AN EVALUATION OF TEST-TAKING PERFORMANCE IN A SELECTION CONTEXT THROUGH MOTIVATIONAL MECHANISMS AND JOB RELATEDNESS PERCEPTIONS

Shelby Wise

A Thesis
Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

December 2017

Committee:
Clare Barratt, Advisor
Eric Dubow
Scott Highhouse
ABSTRACT

Clare Barratt, Advisor

Applicant reactions researchers have posited that perceptions of test fairness may have a direct affect on applicant motivation and test performance thus diminishing test validity as was proposed in the Justice Theory Framework Perspective. The Justice Theory Framework Perspective describes how situational characteristics inform perceptions of ten procedural justice rules which, in turn, contribute to fairness perceptions and more distally reactions during hiring, reactions after hiring, and applicant self-perceptions. Despite the commentary pertaining to this potential model, at this time, few researchers have empirically tested the aforementioned proposition regarding applicant reactions and test validity. To address this gap, this study investigated the relationship between the perceived job-relatedness of selection assessments and test performance through changes in test-taking self-efficacy and test motivation. Additionally, this model was tested while controlling for individual differences related to achievement (i.e. Conscientiousness and generalized self-efficacy). Results of this study indicate that perceptions of job-relatedness vary by selection test used (i.e. personality test, cognitive ability test, and situational judgment test) and that job-relatedness is significantly related to perceptions of overall process fairness. The model evaluated was also supported wherein job-relatedness predicted test performance through changes in self-efficacy and more distally test motivation. The findings suggest that the applicant reactions-performance relationship may be important in terms of testing proficiency and selection outcomes. Implications of these findings and recommendations for future research are further described.
This work is dedicated to my friends and family for their extensive support throughout this process.
ACKNOWLEDGMENTS

I would first and foremost like to thank my thesis advisor, Clare Barratt, for her help throughout this process. Her support in both the writing and analytics was instrumental in the completion of this project. Additionally, I would like to thank my committee members for their constructive feedback and help in improving my research idea.

I would also like to acknowledge my friends and classmates for their support. Graduate school would not be the same without you.
TABLE OF CONTENTS

INTRODUCTION .................................................................................................................. 1

Applicant Reactions and Organizational Justice ................................................................. 3

Procedural Justice Frameworks ........................................................................................ 4

Job-Relatedness ................................................................................................................ 6

Facets of Job-Relatedness ............................................................................................... 8

Types of Selection Tests .................................................................................................. 11

Job-Relatedness and Self-Perceptions .............................................................................. 16

Motivation ........................................................................................................................ 17

Test Performance ............................................................................................................. 19

Self-Efficacy ..................................................................................................................... 19

Control Variables and Individual Differences ............................................................... 22

METHOD ......................................................................................................................... 26

Participants ..................................................................................................................... 26

Procedure ....................................................................................................................... 26

Measures .......................................................................................................................... 27

RESULTS ......................................................................................................................... 31

DISCUSSION ................................................................................................................... 42

Implications ..................................................................................................................... 44

Limitations ....................................................................................................................... 46

Conclusion ....................................................................................................................... 46

REFERENCES ................................................................................................................. 48
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Procedural Justice Constructs by Category with Definition</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Means and Standard Deviations for Study Variables by Test Condition</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>Post-Hoc ANOVA Comparison Between Test Conditions</td>
<td>33</td>
</tr>
<tr>
<td>4</td>
<td>Means, Standard Deviations, Correlations, and Reliability Estimates</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>for Aggregated Study Variables</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Summary of Exploratory Factor Analysis Results Based on a</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Principle Components Analysis with Oblimin Rotation</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Standardized path estimates for control variables as regressed onto</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>model variables</td>
<td></td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypothesized Model Visual</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Hypothesized Model with Standardized Path Estimates</td>
<td>39</td>
</tr>
<tr>
<td>3</td>
<td>Full Mediation Model with Standardized Path Estimates</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>Complex Model with Standardized Path Estimates</td>
<td>41</td>
</tr>
</tbody>
</table>
INTRODUCTION

More than one third of Americans have negative reactions to pre-employment testing methods (Schmit & Ryan, 1997) despite research suggesting that testing aids can sufficiently predict job performance (Barrick & Mount, 1991; Schmidt & Hunter, 2000). These numbers are staggering and suggest a large disconnect in the way in which selection tests are utilized by organizations and the predilection of job applicants.

The study of applicant reactions, defined as the thoughts, emotions, and feelings in relation to, and resulting from, the selection process, has received a lot of empirical attention (Bauer, Truxillo, Sanchez, Craig, Ferrara, & Campion, 2001), and this has not ceased over the last fifteen years. In a recent meta-analysis by McCarthy, Bauer, Truxillo, Anderson, Costa, and Ahmed (2017), it was noted that, since the year 2000, at least 145 primary studies have been conducted on this topic. Many of these empirical works have evaluated the perceived likability of various testing methods, related the impressions of testing to important organizational consequences, and assessed intervention methods for increasing overall fairness perceptions (e.g., Bakhshi, Kummar, & Rani, 2009; Cohen-Charash & Spector, 2001; Hausknecht, Day, & Thomas, 2004; Konradt, Garbers, Boge, Erdogan, & Bauer, 2017). However, research also supports the notion that applicant reactions may impact test performance (e.g., Chan et al., 1997; Smither et al., 1993) potentially reducing the validity of selection measures (e.g., Arvey, Strickland, Drauden, & Martin, 1990; McCarthy et al., 2009).

Specifically, applicant reactions researchers have posited that perceptions of test fairness may have a direct affect on applicant motivation and, in turn, test performance thus diminishing test validity (e.g., Arvey et al., 1990; Cascio, 1987; Ryan & Ployhart, 2000; Robertson & Kandola, 1982). In support of this position, research confirms that characteristics specific to a
situation are instrumental in the development of motivation and task specific self-efficacy (e.g., Bandura & Locke, 2003; Mortacchio, 1994; Strecher, DeVellis, Becker, & Rosenstock, 1986). Both of which are constructs found to be critical in how one performs across a variety of situations, (e.g., Chemers & Garcia, 2001; Locke, Frederick, Lee, & Bobko, 1984; Stajkovic & Luthens, 1988); including testing (Bouffard-Bouchard, 1990). In regard to selection specifically, Bandura (1997) suggests that negative reactions to tests and testing methods may be detrimental to an applicant’s testing self-efficacy thus impacting test-taking motivation and impeding test performance.

Although many studies have evaluated the relationship between reactions, test-taking self-efficacy (Bauer, Maertz, Dolen, & Campion, 1998), test motivation (Chan et al., 1997), and performance (McCarthey, Van Iddekinge, Lievens, Kung, Sinar, & Campion, 2013), few studies have systemically evaluated the process by which reactions to selection assessments directly impact test performance (i.e., Chan et al., 1997). The current study strives to expand the understanding of applicant reactions by evaluating competing models to determine if and how various selection methods impact perceptions of job-relatedness as well as how reactions influence performance through test motivation and self-efficacy beliefs specific to applicant testing. This study also goes beyond past applicant reactions research by considering the perception/performance relationship when controlling for individual differences related to achievement (i.e., general self-efficacy, Conscientiousness). In sum, this study aims to assess the relationship between test reactions and test performance through changes in test-taking self-efficacy and test motivation while controlling for individual differences related to test achievement (i.e., Conscientiousness, general self-efficacy) (see Figure 1).
**Applicant Reactions and Organizational Justice**

Applicant reactions, defined as the feelings and thoughts directly resulting from the selection procedure, have been examined primarily through organizational justice theories which seek to identify and describe how impressions of fairness and equity impact organizational and individual outcomes. Considerations of organizational justice have been called a vital and rudimentary requirement for the effective operations of an organization (Greenberg, 1990).

A majority of organizational justice research within industrial/organizational psychology has focused on distributive and procedural justice (Gilliland, 1993). Both types of justice perceptions significantly correlate with several critical organizational outcomes such as organizational commitment (Folger & Konovsky, 1989; Martin & Bennett, 1996), intentions to quit (Dailey & Kirk, 1992), and subordinate ratings of supervisor effectiveness (McFarlin & Sweeney, 1992). Originally, organizational justice research centered around the behaviors and attitudes resulting from resource allocation and the outcomes of high-stakes decisions (e.g., salary, hiring decisions, promotions) with no regard for the perceived fairness of the process in which these decisions were made (Gilliland & Beckstein, 1996; Klugger & Rothstein, 1993; Magner, Welker, & Johnson, 1996). However, the study of process fairness (i.e., procedural justice) has become very popular since it’s theoretical inception in the 1970’s (Bobocel & Gosse, 2015) and its integration into the workplace literature in the late 1980s (Lind & Lissak, 1985; Sheppard & Lewicki, 1987).

Procedural justice perceptions have been linked with many pre- and post-hire behaviors that are of utmost importance to the success of organizations such as increased job performance, job satisfaction, organizational commitment, and trust toward the organization, as well as decreased withdrawal behavior (Colquitt, Conlon, Wesson, Porter, & Ng, 2001). In relation to
selection, applicant perceptions of process fairness have been found to better predict job attitudes than other justice dimensions including distributive justice (Bauer et al., 1998). Research supports the notion that, regardless of hiring or promotion decisions, aspects of the testing process are notable contributors to post-hire organizational outcomes.

While improved perceptions of both distributive and procedural justice have been shown to positively affect applicant attitudes and organizational attractiveness, there is little an organization can do to alter hiring and promotion decisions without sacrificing job performance. Organizations can, however, design selection tools and use hiring procedures that will be positively perceived by applicants thus improving organizational reputation (Smither et al., 1993) and fairness perceptions (Hausknect et al., 2004). Because of the evident utility to organizations and the empirical support for this justice construct, the present study aims to focus on the applicant reactions and perceptions related to procedural justice.

**Procedural Justice Frameworks**

The concept of procedural justice emerged in the mid to late 1970’s (Bobocel & Gosse, 2015) with two formative theories developed by Thibaut and Walker (1975) and Leventhal (1980), which both focused on voice effects. The impact of voice effects, which says that fairness perceptions are enhanced when individuals feel as if they have had an opportunity to express their opinion about the process before the decision has been made, has been corroborated in numerous research studies across a variety of disciplines (Lind, Kanfer, & Earley, 1990).

As research on procedural justice increased and evolved, so did the development of models and theoretical frameworks related to these justice perceptions. However, over the past few decades, the Justice Theory Framework Perspective, as coined by Stephen Gilliland (1993), has largely dominated applicant reactions research in regard to organizational justice in a
selection context (e.g., Chan, Schmitt, Jennings, Clause, & Delbridge, 1988; McCarthy et al., 2017). Gilliland (1993) hypothesized that situational characteristics (e.g., test type, human resource personnel) inform perceptions for ten procedural justice rules. The degree to which these rules are met contributes to fairness perceptions, and more distally, reactions during hiring (i.e., motivation), reactions after hiring (e.g., job performance), and applicant self-perceptions (i.e., self-efficacy). The ten justice rules have been lumped within 3 major categories: formal characteristics, explanation, and interpersonal treatment.

*Formal characteristics* are those fairness perceptions directly related to process characteristics which, in selection, are the formal qualities of the tests or hiring procedure. This includes perceptions of test content and consistency of testing methods across applicants. *Explanation* describes the quality and quantity of information provided to the applicant throughout the process. Within selection this entails providing evidence supporting test validity, being open and clear about hiring procedures, and providing quality feedback in a timely manner. The last major justice rule category, *interpersonal treatment*, describes the way in which organizations interact with applicants. This is the degree to which applicants feel well treated interpersonally and are allowed freedom to communicate concerns throughout the process.

Subsequent research has continued to show support for Gilliland’s (1993) model as well the justice rules relationship to fairness perceptions (Bauer et al., 1998; Chapman et al., 2005; Gilliland, 1994; McCarthy et al., 2017; Ployhart & Ryan, 1998; Smither et al., 1993; Truxillo, Bauer, & Sanchez, 2001). A comprehensive list of the justice rules and their definitions created by Gilliland (1993) are located in Table 1 below.
Job Relatedness

Of the justice rules created by Gilliland (1993), job relatedness stands to be one of the most impactful characteristics when fostering positive test reactions (Hausknecht et al., 2004; Rynes & Connerley, 1993; Schleicher, Venkataramani, Morgeson, & Campion, 2006; Steiner & Gilliland, 1996). As mentioned in Table 1, job relatedness is described as the extent to which the tests and items appear related to the job of interest (Gilliland, 1993). Gilliland believed that job relatedness had the greatest influence on procedural justice perceptions and many other researchers have empirically supported this (Bauer et al., 1998; Huaskenecht, et al., 2004; Rynes & Connerley, 1993; Truxillo, Bauer, & Sanchez, 2001; Truxillo, Bodner, Bertolino, Bauer, & Yonce, 2009; Zibarras & Patterson, 2015).
Table 1. *Procedural Justice constructs by category with definition (Gilliland, 1993)*

<table>
<thead>
<tr>
<th>Category/Item</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Formal Characteristics</strong></td>
<td></td>
</tr>
<tr>
<td>Job Relatedness</td>
<td>The extent to which a test appears to be valid and measure relevant job information</td>
</tr>
<tr>
<td>Opportunity to Perform</td>
<td>Having the opportunity to show your knowledge, skills, and abilities during the testing process</td>
</tr>
<tr>
<td>Reconsideration Opportunity</td>
<td>Opportunity to question the decision and have it modified if an error was made</td>
</tr>
<tr>
<td>Consistency</td>
<td>Ensuring that decisions are consistent across individuals over time</td>
</tr>
<tr>
<td><strong>Explanation</strong></td>
<td></td>
</tr>
<tr>
<td>Feedback</td>
<td>Timely and informative feedback regarding testing</td>
</tr>
<tr>
<td>Selection Information</td>
<td>Information regarding the validity and effectiveness of these tests</td>
</tr>
<tr>
<td>Honesty</td>
<td>Honesty and openness from those doing the recruiting</td>
</tr>
<tr>
<td><strong>Interpersonal Treatment</strong></td>
<td></td>
</tr>
<tr>
<td>Interpersonal Effectiveness</td>
<td>The degree to which applicants are treated with warmth and respect</td>
</tr>
<tr>
<td>Two-Way Communication</td>
<td>Opportunity for applicants to offer input and have their views considered during the process</td>
</tr>
<tr>
<td>Propriety of Questions</td>
<td>Avoiding improper and prejudicial questions during recruitment</td>
</tr>
</tbody>
</table>
The construct of job-relatedness was designed with consideration of seminal procedural justice rules including both the accuracy and resource rules. The accuracy rule, created by Leventhal (1990), states that for the process of decision making to be perceived as fair, decisions need to be based on as much good information as possible. Similarly, the resource rule empirically studied and developed by Sheppard & Lewicki (1987), outlines how decisions should be made using the appropriate expertise and sources deemed as accurate by the individual impacted by the decided outcome. Procedural fairness perceptions in selection are violated when hiring outcomes are based on poor information or when the test content is deemed as irrelevant to the position. These justice rules have been since linked to an extensive array of empirical works supporting the connection between job-relatedness and fairness perceptions (Kluger & Rothstein, 1991; Schmitt, Gilliland, Landis, Devine, 1993; Schmitt, Oswald, Kim, Gillespie, & Ramsay, 2004; Smither et al., 1993; Truxillo et al., 2001; Truxillo, et al., 2009). Meta-analytic work confirms that applicants prefer selection procedures in which the test content is visibly linked to skills and abilities required by the job (Hauskenecht et al., 1994). For the lay individual, it seems that job-related content is perceived as more valid, meaning an applicant can more easily rationalize how the test content relates to future job performance (Gilliland, 1993).

**Facets of Job Relatedness**

Empirically, job relatedness has been measured using both perceptions of face validity and perceived predictive validity (Bauer et al., 2001; Honkaniemi, Feldt, Metsäpelto, & Tolvanen, 2013, Smither et al., 1993). As mentioned previously, job-relatedness measures the degree to which a test appears to be valid and content relevant (Anastasi, 1988). Whereas face validity encompasses the content relevance and is defined as the appearance of a test as practical and measuring what is intended to be measured (Gilliland, 1993), perceived predictive validity
describes the degree to which the test material seems valid in that it is relevant to the prediction of job performance (Smither et al., 1993). Although some research has examined face validity and predictive validity separately, several studies have found them to be discriminate aspects of the overarching construct of job-relatedness (Hausknecht et al., 2004).

**Face validity.** Academically, face validity has been defined in a variety of ways such as the degree to which a test looks to measure what it intends to (Aronson et al., 1990), the suitability and relevance of test content to the purported test use (Nevo, 1985), and the degree to which a test seems job-related (Smither et al., 1993). However, the generic research methods definition says that the test “on its face” measures the construct adequately in that the superficial impressions relate to the construct (Crano, Brewer, & Lac, 2014, pg. 65). Essentially, if a test is intended to be face valid, the meaning and utility should be outwardly apparent to any individual (Nevo, 1985). In this study, face validity will be measured and defined as “the extent to which applicants perceive the content of the selection procedure to be related to the content of the job (Smither et al., 1993, pg. 54).”

While some researchers have regarded face validity as trivial, the importance in relation to selection and applicant reactions cannot be ignored. Nevo (1985) argues that, although many researchers discount or deride the importance of face validity as a test characteristic, face validity is a necessary test quality, especially in situations where stakeholder acceptance is critical such as selection (Mosier, 1947). This argument has been empirically substantiated, especially in applicant reactions research (Chan & Schmitt, 1997; Ployhart, Ziegert, & McFarland, 2003). Face validity has been linked to positive applicant reactions through perceived job-relatedness in a number of highly cited, academic works (Hausknecht et al., 2004; Kluger & Rothstein, 1993; Rynes & Connerley, 1993) and has been found to contribute to the psychometric validity of
testing procedures (Chan & Schmitt, 1997; Smither at al., 1993). It has been posited that face validity positively contributes to an applicant’s motivation to perform thus increasing psychometric test properties as compared to a testing procedure with a less salient intent (Chan & Schmitt, 1997). Overall, research supports the notion that, through the implementation of face valid selection procedures, organizations can increase applicant perceptions of both procedural and distributive justice, attitudes towards tests, and attitudes toward selection more generally (Hausknecht et al., 2004).

**Predictive validity.** In addition to face validity, job-relatedness is also comprised of perceived predictive validity or predictive job relatedness (Bauer et al., 2001). Perceived predictive validity describes the extent to which a test is believed to adequately predict future job performance. Unlike face validity, which describes the test as superficially appearing to relate to test content, perceived predictive validity evaluates the degree to which an applicant believes the test can accurately differentiate between high and low performers. The accuracy rule, as defined by Sheppard and Lewicki (1987), says that hiring decisions, when sources are perceived as expert and accurate, bolster justice perceptions. Sources, or selection strategies, seen by applicants as predictive of work success are much more likely to be accepted as fair and suitable for hiring procedures (Hauskenecht et al., 2004; Smither et al., 1993).

It is important to note that perceived predictive validity is not a psychometric property and differs from a test’s actual ability to predict job performance. Whereas predictive validity is a subfacet of criterion validity and is used to determine whether a scale can accurately predict some type of performance or differentiate degrees of ability, *perceived predictive validity* is assessed through self-report responses regarding the extent to which an applicant believes scores
on a specific test look as though they will relate to performance on the job rather than if the scores actually do relate to future job performance.

**Types of Selection Tests**

As stated, the job-relatedness of a test or selection procedure has become synonymous with positive applicant reactions, however these reactions have been found to vary across selection techniques (Anderson & Witvliet, 2008; Hausknecht et al., 2004; Ryan & Ployhart, 2000). This means that some selection practices are inherently deemed as more or less job-related based on their content and implementation method alone. Research has confirmed that hiring practices using work simulations or directly measured work behaviors are favored over those tests assessing for abstract characteristics, even if those abstract characteristics are predictive of work performance (Dodd, 1977; Robertson & Kandola, 1982; Schmidt, Greenthal, Hunter, Berner, & Seaton, 1977). Even some of the most commonly utilized selection techniques, general mental ability tests, work sample tests, personality tests, and situational judgment tests (SJTs), vary widely in likability in relation to fairness perceptions (e.g., Anderson, Salgado, & Hulsheger, 2010; Hausknecht et al., 2004; Kluger & Rothstein, 1991; Smither et al., 1993). This section aims to outline some of the most utilized testing methods and describe their standing in applicant reactions research.

**Measures of general mental ability.** At this time, the most effective selection tool in the prediction of job success is the measurement of cognitive ability or general mental ability (GMA; Hunter, 1986; Hunter & Schmidt, 1996; Ree & Earles, 1992; Schmidt & Hunter, 1981). These tests are used to evaluate intelligence, defined as one’s ability to understand and reason through abstract concepts as well as problem solve effectively. Individuals high in cognitive ability are able to quickly acquire and apply job-knowledge necessary to their success within an
organization. Obviously, some jobs, those that are more complex, necessitate extensive job-knowledge, but selection assessments measuring cognitive ability successfully predict job performance even for those jobs requiring less cognitive complexity (Schmidt & Hunter, 2000).

Cognitive ability tests come in varying formats (i.e., paper and pencil tests, computer based assessments) with a variety of content (i.e., math problems, vocabulary items). In general, cognitive ability tests are well perceived by applicants in regard to fairness (e.g., Hausknecht et al., 2004; Kluger & Rothstein, 1991; Smither et al., 1993). However, perceived predictive validity is usually higher for measures of GMA than face validity perceptions (Smither et al., 1993) meaning that applicants believe that these tests do predict performance to some extent but the test intent is somewhat concealed. Supporting job-relatedness research, fairness perceptions of cognitive ability tests are higher when math skills, for a job requiring basic math, are evaluated rather than abstract assessments like the Raven Matrices Test (Raven, 1936), which measures fluid intelligence using shapes and complex patterns (Smither et al., 1993).

Personality tests. Another commonly employed selection technique is personality testing, which is utilized by roughly 40% of Fortune 100 companies and 30% of all American companies (Erickson, 2004). Personality tests measure individual differences that describe one’s interactions, behaviors, and characteristics. Over time, personality has been operationalized in many different ways, but the Big Five seems to be the most commonly utilized framework for trait classification (Gatewood, Field, & Barrick, 2010). The Big Five consists of Conscientiousness, Openness to Experience, Emotional Stability, Extraversion, and Agreeableness. In a large meta-analysis, Barrick and Mount (1991) found that Conscientiousness and Emotional Stability were the most predictive personality traits for all jobs. However, other Big-Five personality traits can also be important predictors depending on the specifics of that
job, such as Extraversion for customer service positions. While personality is not as predictive of job performance as general mental ability, personality does add some incremental prediction above GMA. Further, hiring managers overwhelmingly believe that personality is virtually as important as GMA making it a popular assessment method within organizations (Dunn, Mount, Barrick, Ones, 1995).

Obviously, organizational perceptions of personality measurement are positive, but the use of personality tests in the hiring process remains one of the most contentious issues in the area of applicant reactions as these tests, in the eyes of the test taker, seem to lack relation to the job and unnecessarily infringe upon applicant privacy (McFarland, 2003; Smither et al., 1993). When compared to other selection measures, personality is often rated low in fairness, job-relatedness, face validity, and predictive ability of the test to predict job performance (e.g., Anderson et al, 2010; Hausknecth et al., 2004; Rosse, Miller, & Stecher, 1994; Smither et al., 1993). These findings have been generalized across testing contexts and in a variety of countries (Anderson et al., 2010). Personality tests are meant to assess general attitudes, preferences, and behavioral tendencies of an applicant (Rosse et al., 204). This is often operationalized using brief trait statements like “I use others to get ahead” or “I am often worried.” These items are rarely contextualized meaning they do not ask about behavior directly related to the job but behavior more generally. However, research shows that when items are contextualized and made work-relevant, perceptions of face validity and predictive validity are improved (Holtrop, Born, de Vries, & de Vries, 2014).

It has been argued that negative reactions might not come from the assessment of personality exactly, as it is often assessed in interviews (Huffcutt, Conway, Roth, & Stone, 2001) which are generally well-liked by applicants (Hausknecht et al., 2004). Rather, negative
reactions might stem from the way in which these tests assess personality (McFarland, 2003). In interviews, personality is measured with questions contextualized to work (Huffcut et al., 2001), while personality tests use general statements. Essentially, interviews measure personality in a job-related way while personality tests generally do not, likely contributing to the difference in applicant reactions. Personality tests, overall, appear unrelated to the job, are poorly perceived by applicants, and are often deemed as overly intrusive.

**Situational judgment tests.** The selection techniques discussed previously have a long history within the area of selection (Lievens, Peeters, & Schollaert, 2008). More recently, however, applied psychologists and human resource professionals in the United States have more frequently integrated situational judgment tests (SJTs) into the process of employee selection (Ployhart, 2006). SJTs describe a testing method where participants are presented with a variety of situations directly related to the job and asked which actions, from a provided list, are the most appropriate. These are assessments designed to measure one’s ability to make successful judgments within the workplace (Cabrera & Nguyen, 2001). STJs are more similar to a measurement method (Chan & Schmitt 1997; Weekly & Jones 1999) in that they can be specifically designed to emphasize specific predictors like personality or cognitive ability. Generic SJT’s do have a correlational overlap with cognitive ability and personality (Clevenger, Pereira, Wiechmann, Schmitt, & Harvey, 2001), however, SJT’s, through the measurement of interpersonal skills, explain incremental variance in the prediction of job performance, meaning that when used with other effective predictors, like personality and cognitive ability tests, SJTs provide some unique predictive ability (e.g., Chan & Schmitt, 2002; Jackson, LoPilato, Hughes, Guenole, & Shalfroshan, 2017; Lievens & Sackett, 2006; McDaniel, Hartman, Wetzeli, & Grubbs, 2007).
This testing medium is particularly appealing to practitioners because implementation is usually inexpensive and interpersonally related skills can be measured quickly at any point in the selection process (Roth, Bobko, & McFarland, 2005). SJT’s are valid and incrementally helpful in predicting job performance, while also limiting racial subgroup differences as compared to generic cognitive ability tests (Motowidlo, Dunnette, & Carter, 1990). As a con, however, the creation and validation of SJTs can be difficult and more time consuming than other tests, as they are usually designed for a specific job (Lievens, Peeters, & Schollaert, 2008).

As mentioned, SJTs are a more recent development in selection research and practice, having not been formally designed until 1990 (Motowidlo et al.). Because of its late addition to selection testing, this method has been largely neglected in prominent applicant reactions research (Kanning, Grew, Hollenberg, & Hadouch, 2006). A few studies, however, have been conducted to evaluate applicant perceptions of SJTs independently and, overall, applicant reactions to SJTs seem pretty positive (Chan & Schmitt, 1997; Kanning et al., 2006; Richman-Hirsch, Olson-Buchanan, & Drasgow, 2000). These positive applicant reactions have been attributed to the salient relationship of test items to the job of interest (Lievens et al., 2008). Essentially, because SJTs are usually created for each job and employ the use of scenarios specific to work tasks, they inherently appear more job-related than other tests.

As supported by many empirical studies, the job-relatedness of selection procedures equates to applicant perceptions of fairness (e.g., Bauer et al, 1998; Rynes & Connerley, 1993; Steiner & Gilliland, 1996), but not all tests are created equal. Overall, research has shown that work sample tests and interviews are perceived to be the most job-related, followed by cognitive ability tests, and lastly, personality measures (e.g., Hausknecht et al., 2004; Smither et al., 1993). Situational judgment tests have, at this time, not been compared to other measures in regard to
job-relatedness and fairness perceptions. However, based on the test content and research relating SJTs to high fairness perceptions, it would be expected that the job-relatedness for SJTs would also be high, even when compared to other testing methods. One goal of this study is to replicate findings in previous applicant reactions research (e.g., Hausknecht et al., 2004; Smither et al., 1993), showing that job-relatedness perceptions vary by test. However, this study also aims to extend previous applicant reactions research through the incorporation of SJTs.

Hypothesis 1: Job relatedness perceptions will be highest for (a) situational judgment tests, followed by (b) cognitive ability tests and (c) personality tests.

Hypothesis 2: Job-relatedness perceptions will positively relate to overall fairness perceptions.

Job-Relatedness and Self-Perceptions

As mentioned previously, the Justice Theory Framework explains how specifics of the situation (e.g. test type) inform evaluations of several justice rules which drive fairness perceptions, and more distally, various outcomes including self-perceptions (e.g. motivation, test-taking self-efficacy, performance). This theoretical model coincides with other models derived across disciplines such as the model of achievement behavior which states that situational factors significantly contribute to one’s self-perception and intent to effectively orient effort (i.e., motivation; Schunk, 1995). These theories suggest that, within a selection context, the situational factors inherent to the testing situation (i.e. tests used) and the resulting fairness perceptions contribute to an applicant’s evaluation of personal competency, motivation to perform, and degree to which they perseverere. (Bandura, 1997). In support of these theoretical models, research has shown that job-relatedness and its sub facets impact an applicant’s self-
perceptions toward testing through changes in self-efficacy (Bauer et al., 1998), self-esteem (Chan, & Schmit, 1994), and test motivation (Chan et al., 1997).

**Motivation**

Motivation, most generally, is the psychological and situational mechanism that drives someone to act (Ryan & Deci, 2000) in both intensity and direction of behavior (Humphreys & Revelle, 1984). Motivation largely contributes to the degree to which someone orients cognitive resources, feels efficacious towards a given task, and can persevere when presented with difficulty (Bandura & Locke, 2003), which is likely why motivation has been suggested as a primary driver of performance across a variety of situations (Chemers, Hu, & Garcia, 2001; Locke, Frederick, Lee, & Bobko, 1984; Stajkovic & Luthens, 1988), including applicant assessment (Chan et al., 1997). As was posited by applicant reaction researchers (Arvey et al., 1990), positive test reactions, specifically from job-related test materials, have been found to incite test-taking motivation (Hausknecht et al., 2004) thus improving performance on selection tests (e.g., Chan et al., 1997; McCarthy, Van Iddeking, Lievens, Kung, Sinar, & Campion, 2013).

A variety of theories have been used to explain the relationship between test perceptions and motivation including process efficiency theory (Eysenck, Derakshan, Santos, & Calvo, 2007), interference theory (Sarason, Sarason, & Pierce, 1990), expectancy theory (Sanchez, Truxillo, & Bauer, 2000), applicant attribution-reaction theory (AART; Ployhart & Harold, 2004), and social cognitive theory (Stajkovic & Luthans, 1979). Both process efficiency theory and interference theory argue that perceptions of fairness orient an applicant’s allocation of cognitive and attentional resources (Arvey et al., 1990, Bandura, 1982). Therefore, when few extraneous factors cloud cognitive and attentional resources (i.e., test anxiety, perceived injustice), applicants can more appropriately focus in on the task at hand, orient their attention,
and reduce peripheral processing thus increasing motivation and self-efficacy (Arvey et al., 1990; Bandura, 1982; Chan, Schmitt, Sacco, & DeShon, 1998)

Another potential explanation is through Expectancy theory which describes how one’s expectations about the future impact motivation (Derous, Born, & Witte, 2004). In a selection context, expectations regarding procedural and distributive justice have the potential to play into the formation of test reactions (Greenen, Proost, Schreurs, van Dam, & von Grumbkow, 2012), test motivation, and self-efficacy (Bell, Wiechmann, & Ryan, 2006; Chapman, Uggerslev, & Webster, 2003; Derous et al., 2004; Schreurs, Derous, Proost, & De Witte, 2010). According to expectancy theory, if individuals expect the test to be related to the job and, upon review, believe otherwise, they are likely to have reduced perceptions of ability and motivation (Ryan et al., 2000). Within the conceptual model by Bell, Ryan, and Weichmann (2004), past test experiences, beliefs in tests, and reactions to taking tests generally may also drive expectations thus impacting motivation. Similarly, applicant attribution-reaction theory (AART; Ployhart & Harold, 2004) posits that situations specific to selection are generally neutral. It is the attributional processing that informs the creation of perceptions and behavior outcomes including test performance and motivation. In other words, in an attempt to understand novel situations, people define and categorize situation specific information (i.e. interpersonal treatment, tests), which is used to inform self-perceptions related to ability and motivation.

No one theory fully explains motivation, as behavior is a function of many variables (Schunk, 1991). However, it is likely that because of research supporting these theories in relation to applicant reactions (e.g. Bauer et al., 1998; McCarthy et al., 2013; Ployhart, Ehrhart, & Hayes, 2005) as well as the theoretical basis from which they were drawn, all contribute to the reaction-motivation to some extent. The attributions assigned to a situation (AART) are
likely derived from the associated expectations (expectancy theory) which inform applicant focus, attention orientation, and peripheral processing (interference theory). Based on theory as well as the consistent finding that test-taking motivation relates to job-relatedness (Hausknecht et al., 2004) and perceived fairness (Bauer et al., 2001; Hausknecht et al., 2004); it is hypothesized that job-relatedness will predict test-taking motivation.

*Hypothesis 3:* Job-relatedness perceptions will be positively related to test motivation.

**Test Performance**

As mentioned, motivation drives someone to orient cognitive effort and persevere in times of difficulty (Bandura & Locke, 2003) thus generating performance (Chan et al., 1997; Chemers et al., 2001; Locke et al., 1984; Stajkovic & Luthens, 1988). Because motivation is so strongly linked to behavior, it has been stated that when reactions deplete test motivation and decrease test performance, those who would be the best applicants are erroneously selected out of the process thus harming the proficiency of the selection system and the organization (e.g., Arvey et al., 1990; Ryan & Ployhart, 2000). Employee performance is of utmost importance to the success of an organization and the more we know about the mechanisms that may alter performance, the better we can make our hiring processes to enhance organizational achievement. Performance is widely considered an outcome of both ability and motivation (Chan et al., 1997); therefore, it is hypothesized that test motivation will play an important role in the prediction of test performance.

*Hypothesis 4:* Test motivation will be positively related to test performance.

**Self-Efficacy**

As mentioned, the reactions-motivation relationship has been explained through a variety of theories (e.g., process efficiency theory, expectancy theory). However, one of the most
commonly utilized and empirically tested motivational theories is social cognitive theory (SCT), which suggests that motivation is derived from both environmental and cognitive factors (Stajkovic & Luthans, 1979). In essence, people cognitively integrate environmental cues to make a determination about their ability to succeed thus driving their motivation to perform. This defines a major feature of SCT, self-efficacy, which describes the degree to which a person feels capable to affect the environment and produce desired outcomes (Bandura, 1977). According to SCT, test content provides the applicant with environmental cues that must be integrated, using individual cognition, to determine one’s test-taking self-efficacy or the degree to which they feel confident in their ability to be successful on a specific measure (Bandura, 1997).

Self-efficacy is a motivational construct described as a person’s belief that they are able to complete a task or group of tasks to attain specific performance goals (Bandura, 1977). Self-efficacy contributes to how one feels about a task, how they choose to motivate themselves, the goals they set specific to that task, and whether or not they can persevere when presented with difficulties (Bandura & Locke, 2003). Stronger perceptions of self-efficacy are also related to higher personal goal challenges and goal commitment (Bandura, 1991). Those with high, task-oriented self-efficacy visualize success on the specific task while those who do not have increased self-doubt, thus dwelling on the possibility of failure (Bandura, 1993).

Central to SCT is the idea that self-efficacy is derived from both the situation and the way that situation is appraised (Schunk, 1995). Within selection, the testing situation, tests utilized, and the general appraisal of fairness informs one’s self-evaluation of competency thus impacting test-taking motivation and test performance (Bandura, 1997). The relationship between test reactions and self-efficacy can potentially be explained through the accuracy rule which states that for decisions to be deemed as fair, they should be made using as much credible information
as possible (Leventhal, 1990). In the testing situation, however, where the test seems unrelated to the job, it is difficult for applicants to make a judgment regarding a test’s accuracy in predicting performance. This distrust in the testing process reduces one’s feelings of competency and ability to succeed as the decision criteria is unclear.

It has been posited that fairness perceptions aide in one’s development of self-efficacy through attribution theory (Truxillo, Bauer, & Sanchez, 2001; Weiner, 1985). When the test is deemed as unfair and not related to the job of interest, applicants are more likely to attribute their performance to external outcomes in order to reduce threats to the self. Essentially, tests that are not job-related, thus evaluated as unfair, reduce perceptions of control therefore diminishing one’s internal efficacy.

In alignment with Gilliland’s (1993) Justice Theory Framework’s perspective, the relationship between testing perception and self-efficacy has been tested on numerous occasions and found to be substantial (Hausknecht et al., 2004; McCarthy et al., 2013; Truxillo et al., 2002). In summary, situational factors can contribute to the degree to which an individual is motivated and feels efficacious about their test-taking ability (Schunk, 2005). Situational factors in selection, characteristics of the tests and the resulting fairness perceptions, lead an applicant to feel competent, therefore motivated to put in the necessary effort to be successful and persevere through difficult items (Bandura, 1997). As noted, motivation is a function of many variables (Schunk, 1991); thus, it is hypothesized that job-relatedness will positively relate to motivation partially through self-efficacy beliefs derived from the job-relatedness perceptions of the selection procedures. A model containing all hypothesized relationships can be found in Figure 1.
Hypothesis 5: Self-efficacy will partially mediate the relationship between job-relatedness perceptions and test motivation.

**Figure 1. Hypothesized Structural Equation Model of Job-Relatedness Perceptions and Performance Through Motivation Constructs**

**Control Variables: Individual Differences**

Lastly, it is important not only to consider how job-relatedness perceptions of varied selection techniques influence test performance through motivational mechanisms but it is also imperative that we evaluate the extent to which this process is impacted by individual differences. Recently, a meta-analysis was conducted wherein the authors reviewed the current state of applicant reactions research and provided commentary outlining important next steps for research in the applicant reactions domain (McCarthy et al., 2017). One relatively novel and valuable research area mentioned was the topic on individual differences.

In a previous study, Chan and colleagues (1997) empirically examined the directional relationship between face validity, test-taking motivation, and test performance. In support of the proposed hypotheses, face validity did significantly predict performance through changes in test-taking motivation. In the discussion, researchers noted the findings could be evidence for a potential reduction in test validity based on test reactions but it was also plausible that these
differences reflected trait motivation which could, in turn, predict motivation on the job therefore, bolstering rather than hindering test validity. This study begins to address this concern by evaluating a similar model while also controlling for individual differences related to test and job achievement: Conscientiousness (Barrick & Mount, 1991) and generalized self-efficacy (GSE; Judge & Bono, 2001).

**Conscientiousness.** Conscientiousness is a Big Five Personality trait that encompasses how a person controls and regulates daily impulses, their degree of responsibility, and general predilection toward achievement-striving. Conscientiousness is made up of several facets that include orderliness, self-discipline, and cautiousness. Many meta-analytic studies have shown strong evidence for the relationship between Conscientiousness and job performance across a variety of settings (Barrick & Mount, 1991; Chen et al., 2001) and has been declared the primary dispositional predictor of job performance by Mount and Barrick (1995).

It has been recommended that organizations hire on Conscientiousness because of its significant predictive ability (Barrick & Mount, 1991) and its relationship to productivity measures, performance ratings, job tenure, and goal setting motivation (Chen et al., 2001; Judge & Ilies, 2002). Highly conscientious individuals are more inclined to set positive goals, remain resilient in difficult times, and strive to achieve (Barrick & Mount, 1991; Brown, Lent, Telander, & Tramayne, 2011; Chen et al., 2001; Judge & Ilies, 2002) which is what makes them particularly successful in a work setting. These skills also make a person more inclined to succeed within a testing setting. In a meta-analysis by Hausknecht and colleagues (2004), Conscientiousness was found to positively correlate with procedural justice perceptions only .08 but Conscientiousness and test motivation significantly correlate at .20 meaning that
Conscientiousness might not alter test perceptions, but rather the effect test perceptions have on motivation and how one performs.

**Generalized Self-Efficacy.** Self-efficacy was originally defined as the “beliefs in one’s capabilities to mobilize the motivation, cognitive resources, and courses of action needed to meet given situational demands” (Wood & Bandura, 1989, p. 408). Because the definition included “situational demands,” research focused on task-specific self-efficacy, which is related to many positive organizational outcomes (Judge & Bono, 2001; Stajkovic & Luthans, 1998). However, in the late 1980’s/early 1990’s, researchers began to study the trait level of self-efficacy termed generalized self-efficacy (GSE; Eden, 1988, 1996; Gardner & Pierce, 1998; Judge, Erez, & Bono, 1998; Judge, Locke, & Durham, 1997). GSE has been defined as one’s perceptions of competence and perceived ability across various situations (Judge et al., 1998). The assessment of GSE allows researchers to determine differences in the way in which people view themselves and the degree to which they believe they can succeed in a variety of contexts (Chen, Gully, & Eden, 2001).

Empirically, GSE and task-related self-efficacy have been shown to be separate constructs (Stajkovic & Luthans, 1998); however, they share many outcomes and antecedents (Eden, 1988; Gardner & Pierce, 1998; Judge et al., 1997). In contrast, generalized self-efficacy refers to a broader scope than task specific self-efficacy and is much more resistant to transient influences (Eden, 1988). This suggests that those high in GSE will be less impacted by testing perceptions and more inherently efficacious and motivated to succeed. Given the scope of this project, to evaluate the relationship between job-relatedness perceptions and job performance through motivational mechanisms, it is also important to consider how achievement related individual differences modify the utility of the proposed model. If the proposed model has poor
fit when considering these individual differences, it might indicate that testing perceptions are less important in regard to selection utility, as those who are likely to perform well and be hired are resilient to testing reactions.

It has been stated that when reactions deplete test motivation thus impacting test performance, the proficiency of the selection system may be impacted (Arvey et al., 1990; Ryan & Ployhart, 2000); however, it has also been noted that these dispositional or trait differences in motivation may reflect real candidate ability (McCarthy et al., 2013). Employee performance is of utmost importance to the success of an organization and the more we know about the mechanisms that may alter test proficiency as well as the boundary conditions under which proficiency remains, the better we can make our hiring processes to enhance organizational achievement.
METHOD

Participants

For the current study, participants were recruited through Amazon’s Mechanical Turk (MTurk). MTurk is a crowd sourcing internet marketplace that reaches a diverse group of individuals interested in participating in research projects as well as completing other online jobs. Past research suggests that MTurk is an appropriate and reliable way to obtain participants as MTurk workers act similarly to those participants from other commonly used sampling sources (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010).

In an effort to ensure data quality, only those participants who had previously completed 50 HITS, had an approval rate of 95%, and lived in the United States were eligible. Participants were also required to work in customer service at least 24 hours a week. The original sample consisted of 466 participants, but 41 respondents were removed for failing data quality metrics (e.g., more than one quality check missed, completing the survey in less than 4 minutes, test performance 2.5 SD below mean or more). The average age of the 425 participants retained was 34.6 ($SD = 9.7$). Seventy-six percent were white, 10% were African American, 6% were Hispanic, 7% were Asian, and 51% were male.

Procedure

First, eligible participants were asked to complete the items for the control variables (i.e., generalized self-efficacy, Conscientiousness). Similar to the procedure used by Holtz and colleagues (2005), participants were then provided a job description outlining a desirable (e.g., competitive salary, benefits) customer service position for which they were applying. This job description can be found in Appendix A. This type of research design has been shown to have good generalizability to the field when real job incumbents are not available (Hunthausen
Participants were also told that the top 10% of performers on the assessment would receive additional compensation (i.e., 50% of the promised payment) to increase motivation to perform as seen in similar work with non-applicants (e.g., Chan et al., 1998; Holtz et al., 1995).

Next, participants were randomly assigned to 1 of 3 conditions (personality test, cognitive ability test, or situational judgment test) and asked to review and complete 5 items from that test. This method allows us to capture perceptions devoid of reactions stemming from the actual test performance. After completion of the small test sample, participants were asked to respond to the items for face validity, perceived predictive validity, test-taking self-efficacy, and test motivation. All participants completed these items in that order. Lastly, participants were asked to complete the full selection assessment then respond to perceptions of overall fairness.

The assessments had been previously piloted in undergraduate classes to ensure time for completion was not a confound. Each test was expected to take roughly 8 minutes. The cognitive ability test was timed as it reflected the qualities of a speed test as described by Mead and Drasgow (1993). Cognitive ability tests for customer service jobs usually contain relatively easy items that, given an unlimited amount of time, would be correctly answered by most individuals. The time limit, which is usually employed in selection situations for assessments of general mental ability, is necessary in this given situation to differentiate high and low performers on commonly used test items.

**Measures**

**Selection assessments.** All selection tests used in this study are intellectual property of two consulting companies who specialize in the creation and implementation of selection materials. Because of the proprietary nature of these tests, test material and scoring keys can not be published.
Situational judgment test. The SJT contains 17 items meant to evaluate a person’s ability to judge appropriate behavior in specific situations. This is a validated measure used for selection in customer service positions by Work Skills First, Inc. After being given a situation, participants were provided with 2-6 possible behaviors. They were then asked to endorse how effective each behavior is on a 6-point Likert scale. Points are awarded for correctly identifying the behaviors as effective or ineffective. Higher scores indicate better performance. The directions and an example of a test item can be found in Appendix B.

Cognitive ability test. Forty items were aggregated from a larger multi-hurdle selection system used by Polaris Assessments Incorporated specific to customer service positions within a large organization. Participants were given an 8-minute time limit to complete as many test items as they can. The forty items were meant to assess competencies related to basic math, reasoning skills, and attending to details. Example test items can be found in Appendix C. The test and scoring system has been validated by Polaris Assessments Inc. For test scoring, points are awarded for correct answers and higher scores indicate better performance.

Personality test. Sixty-five personality items were taken from a larger selection system used to hire for customer service jobs within a large, well-known organization. These items reflect competencies related to flexibility, communication, creativity, sociability, helping others, reliability, initiative, and customer service demeanor. Personality items were randomly presented with responses ranging from (1) strongly disagree to (5) strongly agree. The scoring system has been validated by Polaris Assessments Inc. Five points are assigned to the best answer and 1 point to the worst. Higher scores indicate better performance. Example test items can be found in Appendix D.
**Face validity.** Face validity was measured using a five-item survey created by Smither and colleagues (1993) in one of the seminal works on applicant attraction. These items are part of a larger questionnaire also evaluating perceived predictive validity, likelihood of improvement, affect, perceived knowledge of results, and organizational attractiveness. Example items include, “I did not understand what the examination had to do with the job” and “it would be obvious to anyone that the examination is related to the job.” Participants are asked to respond on a five-point Likert scale ranging from strongly disagree to strongly agree. The alpha reported by Smither and colleagues (1993) was .86. A full list of items for the academic measure used can be found in Appendix E.

**Perceived predictive validity.** The items being used to measure perceived predictive validity were also from the larger survey created by Smither et al. (1993). This measure contains five items where participants were asked to respond on a 5-point Likert scale ranging from strongly disagree to strongly agree. This measure was found to have an alpha coefficient of .83 in the original study by Smither and colleagues (1993). Sample items include, “failing to pass the examination clearly indicates that you can’t do the job” and “my performance on the examination was a good indicator of my ability to do the job.”

**Testing self-efficacy.** The measure for self-efficacy was created based on recommendations by Bandura (2006). These questions were created to assess self-efficacy for this specific task rather than general self-efficacy. Participants were asked to, “rate your degree of confidence in each statement by responding appropriately to the scale below,” for each of the items. Responses were reported on a five-point Likert-scale where 1 is strongly disagree and 5 is strongly agree. Sample items include, “I believe I can do well on this test” and “I have the ability to be successful on this test.” This method is similar to that used by Gilliland (1994).
**Test motivation.** Test motivation was measured using five items from a ten item scale originally created by Arvey, Strickland, Drauden, and Martin (1990). In the original survey, the 10 items had a coefficient alpha of .85. Participants were asked to respond to these items as if they were taking the selection test as an applicant for the specific job outlined on a 5-point Likert scale ranging from strongly disagree to strongly agree. Example items include, “I want to be among the top scorers on this test” and “doing well on this test is important to me.” The scale was abbreviated in an effort to eliminate redundancy and decrease time strains on the participants. Two items from the larger 10 item version were removed because they were negatively worded. The other three were dropped because of substantial overlap in item content. An example being that the item “I tried my best on this test” was excluded while "I tried to do the very best I could to on this test” was utilized in this abbreviated version.

**Generalized self-efficacy.** GSE was measured using the New General Self-Efficacy Scale (NGSE) developed by Chen, Gully, and Eden (2001). The measure consists of 8 items. Participants were asked to rate the extent to which each statement describes them on a 1-5 Likert-scale where 1 is strongly disagree and 5 is strongly agree. An example item is “I will be able to achieve most of the goals that I have set for myself.”

**Conscientiousness.** Conscientiousness was measured using the 10 item version of the NEO IPIP adapted by Goldberg (1992). Participants were asked to report the degree to which on a 5 item Likert-scale, they agreed with the following statement. Some examples of Conscientiousness items include “[I] get chores done right away” or “[I] waste my time.” The NEO IPIP is a shortened, 50 item, personality inventory adapted by Goldberg (1992). This measure has been found to have high internal consistency in previous work ($\alpha = .86$) and correlate highly with the appropriate scales.
RESULTS

The means, standard deviations, sample sizes, and reliability estimates by experimental group can be found in Table 2. To test Hypothesis 1, that SJT’s would be significantly highest in job-relatedness, followed by cognitive ability tests then personality, a one-way Analysis of Variance (ANOVA) was conducted. Levene’s test indicated that the assumption of homogeneity of variance was not met for this data (F = 15.6, p < .001) so a Welch adjusted F ratio (43.2) was obtained. This main effect was significant indicating substantial differences between at least two of the three test conditions in terms of job-relatedness, F(2,268.12) = 43.2, p < .001, η² = .20. To determine how each experimental condition varied by perceived job-relatedness, Dunnet’s T3 post hoc procedure was implemented (Table 3.). This analysis determined that those in the cognitive ability condition believed the test to be less job-related (M = 3.07, SD = 1.0) than those in either the personality condition (M = 3.73, SD = 0.73) or SJT condition (M = 4.02, SD = 0.90). The difference in job-relatedness perceptions between the personality condition and SJT condition was also significant wherein the SJT was viewed as more job-related. As a result, Hypothesis 1 was partially supported in that SJT’s were viewed as the most job-related but cognitive ability tests were deemed as the least job-related. In order to determine if job-relatedness perceptions were significantly related to fairness perceptions, the correlation was evaluated and found to be significant, r (416) = .60, p < .001, supporting Hypothesis 2.
Table 2. Means and Standard Deviations for Study Variables by Test Condition

<table>
<thead>
<tr>
<th></th>
<th>Cognitive Ability Test</th>
<th>Personality Test</th>
<th>Situational Judgment Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>($n = 140$)</td>
<td>($n = 144$)</td>
<td>($n = 141$)</td>
</tr>
<tr>
<td>Job-Relatedness</td>
<td>3.1 (1.0)$^{bc}$</td>
<td>3.7 (0.7)$^{ac}$</td>
<td>4.0 (0.6)$^{ab}$</td>
</tr>
<tr>
<td>Face Validity</td>
<td>3.2 (1.1)$^{bc}$</td>
<td>4.0 (0.8)$^{ac}$</td>
<td>4.4 (0.7)$^{ab}$</td>
</tr>
<tr>
<td>Perceived Predictive Validity</td>
<td>2.9 (1.0)$^{bc}$</td>
<td>3.5 (0.9)$^{a}$</td>
<td>3.6 (0.9)$^{a}$</td>
</tr>
<tr>
<td>Test-Taking Self-Efficacy</td>
<td>4.0 (0.7)$^{b}$</td>
<td>4.3 (0.6)$^{a}$</td>
<td>4.2 (0.7)</td>
</tr>
<tr>
<td>Test Motivation</td>
<td>4.4 (0.7)</td>
<td>4.5 (0.7)</td>
<td>4.5 (0.7)</td>
</tr>
<tr>
<td>Test Performance</td>
<td>25.4 (9.39)</td>
<td>259.6 (35.58)</td>
<td>28.5 (3.82)</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>4.1 (0.8)</td>
<td>4.1 (0.7)</td>
<td>4.0 (0.7)</td>
</tr>
<tr>
<td>Generalized Self-Efficacy</td>
<td>4.2 (0.7)</td>
<td>4.2 (0.7)</td>
<td>4.1 (0.8)</td>
</tr>
<tr>
<td>Fairness</td>
<td>3.58 (1.1)$^{bc}$</td>
<td>4.03 (0.9)$^{a}$</td>
<td>4.14 (0.8)$^{a}$</td>
</tr>
</tbody>
</table>

*Note.* Superscripts indicate significant difference between conditions where $a$ is cognitive ability tests, $b$ is personality test, and $c$ is situational judgement test.
Table 3. Dunnett T3 Comparison for Test Type

<table>
<thead>
<tr>
<th>Test (I)</th>
<th>Test (J)</th>
<th>Mean Diff (I-J)</th>
<th>St. Error</th>
<th>Lower bound</th>
<th>Upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive Ability</td>
<td>Personality</td>
<td>-0.66*</td>
<td>0.11</td>
<td>-0.92</td>
<td>-0.41</td>
</tr>
<tr>
<td>SJT</td>
<td>Personality</td>
<td>-0.95*</td>
<td>0.10</td>
<td>1.2</td>
<td>-0.71</td>
</tr>
<tr>
<td>Personality</td>
<td>Cog</td>
<td>0.66*</td>
<td>0.11</td>
<td>0.41</td>
<td>0.92</td>
</tr>
<tr>
<td>SJT</td>
<td>Cog</td>
<td>-0.29*</td>
<td>0.08</td>
<td>-0.49</td>
<td>-0.09</td>
</tr>
<tr>
<td>SJT</td>
<td>Personality</td>
<td>0.29*</td>
<td>0.08</td>
<td>0.09</td>
<td>0.49</td>
</tr>
</tbody>
</table>

* The mean difference is significant at the .05 level.
Correlations, means, and standard deviations for all study variables, as aggregated between conditions, can be found in Table 4. In order to test Hypotheses 3-5, structural equation modeling was used. First a confirmatory factor analysis (CFA) was conducted in MPlus (Muthen & Muthen, 1998-2012) to assess the distinctness of the latent constructs. The indicators for job-relatedness were the sum score for the construct of face validity and a separate sum score for the construct of perceived predictive validity. All other latent constructs were made up of their individual test items. Test performance was entered as an observable variable. Model fit was assessed using various fit indices including the chi-square statistic wherein a non-significant value indicates good fit; however, the chi-square statistic is very sensitive to sample size. In order to triangulate fit across methods, other fit statistics were also evaluated including the comparative fit index (CFI; Hu & Bentler, 1995) where a value of .90 indicates acceptable fit and the standardized root mean residual (SRMR) which evidences moderate fit when the value is .08 or lower but a more conservative estimate is .05 or below (Hu & Bentler, 1999). Root mean square error (RMSEA) is indicative of moderate fit when the value is below .08 (Browne & Cudeck, 1993). Using these fit statistics and predetermined cut-offs, the original CFA evidenced poor fit ($\chi^2(391) = 1465.11, p < .001; \text{CFI} = .863; \text{SRMR} = .059; \text{RMSEA} = .08, 90\% \text{CI} [.076-.085]$). The model modifications index suggested that substantial misfit could be found from measurement error within the measure of test-taking self-efficacy as two questions shared residual error. Both items seem to measure a similar portion of the construct space (i.e., I can do well/achieve on this test) so for the next CFA they were allowed to correlate. This model evidenced significantly better fit ($\Delta\chi^2(1) = 124.87, p < .05$), but overall fit was still poor ($\chi^2(390) = 1340.24, p < .001; \text{CFI} = .879; \text{SRMR} = .065; \text{RMSEA} = .076, 90\% \text{CI} [.071-.080]$).
### Table 4. Means, Standard Deviations, Correlations, and Reliability Estimates

| Variable                        | M    | SD  | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    |
|---------------------------------|------|-----|------|------|------|------|------|------|------|------|------|------|
| 1. Job-Relatedness              | 3.61 | .9  | .92  | .91  | .92  | .92  | .92  | .92  | .92  | .92  | .92  |
| 2. Face-Validity                | 3.86 | 1.0 | .90* | .91  | .91  | .91  | .91  | .91  | .91  | .91  | .91  |
| 3. Perceived Predictive Validity| 3.36 | 1.0 | .89* | .61* | .87  | .87  | .87  | .87  | .87  | .87  | .87  |
| 4. Test Taking Self-Efficacy    | 4.44 | 0.7 | .50* | .40* | .50* | .50* | .50* | .50* | .50* | .50* | .50* |
| 5. Test Motivation              | 4.31 | 0.8 | .37* | .38* | .27* | .56* | .56* | .56* | .56* | .56* | .56* |
| 6. Test Performance             | -    | -   | .18* | .25* | .06* | .34* | .46* | .46* | .46* | .46* | .46* |
| 7. Conscientiousness            | 4.07 | 0.7 | .20* | .20* | .16* | .37* | .47* | .30* | .30* | .30* | (.90)|
| 8. Generalized Self-Efficacy    | 4.16 | 0.7 | .19* | .15* | .19* | .52* | .46* | .22* | .68* | .68* | (.93)|
| 9. Fairness                     | 3.58 | 1.1 | .60* | .47* | .60* | .52* | .38* | .22* | .29* | .35* | (.90)|

* = p < .01. Note: Reliability estimates in main diagonal; the alpha coefficient for test performance is a range because of the three test conditions: cognitive ability test (α = .87), SJT (α = .69), and personality test (α = .97).

Although the measures utilized in this study have been shown to have good internal consistency (Arvey et al., 1990; Goldberg, 1992; Smither et al., 1993), in order to further evaluate model misfit, an exploratory factor analysis (EFA) was conducted. A principle components analysis was used with oblimin rotation as previous works show many of these constructs to be correlated (Cheng & Ickes, 2009; Hasknecht et al., 2004; Zimmerman, 2000). Because the indicators for job-relatedness were sum scores, these items were not entered into the EFA nor were the observed performance scores. Upon interpretation of the EFA and scree plot (Cattell, 1966), there were 5 substantial factors with an eigen value above 1 accounting for 68% of the variance. The Kaiser-Meyer-Olkin measure of sampling adequacy was .94 and Bartlett’s test of sphericity was significant (χ² (351) = 7336.25, p < .05) supporting reasonable fit.

The first factor explained 42% of the variance, factor 2 explained 10% of the variance, factor 3 explained 8%, and factors 4-5 explained roughly 4% each. Three of the measures loaded as expected with primary loadings being above .5 and cross-loadings below .3. However, the items for Conscientiousness were split for those positively and negatively worded and two of the positively worded Conscientiousness items loaded substantially onto the generalized self-
efficacy factor as well. Because of the cross-loading, the item “make plans and stick to them” was removed and the EFA was run again.

The next EFA had similar loadings and the Kaiser-Meyer-Olkin measure of sampling adequacy was still .94 and Bartlett’s test of sphericity was significant (\(\chi^2(325) = 6901.83, p < .05\)). Again, all measures loaded as expected except there was still one Conscientiousness item with a substantial cross-loading and a negatively worded Conscientiousness factor. The item “carry out my plans” was removed. Then the EFA was run a final time. This EFA consisted of five factors accounted for 69% of the variance with all items loading as expected except for Conscientiousness having both a positively worded and negatively worded factor. Summary results of the final EFA can be found in Table 5.

Research confirms that negatively worded items have the potential to diverge via factor analysis because of differences in response patterns (DiStefano & Molt, 2006). Because the EFA suggested an additional method factor and the measure of Conscientiousness used (i.e., IPIP) has been shown to have good psychometric properties (Donnellan, Oswald, Baird, & Lucas, 2006), this factor was integrated into the CFA so that the latent Conscientiousness factor was created using a composite score for the positive worded items and a separate composite for the negatively worded items. The final CFA evidenced good fit (\(\chi^2(194) = 570.27, p < .001; \text{CFI} = .931; \text{SRMR} = .051; \text{RMSEA} = .068, 90\% \text{ CI [}.061-.074]).\)
Table 5. Summary of Exploratory Factor Analysis Results based on a principle components analysis with oblimin rotation for 4 measures (N = 379)

<table>
<thead>
<tr>
<th></th>
<th>Generalized Self-Efficacy</th>
<th>Conscientiousness (+)</th>
<th>Conscientiousness (-)</th>
<th>Motivation</th>
<th>Test Taking Self-Efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSE1</td>
<td>.847</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE2</td>
<td>.814</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE3</td>
<td>.871</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE4</td>
<td>.898</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE5</td>
<td>.809</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE6</td>
<td>.807</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE7</td>
<td>.655</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE8</td>
<td>.716</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConPos1</td>
<td></td>
<td>.700</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConPos2</td>
<td></td>
<td>.622</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConPos3</td>
<td></td>
<td>.714</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConNeg1</td>
<td></td>
<td>.322</td>
<td>.681</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConNeg2</td>
<td></td>
<td></td>
<td>.788</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConNeg3</td>
<td></td>
<td></td>
<td>.757</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConNeg4</td>
<td></td>
<td></td>
<td>.774</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ConNeg5</td>
<td></td>
<td></td>
<td>.764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Motivation1</td>
<td></td>
<td></td>
<td></td>
<td>.701</td>
<td></td>
</tr>
<tr>
<td>Motivation2</td>
<td></td>
<td></td>
<td></td>
<td>.757</td>
<td></td>
</tr>
<tr>
<td>Motivation3</td>
<td></td>
<td></td>
<td></td>
<td>.835</td>
<td></td>
</tr>
<tr>
<td>Motivation4</td>
<td></td>
<td></td>
<td></td>
<td>.757</td>
<td></td>
</tr>
<tr>
<td>Motivation5</td>
<td></td>
<td></td>
<td></td>
<td>.659</td>
<td></td>
</tr>
<tr>
<td>TTE1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.584</td>
</tr>
<tr>
<td>TTE2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.873</td>
</tr>
<tr>
<td>TTE3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.812</td>
</tr>
<tr>
<td>TTE4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.610</td>
</tr>
<tr>
<td>Eigenvalues</td>
<td>C</td>
<td>1.03</td>
<td>2.04</td>
<td>1.17</td>
<td>2.61</td>
</tr>
<tr>
<td>% of variance</td>
<td>41.1</td>
<td>4.1</td>
<td>8.2</td>
<td>4.7</td>
<td>10.5</td>
</tr>
</tbody>
</table>

Note: Factor loadings < .3 have been suppressed.
All structural models were conducted using maximum likelihood bootstrapping (5,000 samples were drawn) to estimate confidence intervals (95%) and estimate standard errors. Next, the hypothesized model was tested (see Figure 1) and found to have acceptable fit ($\chi^2(198) = 594.96, p < .001; \text{CFI} = .928; \text{SRMR} = .075; \text{RMSEA} = .069, 90\% \text{ CI [.062-.075]}$). Not noted in Figure 1, the control variables were allowed to correlate for all structural models. The control variables were also regressed onto test performance, test-taking motivation, and test-taking self-efficacy. All relationships are reported using standardized parameter estimates.

Job-relatedness perceptions did significantly predict test-taking self-efficacy ($\beta = .60, p < .001$) but did not significantly predict test-taking motivation ($\beta = -.04, p > .05$). Therefore, Hypothesis 5 was supported while Hypothesis 3 was not. In addition, test-taking self-efficacy did predict test-taking motivation ($\beta = .59, p < .01$), further supporting the proposed mediation and Hypothesis 5. Test motivation also significantly predicted test performance ($\beta = .45, p < .001$) in support of Hypothesis 4. The indirect effect of job-relatedness on performance through motivation was evaluated and not significant ($\beta = -.02, (95\% \text{ CI [-.16, .12]}, p > .05$). However, the indirect effect of job-relatedness on test performance through test-taking self-efficacy and motivation was significant ($\beta = .17, (95\% \text{ CI [.02, .28]}, p < .05$). These results were confirmed when controlling for Conscientiousness, which did not significantly predict test-taking self-efficacy ($\beta = -.08, p > .05$) or test performance ($\beta = .12, p > .05$) but did significantly predict test motivation ($\beta = .39, p < .05$). The control variable generalized self-efficacy did significantly predict test-taking self-efficacy ($\beta = .58, p < .01$) but not motivation ($\beta = -.13, p > .05$) or performance ($\beta = -.11, p > .05$). The model accounted for 63% of the variance in test-taking self-efficacy, 51% of the variance in test motivation, and 22% of the variance in test performance. Standardized parameter estimates for the hypothesized model can be found in Figure 2.
Standardized parameter estimates for the control variables as regressed onto test-taking self-efficacy, motivation, and test performance can be found in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>Hypothesized Model</th>
<th>Full Mediation Model</th>
<th>Complex Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Con</td>
<td>GSE</td>
<td>Con</td>
</tr>
<tr>
<td>Test Taking Self-Efficacy</td>
<td>-.08</td>
<td>.58*</td>
<td>-.08</td>
</tr>
<tr>
<td>Motivation</td>
<td>.39</td>
<td>-.13</td>
<td>.38*</td>
</tr>
<tr>
<td>Test Performance</td>
<td>.12</td>
<td>-.11</td>
<td>.13</td>
</tr>
</tbody>
</table>

*p < .01

To determine the extent to which the hypothesized model fit the specific data, two alternative models were also tested. The first alternative was a full mediation wherein test performance was predicted by motivation, motivation was predicted by test-taking self-efficacy, test-taking self-efficacy was predicted by job-relatedness. This model, although more parsimonious, evidenced good fit ($\chi^2(199) = 595.186$, $p < .001$; CFI = .928; SRMR = .075; RMSEA = .068, 90% CI [.062-.075]). The chi-square difference test comparing the hypothesized
model to the more parsimonious alternative model was not significant suggesting that the more parsimonious, full mediation model was more appropriate ($\Delta \chi^2(1) = 0.23, p > .05$). In this model, job-relatedness significantly predicted test-taking self-efficacy ($\beta = .60, p < .001$). Test-taking self-efficacy significantly predicted test motivation ($\beta = .55, p < .001$). Test motivation significantly predicted test performance ($\beta = .45, p < .001$). The indirect effect of job-relatedness on test performance through test-taking self-efficacy and test motivation was also significant ($\beta = .15, 95\% \text{ CI} [.12, .38], p < .001$). The model accounted for 63\% of the variance in test-taking self-efficacy, 50\% of the variance in test motivation, and 22\% of the variance in test performance. Standardized parameter estimates for the more parsimonious full mediation model can be found in Figure 3.

![Figure 3. Standardized path estimates for the full mediation model (N = 411). *p < .01](image)

In addition, a more complex model was tested wherein test performance was directly predicted by job-relatedness and test-taking motivation. Test motivation was predicted by test-taking self-efficacy and job-relatedness. Test-taking self-efficacy was predicted by job-relatedness. The complex model also evidenced good fit ($\chi^2(197) = 594.99, p < .001$; $\text{CFI} = .927$; $\text{SRMR} = .075$; $\text{RMSEA} = .069, 90\% \text{ CI} [.063-.075]$). The chi-square difference test evaluating the hypothesized model and more complex model was not significant ($\Delta \chi^2(1) = 0.03, p > .05$)
suggesting that hypothesized model was more appropriate. However, the chi-square difference test in comparison to the full mediation model was also not significant ($\Delta \chi^2(2) = 0.196, p > .05$).

Standardized parameter estimates for the more complex model can be found in Figure 4. Based on the model comparisons, similarity in fit statistics, and increased parsimony associated with the full mediation model, the alternative full mediation was selected as more appropriate than either the hypothesized model, the more complex model, or the different model.

*Figure 4. Standardized path estimates for the more complex model ($N = 411$).  
*p < .01
As expected, results indicate that selection tests vary in their perceived job-relatedness. However, unlike previous research, which has found cognitive ability tests to be perceived as more job-related than personality tests (Hausknecht et al., 2004; Smither et al., 1993), our findings were that situational judgment tests were perceived as being most job-related followed by personality tests and cognitive ability tests. The differences in findings could potentially be attributed to the specific job used (i.e., customer service worker). Customer service jobs are generally perceived as highly interpersonal. Of the Big Five personality traits, specific to customer service jobs, Agreeableness has been shown to be incrementally predictive (Frei & McDaniel, 1998). It is also possible that the diminished perception of job-relatedness for the cognitive ability test was because of the test used. Although this test has been validated for customer service jobs, the items are related to math, spelling, reading tables/charts, and grammar, which may not seem perceived as directly related to the tasks conducted by customer service workers. On the other hand, the personality test used was compiled of questions directly related to interpersonal competencies and many were contextualized to a work setting potentially improving the perceived job-relatedness. However, the results are mixed regarding the utility of contextualized measures to improve test reactions (Holtop, Born, de Vries, & de Vries, 2014; Holtz, Ployhart, & Dominguez, 2005).

The situational judgment test was deemed as significantly more job-related than the other test types. Because SJTs are a more novel testing strategy than the other two utilized herein, few studies have compared perceptions of these tests to one another. These findings suggest that the use of SJTs may be helpful when fostering positive test reactions. SJTs can be written to evaluate
a variety of abilities (i.e., personality, cognitive ability; Christian, Edwards, & Bradley, 2010) so it may be beneficial to use SJTs when possible.

Again, although the tests did not relate to perceived job-relatedness as expected, the tests did vary in applicant’s perceptions. In addition, perceived job-relatedness did significantly relate to overall fairness perceptions of the testing process. This correlation was not only significant but large ($r = .60$) suggesting that overall fairness may be heavily impacted by the degree to which a test seems related to the job of interest. These findings further indicate that if organizations are interested in applicants’ perceptions of process fairness regarding the selection strategy, it would be wise to utilize testing methods that are directly related to the job.

These findings also indicate that face validity perceptions do impact test performance indirectly through changes in self-efficacy and more distally test motivation after controlling for the effects of Conscientiousness and GSE. This model supports the relationship previously posited by applicant reactions researchers (Hausknecht et al., 2004; Macan et al., 1994; Smither et al., 1993) and found in part by Chan and colleagues (1997). Although, in the hypothesized model it was expected that test-taking self-efficacy would only partially mediate the relationship between reactions and test performance, the full mediation model was found to be more appropriate due the minimal changes in fit statistics and increased parsimony. These findings suggest that, not only is test performance affected by applicant reactions, but this relationship can be explained, at least in part, through changes in test-taking self-efficacy and test motivation.

As noted, self-efficacy, is the degree to which a person feels capable to affect the environment and produce desired outcomes (Bandura, 1977) which, in turn, informs how one feels about a task, how they choose to motivate themselves, the goals they set specific to that task, and their resilience when presented with difficulty (Bandura & Locke, 2003). According to
Social Cognitive Theory, in the case of selection, environmental factors such as reactions to tests and testing methods may impact an applicant’s testing self-efficacy (Bandura, 1997) thus informing intent to perform and test performance outcomes. Although it has been noted that motivation is very complex and made up of variables other than self-efficacy (Schunk, 1991), in the case of applicant reactions, it may be that self-efficacy is the primary driver to one’s motivation which substantially contributes to test performance.

Implications

Overall, this study offers several contributions to the academic literature and the proficiency of organizational selection. First, although the directional relationship between face validity and test performance through motivation has been examined before (Chan et al., 1997), this model has not been tested while considering individual differences nor was self-efficacy introduced as a mediator driving those changes in motivation. It has been argued that negative reactions could reduce the validity of assessments through decreased test performance (e.g. Arvey et al., 1990; McCarthy et al., 2009). However, in the previous study evaluating this relationship (Chan et al., 1997), it was noted that their specific findings could reflect trait motivation, which could, in turn, predict motivation on the job therefore bolstering, rather than hindering test validity. Through the inclusion of individual differences related to both resilience and effort toward achievement (Brown, Lent, Telander, & Tramayne, 2011; Judge & Ilies, 2002) as well as test and job performance (e.g. Barrick & Mount, 1991), we were better able to determine the extent to which the applicant reaction-performance relationship holds devoid of individual differences. The hypothesized model, and the more appropriate fully mediated model, showed good fit regardless of these dispositional differences supporting the notion that reactions matter even for those more likely to achieve thus providing evidence that depleted performance
might cause good applicants to be erroneously selected out of the process diminishing the proficiency of the selection system.

In addition, Chan and colleagues (1997) noted that, in the past, researchers have subsumed individual differences such as motivation and test perceptions as random error when measuring performance thus limiting our ability to improve measurement and fully understand the performance construct. This study provides a more comprehensive understanding of the attitudinal mechanisms related to performance that can enable more appropriate measurement and prediction of job success.

These findings also support the premise that organizations should be using job-related tests. These results, along with the work of previous research, find perceptions of job-relatedness to significantly relate to overall fairness perceptions (Hausknechet et al., 2004). Prior research in this area also show job-relatedness perception to contribute to organizational attractiveness (Smither et al., 1993), intentions to seek legal action (Bauer et al., 1998; Smither et al., 1993), and intentions to accept a job offer (Chapman et al., 2005; Hausknecht et al., 2004) as well as the increased psychometric properties of tests (Chan & Schmitt, 1997). The findings confirm that perceptions of job-relatedness vary by test and potentially by job. Organizations should be considering job-relatedness, specifically face validity and perceived predictive validity, along with other methodological test qualities in the development of effective selection assessments.

In sum, this research supports the notion that both organizations and researchers should consider the perceived job-relatedness when creating selection assessments. As was posited, the validity of assessments could be at risk for those tests deemed as unrelated to the job by applicants (Arvey et al., 1990; McCarthy et al., 2009; Ryan & Ployhart, 2000). However, at this time, the extent to which rank ordering is changed based on the reactions-performance
relationship has not been assessed. It is possible that if all people are equally impacted by job-relatedness perceptions, the effect on performance will wash out making this model unimportant for the proficiency of selection tests. Based on the findings of prior research and this study specifically, future applicant reactions research should evaluate the degree to which reactions impact performance and how that alters the proficiency of selection systems.

Limitations

As noted, this sample was collected from MTurk and did not consist of real job applicants, which could impact the studies ecological validity. In an attempt to reduce this effect, participants were provided with a desirable job description that has been found to have good generalizability when real job applicants are not available (Hunthauser et al., 2003; Schmit et al., 1995). Participants were also incentivized to perform through a substantial bonus payment that has been shown to increase testing motivation more similar to a real job scenario (Chan et al., 1998; Holtz et al., 1998). However, future studies should attempt to recreate this model within a sample of actual job applicants as motivation may more largely derived from the desire to get the job and not the test content.

In addition, responses were aggregated for all experimental groups (i.e., personality, cognitive ability, and situational judgment test). Although, at this time theory suggests that face validity perceptions impact motivation and performance regardless of test type (Gilliland, 1993; Chan & Schmitt, 1997), it is possible that this model works differently for various testing methods. Future research should assess test type as a potential moderator within this model.

Conclusion

Overall, this research indicates that applicant reactions, specifically the degree to which a test seems job-related, impacts test performance through changes in self-efficacy and test
motivation. This relationship was supported while also controlling for achievement related traits further suggesting that reactions may have a significant impact on how one performs and even more importantly, who is selected for the job. Further research should continue to determine the degree to which test reactions impact performance, the potential moderators altering this relationship, and how the proficiency of a selection system is impacted by applicant reactions.
REFERENCES


Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel psychology, 54*(2), 387-419.


Byham (Eds.), *Applying the assessment center method* (pp. 161-183). New York: Pergamon Press.


tell you about justice: The influence of peer communication on applicant reactions. *Revista de Psicología del Trabajo y de las Organizaciones*, 29(1), 37-44.


Leventhal, G. S. (1980). What should be done with equity theory?. In Social exchange (pp. 27-55). Springer US.


*Personality and social psychology*, 952.


Imagine that you are applying for a new job in customer service within an organization that is of great interest to you. This position offers a higher pay, good benefits, and more vacation time than your current position. You are being tasked with taking an assessment to determine fit for this specific job. Please read the directions and take each test seriously as your scores determine acceptance into the hypothetical position of interest as well as a potential for a real pay bonus.
APPENDIX B. SITUATIONAL JUDGEMENT

*The full measure, without the answer key, can be viewed upon request.

The survey presents you with several situations that an employee might encounter at work. Read each situation and the possible actions that one might take in response to each situation. Some actions are effective responses to the situation. Other actions are less effective responses to the situation. Please rate each action on its effectiveness using the following rating scale:

1 - Extremely Ineffective  
2 - Moderately Ineffective  
3 - Slightly Ineffective  
4 - Slightly Effective  
5 - Moderately Effective  
6 - Extremely Effective

<table>
<thead>
<tr>
<th>1. A customer is disrespectful and rude to you</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ask the customer to leave until he is calm.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Treat the customer as the customer treats you.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Try to solve the problem as quickly as possible to get the customer out of the store.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Copyright © 2012 Work Skills First, Inc.
APPENDIX C. COGNITIVE ABILITY TEST

The following assessment is timed. You will have 8 minutes to complete this section. Please read the items carefully and answer as many as you can. You are allowed to use scratch paper but please do not use any technological device such as a calculator or cell phone. An example items is below.

Please select the best answer for the questions below.

1. A customer would like to purchase a videocassette tape for $3.99 and a stuffed animal for $8.93. What is the total for these two items without tax?

   A. $14.18
   B. $10.36
   C. $12.92
   D. $11.47

   Please click next when you are ready to begin this timed assessment.

Example Items:

Answer the following questions by filling in the blank lines with the correct answers. Feel free to use scratch paper to figure out your answers. Please do not include any additional words, numbers, or symbols other than the exact answer.

1. 23 x 11 =

If Statements 1 and 2 are both true, is Statement 3 true or false?

   1. Statement 1: Madison is older than Frances
   2. Statement 2: Penny is older than Madison
   3. Statement 3: Frances is older than Penny

What should be the next number in the series? Please use the forward slash (/) if you need to create a fraction.

1. 3, 9, 15, ?
APPENDIX D. PERSONALITY TEST

Please respond to the items below regarding the extent you either agree or disagree as each statement pertains to you.

Example Items:

1. I hate working with ignorant people.
2. I tend to see difficulties everywhere.
3. I tend to focus on my work more intensely than other people.
4. I prefer not to take risks when choosing a job or an organization to work for.
5. I am not easily distracted.
6. I catch on to things quickly.
APPENDIX E. STUDY MEASURES

Face validity:

Considering the assessment(s) you just took, please respond to the statements below regarding the extent to which you agree or disagree.

- I do not understand what the examination has to do with the job.
- I can not see any relationship between the examination and what is required on the job.
- It would be obvious to anyone that the examination is related to the job.
- The actual content of the examination is clearly related to the job.
- There was no real connection between the examination and the job.

Perceived predictive validity:

- Failing to pass the examination clearly indicates that you can't do the job.
- Performance on the examination would be a good indicator of my ability to do the job.
- The employer can tell a lot about the applicant's ability to do the job from the results of this test.
- I am confident that the test can predict how well an applicant will perform on the job.
- Applicants who perform well on this type of test are more likely to perform well on the job than applicants who perform poorly.

Test-Taking Self-efficacy:

- I believe I can do well on this test.
- I believe I will get the job based on the results of this test.
- I believe I could accurately represent myself as a good employee on this test.
- I have the ability to be successful on this test.

Test Motivation:

- Doing well on this test would be important to me.
- I would try to do the very best I could to on this test.
- I would be extremely motivated to do well on this test.
- I want to be among the top scorers on this test.
- I would not put much effort into this test.

Overall Fairness:

- I think that the testing process is a fair way to select people for the job.
- I think that the test was fair.
- Overall, the method of testing used was fair.
Generalized self-efficacy:

- I will be able to achieve most of the goals that I have set for myself.
- When facing difficult tasks, I am certain that I will accomplish them.
- In general, I think that I can obtain outcomes that are important to me.
- I believe I can succeed at most any endeavor to which I set my mind.
- I will be able to successfully overcome many challenges.
- I am confident that I can perform effectively on many different tasks.
- Compared to other people, I can do most tasks very well.
- Even when things are tough, I can perform quite well.

Conscientiousness:

Please respond to the items below regarding the extent you agree or disagree with the statement as it pertains to you using the scale provided below.

- Am always prepared.
- Pay attention to details.
- Get chores done right away.
- Carry out my plans.
- Make plans and stick to them.
- Waste my time.
- Find it difficult to get down to work.
- Do just enough work to get by.
- Don't see things through.
- Shirk my duties.
APPENDIX F. HSRB APPROVAL

Subject Information and Consent Form

This research is being conducted at Bowling Green State University by the primary investigator, Shelby Wise, and the assistant investigator, Dr. Clare Barratt, in the psychology department.

This research aims to further understand how test reactions, in relation to hiring tests, impact test performance. We hope these findings will inform our understanding of current testing practices within organizations.

Participants should be MTurk workers who are between 18 and 65 years of age and work at least 24 hours a week on average in the United States. The survey contains three screening items to assess fit for this project. You will not be paid for the three screening items. If selected, the full survey should take approximately 15 minutes and participants will be compensated $1.50 for conscientious responding, meaning quality attempts to read directions and complete the survey items.

This study also offers a potential bonus of $0.75. The top 10% of performers (those scoring in the top 10% on the hiring assessment) in each condition will receive the bonus. These assessments have been developed and utilized to hire people within well-known organizations. Scoring is based on data linking assessment items to job performance.

This is completely voluntary and you are free to withdraw at any time with no penalty. Discontinuing participation will in no way impact your relationship with Bowling Green State University. Completion of the survey indicates consent to participate.

We have no access to any of your personal information and this will be completely anonymous. It is recommended that after the completion of the survey you clear your browser and page history.

There are no risks associated with this survey that would be experienced outside of everyday life.

If you have any questions regarding your participation in this study, please contact Dr. Clare Barrett at 419-372-2301 or cbarrat@bgsu.edu or Shelby Wise at srwise@bgsu.edu or 419-372-4201. You may also contact the Chair, Institutional Review Board (IRB) at 419-372-7716 or orc@bgsu.edu, if you have any questions about your rights as a participant in this research.

Thank you for your time. By clicking next, I indicate that:

I have been informed of the purposes, procedures, risks and benefits of this study. I have had the opportunity to have all my questions answered and I have been informed that my participation is completely voluntary. I agree to participate in this research.