural equation modeling. *Structural Equation Modeling, 10*, 544–565.

Green, S. B., Akey, T. M., Fleming, K. K., Hershberger, S. L., & Marquis, J. G. (1997). Effect of the number of scale points on chi-square fit indices in confirmatory factor analysis. *Structural Equation Modeling, 4,* 108–120.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*, 4-27.

Greenwood, M. & Woods, H.M. (1919). The incidence of industrial accidents upon individuals with special reference to multiple accidents. *Industrial Fatigue Research Board, Medical Research Committee*, Report No. 4. Her Majesty's Stationery Office, London.

Haidt, J. P. (2001). The emotional dog and its rational tail: A social instuitionist approach to moral judgement. *Psychological Review, 108*, 814-834.

Hansen, C. P. (1989). A causal model of the relationship among accidents, biodata, personality, and cognitive factors. *Journal of Applied Psychology, 74,* 81–90.

Harris, P. R., Griffin, D. W., & Murray, S. (2008). Testing the limits of optimistic bias: Event and person moderators in a multilevel framework. *Journal of Personality and Social Psychology*, *95*, 1225.

Hemenway, D. (2013). Three common beliefs that are impediments to injury prevention. *Injury prevention*, *19*, 290-293.

Henning, J. B., Stufft, C. J., Payne, S. C., Bergman, M. E., Mannan, M. S., & Keren, N. (2009). The influence of individual differences on organizational safety attitudes. *Safety Science*, *47*, 337-345.

Hinrichsen, J. J, Gryll, S. L, Bradley, L. A, & Katahn. M. (1975). Effects of impression management efforts on FIRO-B profiles. *Journal of Consulting and Clinical Psychology 43,* 269.

Hofmann, D.A., & Mark, B. (2006). An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Personnel Psychology, 59,* 847-869.

Hough, L. M., Eaton, N. K., Dunnette, M. D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*, 581-595.

Hunter, D. R., & Stewart, J. E. (2012). Safety locus of control and accident involvement among army aviators. *The International Journal of Aviation Psychology*, *22*, 144-163.

Iverson, R. D., & Erwin, P. J. (1997). Predicting occupational injury: The role of affectivity. *Journal of Occupational and Organizational Psychology, 70*, 113–128.

James, L. R. (1998). Measurement of personality via conditional reasoning. *Organizational Research Methods*, *1*, 131-163.

James, L. R., & LeBreton, J. M. (2011). *Assessing the implicit personality through conditional reasoning*. Washington, D.C.: American Psychological Association.

James, L. R., & Mazerolle, M. D. (2002). *Personality in work organizations.* Thousand Oaks, CA: Sage.

James, L. R., & McIntyre, M. D. (2000). *Conditional Reasoning Test of Aggression test manual.* San Antonio, TX: Psychological Corporation.

James, L. R., McIntyre, M. D., Glisson, C. A., Bowler, J. L., & Mitchell, T. R. (2004). The conditional reasoning measurement system for aggression: An overview. *Human Performance, 17,* 271–295.

James, L. R., McIntyre, M. D., Glisson, C. A., Green, P. D., Patton, T. W., LeBreton, J. M., & ... Williams, L. J. (2005). A Conditional Reasoning Measure for Aggression. *Organizational Research Methods*, *8*, 69-99.

Jones, J. W. & Wuebker, L. J. (1993). Safety locus of control and employees' accidents. *Journal of Business and Psychology, 7*, 449–457.

Jones, J. W., & Wuebker, L. J. (1985). Development and validation of the safety locus of control scale. *Perceptual and Motor Skills*, *61*, 151-161.

Joseph, C., Reddy, S., & Kashore Sharma, K. (2013). Locus of control, safety attitudes and involvement in hazardous events in Indian army aviators. *Aviation Psychology and Applied Human Factors*, *3*, 9-18.

Kang, T., Cohen, A. S., & Sung, H. (2009). Model selection indices for polytomous items. *Applied Psychological Measurement*, *33*, 499-518.

Kilstrom, J. F. (1999). The psychological unconscious. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 424-442). New York, NY: Guilford.

Klen, T., & Ojanen, K. (1998). The correspondence of self-reported accidents with company records. *Safety science*, *28*, 45-48.

Kline, R. B. (2010). *Principles and Practice of Structural Equation Modeling: Third Edition*. New York, NY: The Guilford Press.

Kouabenan, D. R. (1998). Beliefs and the perception of risks and accidents. *Risk Analysis, 18*, 243–252.

Krebs, D. L., & Denton, K. (1997). Social illusions and self-deception: The evolution of biases in person perception. In J. A. Simpson, D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 21-48). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Leary, M. R., Tchividijian, L. R., & Kraxberger, B. E. (1994). Self-presentation can be hazardous to your health: Impression management and health risk. *Health Psychology*, *13*, 461-470.

LeBreton, J. M., Barksdale, C. D., Robin, J., & James, L. R. (2007). Measurement issues associated with conditional reasoning tests: Indirect measurement and test faking. *Journal of Applied Psychology*, *92*, 1-16.

Leigh, J. P. (2011). Economic burden of occupational injury and illness in the United States. *Milbank Quarterly*, *89*, 728-772.

Levashina, J., & Campion, M. A. (2007). Measuring faking in the employment interview: Development and validation of an interview faking behavior scale. *Journal of Applied Psychology, 92,* 1638–1656.

Levenson, H. (1974). Activism and powerful others: Distinctions within the concept of internal-external control. *Journal of Personality Assessment*, *38*, 377-383.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*, 267-286.

Lord, F. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

MacIntosh, R., & Hashim, S. (2003). Variance estimation for converting MIMIC model parameters to IRT parameters in DIF analysis. *Applied Psychological Measurement*, *27*, 372–379.

Mark, B.A., Hughes, L.C., Belyea, M., Chang, Y., Hofmann, D., Jones, C.B., & Bacon, C.T. (2007). Does safety climate moderate the influence of staffing adequacy and work conditions on nurse injuries? *Journal of Safety Research, 38,* 431-446.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review, 96,* 690–702.

Meertens, R. M., & Lion, R. (2008). Measuring an individual's tendency to take risks: The risk propensity scale. *Journal of Applied Social Psychology*, *38*, 1506-1520.

Mersman, J.L., & Shultz, K.S. (1998). Individual differences in the ability to fake on personality measures. *Personality and Individual Differences, 24*, 217–227.

Michel, J. S., Pace, V. L., Edun, A., Sawhney, E., & Thomas, J. (2014). Development and validation of an explicit aggressive beliefs and attitudes scale. *Journal of Personality Assessment*, *96*, 327-338.

Mischel,W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*, 246–268.

Moore, K. D. (1998). *Patterns of inductive reasoning: Developing critical thinking skills* (4th ed.). Dubuque, IA: Kendall/Hunt Publishing.

Moustaki, I., Jöreskog, K. G., & Mavridis, D. (2004). Factor models for ordinal variables with covariate effects on the manifest and latent variables: A comparison of LISREL and IRT approaches. *Structural Equation Modeling*, *11*, 487–513.

Murray, M., Fitzpatrick, D., & O'Connell, C. (1997). Fisherman's blues: factors related to accidents and safety among Newfoundland fisherman. *Work and Stress, 11*, 292–297.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript.

Muthén, B. O., Kao, C., & Burstein, L. (1991). Instructional sensitivity in mathematics achievement test items: Applications of a new IRT-based detection technique. *Journal of Educational Measurement, 28*, 1–22.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor-analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology*, *45*, 19–30.

Neal, A., & Griffin, M. A. (2004). Safety climate and safety at work. In J. Barling & M. R. Frone (Eds.), *The psychology of workplace safety* (pp. 15-34). Washington, DC: American Psychological Association.

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety Science, 34,* 99-109.

Nieminen, L. R. (2012). *The development and validation of a conditional reasoning test of withdrawal.* (Dissertation), Wayne State University.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of human judgment.* Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Oliver, A., Cheyne, A., Tomás, J. M., & Cox, S. (2002). The effects of organizational and individual factors on occupational accidents. *Journal of Occupational and Organizational Psychology*, 75, 473-488.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology*, *60*, 995-1027.

Parker, D. L., Carl, W. R., French, L. R., & Martin, F. B. (1994). Characteristics of adolescent work injuries reported to the Minnesota Department of Labor and Industry. *American Journal of Public Health*, *84*, 606-611.

Parker, S. K., & Turner, N. (2002). Work design and individual work performance: Research findings and an agenda for future inquiry. In S. Sonnentag (Ed.), *Psychological management of individual performance* (pp. 69–93). Chichester, England: Wiley.

Pauls, C. A., & Crost, N. W. (2005). Cognitive Ability and Self-Reported Efficacy of Self-Presentation Predict Faking on Personality Measures. *Journal of Individual Differences*, *26*, 194-206.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Reason, J. (1990). *Human error*. New York: Cambridge University Press.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114*, 552–566.

Roe-Burning, S., & Straker, G. (1997). The association between illusions of invulnerability and exposure to trauma. *Journal of Traumatic Stress*, *10*, 319-327.

Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83,* 634–644.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement: *Psychological Monographs: General & Applied, 80*, 1–28.

Rundmo, T., & Hale, A. R. (2003). Manager's attitudes towards safety and accident prevention. *Safety Science*, *41*, 557-574.

Salgado, J., Viswesvaran, C., & Ones, D.S. (2001). Predictors used for personnel selection: An overview of constructs, methods, and techniques. In N. Anderson, D.S. Ones, H.K. Sinangil & C. Viswesvaran (Eds.), *Handbook of Industrial, Work, and Organizational Psychology: Personnel Psychology* (pp. 165-199). Thousand Oaks, CA: Sage.

Samreen, H., & Zubair, A. (2013). Locus of control and death anxiety among police personnel. *Pakistan Journal of Psychological Research*, *28*, 261-275.

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling*, *21*, 167-180.

Schwab, D. P. (1971). Issues in response distortion studies of personality inventories: A critique and replicated study. *Personnel Psychology, 24,* 637-647.

Schoen, J. L., Bowler, J. L., & Schilpzand, M. C. (2016). Conditional Reasoning Test for Creative Personality: Rationale, Theoretical Development, and Validation. *Journal of Management* (online first).

Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. *Journal of Applied Psychology, 78,* 966-974.

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, *14*, 1475-1479.

Sheehy, N. P., & Chapman, A. J. (1987). Industrial accidents. In C. L. Cooper & I. T. Robertson (Eds.), International Review of Industrial and Organizational Psychology 1997. Oxford, UK: Wiley.

Steinberg, L., & Monahan, K. C. (2007). Age differences in resistance to peer influence. *Developmental Psychology*, *43*, 1531-1543.

Stevens, C. K., & Kristof, A. L. (1995). Making the right impression: A field study of applicant impression management during job interviews. *Journal of Applied Psychology, 80,* 587–606.

Stewart, I. I., & John, E. (2006). *Locus of Control, Attribution Theory, and the "Five Deadly Sins" of Aviation* (No. ARI-TR-1182). Army Research Institute for the Behavioral and Social Sciences Fort Rucker Al Rotary-Wing Aviation Research Unit.

Stone, E. R., Choi, Y., de Bruin, W. B., & Mandel, D. R. (2013). I can take the risk, but you should be safe: Self-other differences in situations involving physical safety. *Judgment and Decision Making*, *8*, 250-267.

Sutherland, V. J., & Cooper, C. L. (1991). Personality, stress and accident involvement in the offshore oil and gas industry. *Personality and Individual Differences*, *12*, 195-204.

Thornton, G. C, & Gierasch, P. F. (1980). Fakability of an empirically derived selection instrument. *Journal of Personality Assessment, 44,* 48-51.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124-1131.

van der Pligt, J. (1996). Risk perception and self-protective behavior. *European Psychologist*, *1*, 34.

Veazie, M. A., Landen, D. D., Bender, T. R., & Amandus, H. E. (1994). Epidemiologic research on the etiology of injuries at work. *Annual review of public health*, *15*, 203-221.

Viswesvaran, C., & Ones, D.S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197–210.

Walker, C. B. (1985, February). *The fakability of the Army's Military Applicant Profile (MAP).* Paper presented at the Annual Conference of the Human Resources Management and Organizational Behavior Association, Denver.

Weber, E. U., Blais, A., & Betz, N. (2002). A domain specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*, 263–290.

Wegner, D. M., & Vallacher, R. R. (1977). *Implicit psychology*. New York: Oxford.

Wegner, D. M., & Vallacher, R. R. (1981). Common-sense psychology. In J. P. Forgas (Ed.), *Social cognition: Perspectives on everyday understanding* (pp. 225-246). London: Academic Press.

Weiss, B., & Feldman, R. S. (2006). Looking good and lying to do it: Deception as an impression management strategy in job interviews. *Journal of Applied Social Psychology, 36,* 1070–1086.

Westen, D., & Gabbard, G. O. (1999). Psychoanalytic approaches to personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 57 101). New York: Guilford.

Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C.,& Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review*, *105*, 230-250.

Workcover New South Wales. (2002). *Attitudes and work practices of electricians in the NSW construction industry.* Sydney, New South Wales, Australia.

Zickar, M.J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551–563.

Zimbardo, P. G., Keough, K. A., & Boyd, J. N. (1997). Present time perspective as a predictor of risky driving. *Personality and Individual Differences, 23*, 1007–1023.

Appendix A

Domain-Specific Risk-Taking (Adult) Scale - RT scale

Directions: For each of the following statements, please indicate the likelihood that you would engage in the described activity or behavior if you were to find yourself in that situation. Provide a rating from *Extremely Unlikely* to *Extremely Likely*, using the following scale:

1	2	3	4	5	6	7
Extremely Unlikely	Moderately Unlikely	Somewhat Unlikely	Not Sure	Somewhat Likely	Moderately Likely	Extremely Likely

1. Admitting that your tastes are different from those of a friend. (S)
2. Going camping in the wilderness. (R)
3. Betting a day's income at the horse races. (F)
4. Investing 10% of your annual income in a moderate growth mutual fund. (F)
5. Drinking heavily at a social function. (H/S)
6. Taking some questionable deductions on your income tax return. (E)
7. Disagreeing with an authority figure on a major issue. (S)
8. Betting a day's income at a high-stake poker game. (F)
9. Having an affair with a married man/woman. (E)
10. Passing off somebody else's work as your own. (E)
11. Going down a ski run that is beyond your ability. (R)
12. Investing 5% of your annual income in a very speculative stock. (F)
13. Going whitewater rafting at high water in the spring. (R)
14. Betting a day's income on the outcome of a sporting event (F)
15. Engaging in unprotected sex. (H/S)
16. Revealing a friend's secret to someone else. (E)
17. Driving a car without wearing a seat belt. (H/S)
18. Investing 10% of your annual income in a new business venture. (F)
19. Taking a skydiving class. (R)
20. Riding a motorcycle without a helmet. (H/S)
21. Choosing a career that you truly enjoy over a more secure one.[12] (S)
22. Speaking your mind about an unpopular issue in a meeting at work. (S)
23. Sunbathing without sunscreen. (H/S)
24. Bungee jumping off a tall bridge. (R)
25. Piloting a small plane. (R)
26. Walking home alone at night in an unsafe area of town. (H/S)
27. Moving to a city far away from your extended family. (S)
28. Starting a new career in your mid-thirties. (S)
29. Leaving your young children alone at home while running an errand. (E)
30. Not returning a wallet you found that contains $200. (E)

Note. E = Ethical, F = Financial, H/S = Health/Safety, R = Recreational, and S = Social.

Appendix B

Risk Propensity

Directions: Please indicate to what extent you agree or disagree with the following statements. Please do not think too long before answering, usually your first inclination is also the best one.

1	2	3	4	5	6	7
Extremely Unlikely	Moderately Unlikely	Somewhat Unlikely	Not Sure	Somewhat Likely	Moderately Likely	Extremely Likely

1. Safety first.
2. I do not take risks with my health.
3. I prefer to avoid risks.
4. I take risks regularly.
5. I really dislike not knowing what is going to happen.
6. I usually view risks as a challenge.
7. I view myself as a risk avoider.

Appendix C

Risk-Taking Report

Directions: You will be asked whether you have experienced the following instances. Please select "Yes" or "No" to indicate whether you have experienced each item below in the last 12 months AND/OR in your lifetime.

NOTE: If you have selected "Yes" for in the "In past 12 months", you should also select "Yes" for "In your lifetime".

1. Have you been ticketed for speeding?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

2. Have you owned a motorbike?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

3. Have you been involved in a physical fight?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

4. Have you passed out from drinking or drugs?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

5. Have you had unprotected sex with someone you didn’t know very well?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

6. Have you attended an emergency unit with an injury?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

Appendix D
Substance Use and/or Abuse

Directions: You will be asked whether you have experienced the following instances. Please select "Yes" or "No" to indicate whether you have experienced each item below in the last 12 months AND/OR in your lifetime.

NOTE: If you have selected "Yes" for in the "In past 12 months", you should also select "Yes" for "In your lifetime".

1. Have you passed out from drinking or drugs?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

2. Have you smoked cigarettes or cigars?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

3. Have you used marijuana?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

4. Have you tried “party drugs” (e.g., ecstasy) or hard drugs (e.g., cocaine or heroin)?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

5. Have you taken a higher dosage of medicine than recommended by your doctor or the package insert?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

6. Have you drunk more than you could handle?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No

7. Have you driven a car or motorbike after drinking alcohol?
   In the past 12 months: Yes or No
   In your lifetime: Yes or No